Quelques développements récents en traitement du signal

Pierre Comon

HABILITATION
A DIRIGER DES RECHERCHES

UNIVERSITE DE NICE SOPHIA-ANTIPOLIS
U.F.R. SCIENCES

QUELQUES DEVELOPPEMENTS RECENTS EN TRAITEMENT DU SIGNAL

Pierre COMON

Présentée le **18 septembre 1995** devant le jury:

Mme Odile Macchi	Présidente et Rapporteur
Mr Gérard Favier	Examinateur
Mr Michel Granger	Rapporteur
Mr Laurent Kopp	Examinateur
Mr Jean-Louis Lacoume	Rapporteur
Mr Joel Le Roux	Examinateur

Quelques développements récents en traitement du signal

HABITITATION A DIRIGER DES RECHERCHES

Pierre Comon

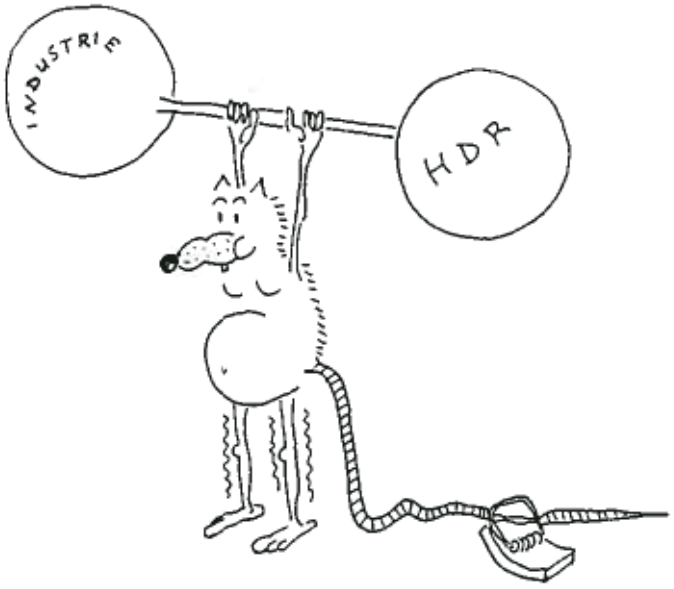
Imprimé le 16 août 1995

Bien évidemment, je remercie les membres du jury pour le temps qu'ils ont accepté de sacrifier à l'analyse de mon travail. J'espère que cette analyse n'aura pas été stérile.

Cependant, ce n'est pas l'essentiel de mon message. En effet un dénominateur commun a réuni les membres du jury: la confiance qu'ils ont bien voulu accorder à certains de mes travaux. Cette confiance est essentielle dans la vie d'un chercheur car sa carence compromet l'efficacité de son travail. Je tiens à remercier chacun d'entre eux pour ce concours implicite.

Je n'oublie pas ma petite famille, qui a souvent dû faire le sacrifice de ses loisirs pour une cause discutable, ainsi que bien d'autres contributeurs indirects sans qui mon travail aurait été entravé. Je pense notamment à G. Bienvenu, défenseur de la recherche amont en milieu industriel.

A mes amis
Ceux que les diplômes indiffèrent
Ceux qui s'éloignent sans s'en rendre compte



Imprimé le 24 juillet 1995

Table des matières

Chapitre 1

Introduction

1.1 Organisation du document

L'habilitation à diriger des recherches est régie par l'arrêté du 23 novembre 1988, modifié par un arrêté du 13 février 1992. Pour plus de clarté, le ministère de l'éducation nationale a publié une circulaire le 27 octobre 1992 afin de prévenir les mauvaises interprétations éventuelles des arrêtés. Il y est notamment précisé que:

L'habilitation n'est pas une thèse. Il s'agit d'une procédure qui [...] doit rester légère. [...] On ne saurait en particulier exiger du candidat [...] la rédaction d'un véritable mémoire ni d'une seconde thèse, après celle du doctorat.

Bien que les textes officiels engagent les candidats à se contenter d'une synthèse rapide suivie de publications, j'ai été encouragé par mon entourage à aller plus loin. Cependant, ne voulant pas non plus m'engager sur la rédaction d'une seconde thèse, j'ai délibérément choisi de ne détailler qu'une partie de mes travaux. Ainsi, le contexte technique et l'état de l'art exposés dans le chapitre 4 ne portent que sur un des quatre volets de ma recherche.

Le chapitre 2 présente de façon synthétique l'ensemble des travaux que j'ai accomplis depuis 1983, et mentionne brièvement mes activités périphériques à la recherche proprement dite, telles que l'organisation d'évènements, l'obtention de contrats, ou les expertises scientifiques. La vocation de ce chapitre n'est pas d'entrer dans des détails techniques, contrairement aux chapitres suivants; elle doit être acceptée comme étant très partiale, puisque limitée à mes propres travaux.

Les chapitres 3 et 4 détaillent un des quatre volets de l'activité présentée

dans le chapitre 2, qui porte essentiellement sur l'usage des statistiques d'ordre supérieur à deux. Pour plus de clarté et dans un souci de complétude, le premier de ces deux chapitres rassemble un certain nombre de résultats en principe connus, mais pour la plupart disséminés dans des ouvrages ou des revues spécialisés. Le second chapitre présente en détail quelques contributions, en prenant cette fois en compte l'état de l'art de façon objective. Les sections 4.1, 4.2, et 4.3 peuvent être vues comme trois projets de recherche, du plus finalisé au plus ambitieux.

Enfin, et puisque ceci est requis d'après les arrêtés ministériels, le chapitre 5 est dédié à la présentation de projets de recherche futurs.

1.2 Présentation succincte

Les activités que j'ai menées depuis une dizaine d'années relèvent essentiellement du domaine du traitement du signal, mais aussi de l'analyse de données, et de l'analyse numérique. Ces activités peuvent être regroupées dans quatre volets qui sont exposés ci-après de façon succincte.

Le premier volet, intitulé "Traitement d'antenne", concerne les transformations effectuées sur les signaux issus d'antennes à capteurs discrets, utilisées en acoustique Sonar et en Radar notamment. Une des tâches à réaliser est l'élimination de bruit de fond, en faisant appel à des opérateurs linéaires, principalement des projecteurs. Ces opérateurs doivent être adaptés à la tâche à effectuer mais aussi aux statistiques des signaux mesurés afin de réduire les erreurs. La versatilité des espaces vectoriels que l'on peut construire autour des signaux observés entraîne autant de projecteurs de nature différente. Une autre tâche majeure (qui rejoint la précédente à bien des égards) est celle de la focalisation électronique de l'antenne dans une direction donnée, afin de réduire l'influence des sources rayonnantes avoisinantes, et notamment des brouilleurs.

Pour ces deux tâches, il est très utile de pouvoir accéder aux performances des traitements, même de façon approchée, en termes de gain en rapport signal à bruit ou de résolution angulaire, par exemple. Par ailleurs, grâce à l'essor des calculateurs, le recours aux statistiques d'ordre élevé est maintenant possible et autorise l'utilisation d'opérateurs non linéaires.

Le deuxième volet concerne les Statistiques d'Ordre Elevé (SOE), et en particulier l'étude du nouveau concept d'Analyse en Composantes Indépendantes (ACI) que j'ai développé en 1990. L'ACI peut être considérée comme une alternative à l'Analyse en Composantes Principales (ACP) en

analyse de données; ce concept a de nombreux avantages sur l'ACP dans un certain nombre de cas de figure, que l'on rencontre en particulier en traitement d'antenne. Toutefois, de sérieux problèmes d'une part théoriques et d'autre part algorithmiques restent encore incomplètement résolus.

J'ai montré notamment que l'ACI ne peut être définie que relativement à la maximisation d'un critère de contraste, et que tous les contrastes ne sont pas équivalents. Certains contrastes jouent cependant un rôle privilégié. L'analyse des contrastes relève de la théorie de l'information. L'ACI est aussi utile dans les problèmes de déconvolution ou d'identification multivariables. En outre l'ACI soulève des problèmes plus généraux sur les décompositions tensorielles (factorisation, diagonalisation).

Le troisième volet comprend l'étude d'algorithmes numériques. Ces algorithmes peuvent être de type "en ligne", c'est à dire traiter les données en temps réel et mettre à jour une solution récursivement (le terme "adaptatif" semble ambigu dans ce contexte), soit de type "hors ligne" et traiter les données en temps différé. Considérons l'espace des matrices dont les dimensions sont de l'ordre de N. On sait qu'un certain nombre de décompositions de matrices nécessitent de l'ordre de N^3 opérations; c'est le cas notamment de la factorisation QR et du calcul des éléments propres. Toutefois, cette évaluation de complexité n'est valable qu'en régime hors ligne. En régime en ligne, la complexité peut être diminuée d'un ordre de grandeur.

De même, si la matrice considérée possède une structure exploitable, telle que Töplitz, Hankel, ou produit de Töplitz, qui sont des structures rencontrées couramment en traitement du signal, la complexité hors ligne peut être également considérablement réduite, au prix parfois d'une perte de stabilité de l'algorithme. De plus, les algorithmes rapides existants tels que celui communément appelé algorithme de Schur, ne permettent pas résoudre au sens des moindres carrés les systèmes linéaires structurés singuliers. Une qualité importante des algorithmes numériques est aussi leur stabilité numérique apparente, fonction à la fois du conditionnement des problèmes, et de l'arithmétique (nécessairement finie) des machines.

D'autre part j'ai évoqué au paragraphe précédent les problèmes algorithmiques rencontrés avec la mise au point de l'ACI. Cette décomposition demande en effet de l'ordre de N^4 à N^6 opérations, suivant la façon dont on procède, cette complexité étant évidemment à mettre en parallèle avec les $O(N^3)$ opérations requises pour le calcul de l'ACP. Le coût élevé de cette décomposition rend encore plus attrayantes les solutions rapides et parallélisables.

Le quatrième et dernier volet est intitulé “Apprentissage supervisé”. En traitement du signal, les problèmes de détection ou d'estimation sont traités habituellement avec l'aide d'un modèle probabiliste. Le cas le plus simple est celui de l'estimation d'un signal noyé dans un bruit, où l'on suppose que le bruit est gaussien. Dans la pratique, l'hypothèse du bruit gaussien additif est parfois beaucoup trop simpliste. On ne peut pourtant recourir à des modèles plus performants que si les connaissances a priori que l'on a du phénomène physique le permettent. Or cela ne représente peut être que la moitié des situations rencontrées. C'est une des raisons pour lesquelles les réseaux de neurones ont fait l'objet de tant d'engouement ces dernières années.

Mon enthousiasme sera plus réservé, pour la simple raison que ces nouvelles techniques ne semblent pas toujours apporter de solutions meilleures que celles que fournissent les approches plus classiques qui, contrairement à ce qui est souvent clamé, peuvent fort bien traiter ce genre de problème. Le contexte des réseaux de neurones est celui de l'apprentissage d'un traitement de façon “supervisée”, c'est à dire à l'aide d'un ensemble d'exemples contenant des couples (entrée, sortie désirée). J'ai montré qu'il est aussi possible d'élaborer des solutions classiques dans un tel contexte, et que ces dernières peuvent être moins coûteuses. En outre, les performances sont plus facilement prédictibles, et on peut s'attendre à ce qu'elles soient plutôt meilleures.

Ces quatre volets vont être maintenant développés et commentés dans le chapitre qui suit. Ceci sera notamment l'occasion d'introduire les publications parues dans la littérature ouverte.

Chapitre 2

Présentation des travaux

Dans ce chapitre, on propose un exposé des principales publications, étayé de courts commentaires techniques de quelques lignes. Les publications sont regroupées par type (articles en langue anglaise, française, actes de conférences...), comme il de coutume de le faire, et classées par ordre chronologique au sein de chaque type (cf chapitre 6). Je reprends successivement dans ce chapitre les quatre volets annoncés en introduction.

2.1 Traitement d'antenne

Une façon d'éliminer le bruit dans une mesure est de placer des capteurs supplémentaires ne mesurant que des bruits. Les signaux ainsi mesurés sont souvent appelés "références de bruit seul". Imaginons par exemple que nous voulions enregistrer une conversation dans une voiture. En plus du microphone situé dans l'habitacle, nous pouvons placer des microphones ou des accéléromètres en différents points stratégiques du moteur ou du châssis. Par régression, il est possible d'estimer le bruit perturbant l'enregistrement et de le soustraire à ce dernier. Pour cette raison, ce procédé exploitant des références de bruit seul est souvent désigné par "soustraction de bruit". Evidemment, un tel procédé est d'autant plus performant que les références de bruit sont corrélées avec le bruit perturbant le signal utile.

J'ai montré dans *[11] [15] [59]* que le rapport signal à bruit peut se dégrader si la régression n'est pas calculée avec suffisamment de précision. Dans ces travaux, j'ai proposé un critère de performance quantitatif pouvant assurer un gain positif. Grâce à ce critère, un filtre "robuste" toujours performant peut être construit. Sans entrer dans les détails, le résultat peut se

résumer ainsi: plus la cohérence entre le bruit perturbateur et les références est faible, plus les observations devront être longues. Dans *[10]*, nous avons montré que la même technique peut être utilisée pour identifier l'ordre de filtres à réponse impulsionnelle finie. Dans *[19]*, j'ai repris de façon plus didactique et plus succincte la technique d'évaluation des résidus dans les problèmes de régression linéaire.

En l'absence de références de bruit, un autre recours est de multiplier les mesures du signal pollué. Par focalisation électronique, on peut ensuite construire un récepteur plus directionnel et réduire ainsi la présence du bruit dans la mesure. Parmi les méthodes de focalisation connues, citons la méthode de Capon maximisant le rapport signal à bruit, et les méthodes Haute-Résolution (HR). Dans *[57]*, nous révisons la présentation usuelle du filtre spatial de Capon, alors que dans *[55]* je me penche sur les performances de la méthode HR la plus répandue, basée sur l'Analyse en Composantes Principales (ACP).

Avant de terminer cette section, on peut mentionner quelques travaux un peu plus généraux pouvant aussi être utilisés en traitement d'antenne. Le premier *[23]* souligne de façon didactique les différences élémentaires souvent ignorées entre la théorie de l'estimation de paramètres réels, et complexes. Quelques exemples sont donnés, notamment lorsque le bruit suit une loi complexe circulaire. De nouvelles définitions de la notion de circularité sont introduites dans *[17]*.

Dans *[61]*, on avait présenté différentes techniques d'extrapolation de signaux lorsque les mesures sont interrompues (pannes d'enregistrement par exemple). Dans *[5]*, un nouvel algorithme d'identification de filtres à Réponse Impulsionnelle Finie (RIF) à phase non minimale est proposé, basé sur des statistiques d'ordre supérieur à deux. Pour des données de courte durée, cet algorithme améliore considérablement les autres approches récemment proposées dans la littérature.

J'ai proposé récemment un sujet de recherche spécifiquement axé sur l'estimation de temps de retard, problème qui apparait fréquemment dans le domaine du sonar, sous des formes plus ou moins complexes. L'idée de départ était de recourir aux SOE *[66] [21]*. Deux pistes avaient été proposées dans *[21]*. La première, basée sur une approche bande étroite, n'avait pas été recommandée en raison de la difficulté que représente la fusion des bandes en présence de bruit. La seconde consistait en l'identification d'un modèle MA ou ARMA monique suivie d'une ACI. Grâce à l'obtention d'une bourse CNRS cofinancée par Thomson, nous avons pu concrétiser cette idée. Une première approche de type spectral *[32]* a été nantie d'une procédure de

retour dans le domaine temps permettant une vraie intégration large bande *[31]*. Une présentation plus élégante est donnée dans *[26]*. Nous avons aussi développé l'autre approche consistant à identifier dans un premier temps un modèle linéaire multivariable, puis à remonter dans un second temps aux retards par interpolation *[27]*.

Je participe (modestement) à la rédaction d'un ouvrage de synthèse sur le traitement d'antenne *[62]*, sous l'impulsion de Laurent Kopp.

2.2 Statistiques d'Ordre Elevé (SOE)

Le concept d'ACI a été initialement suggéré par C.Jutten vers 1987, et je l'ai défini de façon précise en 1989. L'algorithme proposé à l'origine par C.Jutten était présenté de façon heuristique, et utilisait des cellules de calcul neuromimétiques. J'ai dans un premier temps analysé le fonctionnement de cette technique *[51] [6]*, ce qui a révélé sa sous-optimalité. Un autre algorithme a été proposé par ailleurs dans *[48] [50]*, et breveté *[73]*. Dans *[50]*, une version adaptative (en ligne) peu coûteuse était décrite. A ce stade, il s'agissait de séparer des signaux inconnus linéairement mélangés par une transformation instantanée.

Ces travaux ont démontré la généralité du problème et motivé la définition du concept d'ACI, que j'ai d'abord proposé dans *[21]* et *[43]*. Cette définition permet de s'affranchir du modèle d'observation linéaire avec bruit gaussien, sous-jacent dans toutes les autres approches connues abordant le problème de la séparation aveugle de signaux. Ceci a été rendu possible grâce à l'introduction d'une fonctionnelle de "contraste". Les résultats exposés dans *[43]* sont développés plus en détail et de façon plus complète dans *[2]*. La complexité du calcul de l'ACI, comparée pour diverses approches, a été abordée dans *[45]* et *[47]*.

L'ACI est directement applicable aux méthodes HR en traitement d'antenne, ainsi que dans bien d'autres domaines: c'est une décomposition pouvant prendre la place de l'ACP. J'ai esquissé quelques applications dans *[21]* et *[46]*, notamment la détection et la localisation de sources rayonnantes, la compression de données, et la classification. On peut notamment définir le concept d'Analyse en Sous-espaces Indépendants (ASI) *[29]*, qui sera repris en section 2.4. Ce dernier permet de réduire le nombre minimal d'échantillons en dimension $d > 1$, nécessaire pour estimer une densité de probabilité avec une précision donnée.

L'ACI peut aussi être mise à profit dans les problèmes de déconvolution

[2] [43] et d'identification aveugles multivariables *[49] [5]*. Ces utilisations de l'ACI conjointement à d'autres techniques d'identification permettent de mettre en œuvre la séparation aveugle de signaux large-bande linéairement mélangés par une transformation convolutive à phase non minimale (cf. section précédente). Une application prometteuse est celle de l'amélioration du contrôle aérien dans les aéroports civils, notamment celui d'Orly *[3]*.

En essayant d'utiliser les SOE pour mettre en œuvre des algorithmes de localisation haute-résolution, nous avons été confrontés a l'observation suivante: les variables aléatoires spectrales traitées ne possédaient pas toujours la propriété dite de circularité. Cette constatation a été présentée assez formellement dans *[17]*. Mais d'autres travaux ont exploré récemment ce domaine plus en profondeur [176]. En coopération avec J.L.Lacoume, on tentera dans l'ouvrage en préparation *[63]* de donner un aperçu synthétique de l'ensemble de ces propriétés.

Un autre sujet a attiré mon attention ces dernières années, à force de pratiquer l'usage des SOE. En effet, il est clair qu'il est des domaines plus fertiles pour les SOE (comme les télécommunications) que d'autres (comme le sonar). La question qu'on aimerait se poser avant de lancer une étude de performances d'un algorithme sur une base de données réelle est de savoir si les données s'écartent suffisamment du caractère gaussien pour qu'il y ait un espoir de faire mieux que les techniques classiques d'ordre deux. Or, la quasi-totalité des tests de gaussianité supposent que les signaux à tester sont blancs. En outre, peu nombreuses, les techniques restantes demeurent très gourmandes en calculs, et exigent des durées d'intégration importantes, ce qui peut être incompatible avec la durée de stationnarité des phénomènes observés. C'est pourquoi j'ai développé un test original, qui reste valable quand on s'écarte de l'hypothèse de blancheur (dans la pratique, les spectres des signaux doivent être continus et à large support). Les résultats semblent pour l'instant encourageants *[30]*.

Enfin, les cumulants (ou les moments) de variables aléatoires vectorielles peuvent être vus comme des tenseurs symétriques. Avec cette vision subjective, un problème tel que l'ACI devient un problème de diagonalisation tensorielle *[35]*. Or, s'il y a apparemment très peu d'ouvrages connus sur les décompositions en algèbre multilinéaire, de nombreux travaux ont vu le jour depuis le début du siècle sur les polynômes homogènes (Gauss, Cayley, Noether, Dieudonné...). C'est alors en remarquant qu'on peut associer bijectivement tout tenseur symétrique à un polynôme homogène, qu'on peut transposer de nouveau le problème *[34]*. Avec cette approche, au lieu de chercher à diagonaliser approximativement un tenseur symétrique *[36]*, on

s'aperçoit que l'on peut le diagonaliser exactement, mais en général de plusieurs manières *[24]*.

2.3 Algorithmes numériques

J'ai expliqué dans l'introduction de ce document, ainsi que dans la première section de ce chapitre, le rôle que joue l'ACP en traitement d'antenne, et plus généralement en traitement du signal. Le calcul des éléments propres a été longtemps considéré comme prohibitif en raison de son coût élevé, i.e, de l'ordre de N^3 opérations pour une matrice symétrique $N \times N$. Toutefois, cette complexité peut être considérablement décrue lorsque cette décomposition doit être calculée en ligne, ou lorsqu'un petit nombre d'espaces propres dominants sont requis *[55] [54] [53] [52]*. De plus, ces solutions n'interdisent pas d'implantation parallèle pour autant *[12]*. L'article *[7]* récapitule les résultats que j'ai obtenus et présente une étude comparative de divers algorithmes adaptatifs existants.

Dans le contexte de l'ACI, nous avons suggéré la même famille d'algorithmes dans [45] pour calculer les matrices propres de l'opérateur quadricovariance, afin de réduire la complexité.

En régime non stationnaire (localement stationnaire), le calcul de la régression linéaire dans le domaine spectral (pour la soustraction de bruit par exemple) nécessite la résolution d'une suite de systèmes linéaires voisins les uns des autres. Lorsque ces systèmes sont pleins, nous avons proposé un certain nombre de solutions *[13] [9]*. En revanche, en régime hors ligne, on peut également utiliser une des architectures parallèles maintenant bien connues *[8] [14] [22]*.

En régime stationnaire (ou presque stationnaire au sens des rangs de déplacements), le calcul de la régression linéaire requiert la résolution de systèmes structurés (Töplitz par exemple). Les algorithmes à notre disposition sont principalement issus de ceux qui sont maintenant connus sous le nom de Levinson et de Schur. Malheureusement, ces derniers ne fonctionnent efficacement que sur des classes de systèmes fortement réguliers. Pour cette raison, il est important de savoir évaluer la stabilité de ces algorithmes en fonction du conditionnement des systèmes. Dans *[4] [41]*, nous avons montré que même les systèmes structurés singuliers devraient pouvoir être résolus rapidement au sens des moindres carrés, car une forme de structure est préservée dans le calcul de la pseudo-inverse. D'autres travaux sont en cours sur ce sujet *[1]*.

Mais la rapidité n'est pas le seul critère à prendre en compte dans la mise au point d'un algorithme. La robustesse numérique est au moins aussi importante. Il se trouve précisément qu'en général plus les algorithmes sont rapides, plus ils sont numériquement instables. Nous examinons dans *[16]* la stabilité numérique de l'algorithme de Levinson, un des algorithmes les plus utilisés pour la résolution des systèmes Töplitz. Ces travaux mériteraient d'être poursuivis également pour l'algorithme de Schur, et les algorithmes correspondants adaptés aux systèmes Töplitz par blocs.

2.4 Apprentissage supervisé

Comme je l'ai précisé en introduction, l'apprentissage supervisé consiste à identifier une relation entre deux ensembles de variables, $\mathcal{E}$ et $\mathcal{F}$, à partir d'exemples $\{x(n) \in \mathcal{E}, y(n) \in \mathcal{F}, 1 \leq n \leq N\}$ en nombre fini. Considérons le cas d'une application de $\mathcal{E}$ dans $\mathcal{F}$. Il est souvent clamé par une partie de la communauté que les réseaux de neurones sont parfaitement adaptés à ce genre de problème, puisque c'est également de cette façon qu'un enfant apprend, supervisé par ses parents. Sans réfuter cette affirmation, je ne pense pas que les techniques classiques soient incapables d'aborder ce genre de problème. J'ai montré dans *[20]* comment les problèmes de classification (où le cardinal de $\mathcal{F}$ est fini) pouvaient fort bien être résolus par l'approche bayesienne.

J'ai suggéré également une méthode permettant d'identifier des applications plus générales *[72]*. Cette méthode peut aussi être utilisée pour identifier des relations multivoques entre ensembles. Pour devancer les critiques, l'implantation de l'apprentissage sur un réseau de cellules est également possible dans l'approche proposée *[18]*. Il convient donc de comparer cette approche "classique" et l'approche "neuronale". C'est ce que j'ai tenté de faire dans *[18]*, en me limitant au réseau de neurones le plus répandu, le "Perceptron MultiCouche" (PMC). J'ai analysé en détail ses défauts et ses qualités.

Ma conclusion penche en faveur de mon approche, pour essentiellement deux raisons. La première est qu'il est difficile de prédire les performances du PMC sans faire appel à une interprétation probabiliste, ce qu'il est malgré tout possible de faire lorsque la taille du réseau PMC et le nombre d'exemples tendent tous deux vers l'infini *[44] [37] [25]*. Cette analyse montre aussi au passage que les performances du PMC sont toujours moins bonnes que la solution bayesienne. La seconde raison est que l'apprentissage du PMC est très lent: le temps requis pour atteindre une précision donnée ϵ, n'est pas

fonction polynomiale de ϵ, même dans le cas très favorable où le problème est convexe *[18]*. Dans le jargon de la théorie de l'apprentissage, on dirait que l'apprentissage du PMC est NP (Non Polynomial). On peut se référer à [126] pour une introduction acceptable aux réseaux de neurones, et à [125] pour une présentation moderne du point de vue traitement du signal.

Cependant, l'estimation des densités de probabilité soulève des problèmes d'ordre pratique *[33]*. En effet, les résultats connus sont essentiellement de nature asymptotique *[29]*; malgré les nombreux travaux menés par les statisticiens, que je mentionne dans *[29]*, le choix de certains paramètres est encore laissé sous le contrôle de l'intuition, faute de mieux. Par exemple, dans les estimateurs à noyau fixe, on sait que le facteur largeur doit être de l'ordre de $N^{-1/d+4}$, si d est la dimesion de l'espace, et N le nombre d'exemples. Mais le coefficient de proportionnalité est difficile à déterminer, car il dépend du Laplacien de la densité cherchée.

Un autre obstacle pratique à l'utilisation des estimateurs à noyau est leur grande exigeance en place mémoire pour de grands échantillons. Nous avons proposé dans *[28]* une approche tirant parti d'un groupement automatique préalable. L'idée est que, dans certains problèmes, le temps d'apprentissage ne compte pas, mais l'optimalité de l'exploitation de la mémoire est d'importance. Par exemple, un constructeur de jouets, désirant mettre au point une carte de reconnaissance de la parole bon marché, imposera des contraintes de mémoire importantes en mode opérationnel, mais ne tiendra pas compte du fait que l'apprentissage prend 40mn sur un PC, ou 3 jours sur une station UNIX, si cet apprentissage se fait une fois pour toute en usine.

Enfin, le dernier problème auquel je me suis interessé dans ce domaine est celui de la taille minimale de la base de données. Pour estimer une densité en dimension d, il est clair qu'il faudra un nombre d'exemples fonction exponentielle de d. Certains auteurs ont donné des indications pratiques dans ce sens. Notamment, il est raisonnable d'admettre que le nombre minimal d'exemples N_{min} en dimension d est donné par $\log_{10} N_{min} = 0.6\,(d - \frac{1}{4})$. Evidemment, cette règle est valable si aucune modèle paramétrique ne désire être utilisé. De plus cette règle est une condition suffisante, permettant d'obtenir une estimation de variance relative acceptable. Elle peut ne pas être nécessaire, en particulier si les données s'avèrent être concentrées autour d'une variété de dimension inférieure à d. Le problème est que ceci n'est pas connu à l'avance, dans la quasi-totalité des cas.

Or il arrive fréquemment qu'on ait à construire un classifieur en dimension trop élevée par rapport à la borne précédente (exemple: $d = 10 \Rightarrow$

$N = 708000$). L'idée proposée dans *[2] [29]* lorsqu'on est confronté à ce problème est la suivante. On cherche un changement de base (inversible, mais non orthogonal) de façon à ce que la variable aléatoire z de dimension d puisse se décomposer en deux variables aléatoires x et y approximativement indépendantes de dimension plus faibles dans le nouveau système:

$$p_z(A\,u) \approx p_x(u_1)\,p_y(u_2).$$

Cette décomposition a été baptisée Analyse en Sous-espaces Indépendants (ASI), par analogie à l'ACI décrite plus haut, et fera l'objet de coopérations avec d'autres laboratoires français et étrangers.

2.5 Autres travaux

Expérience industrielle Le poste que j'occupe depuis plus de sept ans à Thomson requiert avant tout l'obtention de contrats de recherche. En effet, la pérennité de l'activité de "recherche amont" n'est assurée qu'à ce prix, les autres activités de la société étant conditionnées par leur rentabilité à court terme. En outre, la conjoncture économique ne fait qu'exacerber ces contraintes.

La circulaire ministérielle du 27 octobre 1992 demande que soient mentionnées dans le document d'habilitation les obtentions de contrats de recherche. Il est difficile pour certains universitaires d'imaginer le coût que représente une étude dans l'environement de travail qui est celui d'une grande entreprise comme Thomson. En sept ans, j'ai démarché une dizaine de contrats pour un budget total de l'ordre de 8 millions de francs hors taxes (les financements internes ne sont pas inclus dans ces chiffres). Un tel budget est considéré comme modéré dans l'industrie.

Par ailleurs, j'ai eu l'honneur de rentrer au collège scientifique de Thomson S.A. en 1992.

Vie universitaire Sur le plan universitaire, j'ai encadré une thèse sur le thème des algorithmes rapides pour les systèmes structurés, qui a été soutenue en septembre 1993 (E. Kazamarande), et en encadre une autre actuellement sur le thème de l'estimation de temps de retards différentiels (B. Emile). Il n'est malheureusement pas envisageable d'encadrer plus d'un étudiant à la fois tout en conservant comme emploi principal la fonction qui est la mienne à Thomson-Sintra.

J'ai également participé au jury de 12 thèses, la plupart du temps en tant que rapporteur.

Enseignement De septembre 1981 à juin 1982, j'ai assuré pendant une année scolaire des travaux dirigés d'algèbre élémentaire, tournée vers les applications en électricité. Entre octobre 1982 à mars 1984, j'ai été responsable pendant deux années consécutives de travaux dirigés à l'INPG, sur la théorie des distributions et leur application dans le monde de l'ingénieur. Entre octobre 1989 et mars 1993, j'ai assuré la dispense d'un petit cours-TD à l'ESSI sur la théorie de la détection et le traitement du signal Sonar. A partir de la rentrée 1995, je serai chargé d'un cours de DEA.

Organisation d'évènements J'ai été convié à participer à l'organisation du "Workshop on High-Order Statistics" qui s'est tenu à Chamrousse en juillet 1991. J'ai co-organisé les conférences IEEE sur les statistiques d'ordre élevé en juin 1993 à Lake Tahoe, Californie, et en juin 1995 à Begur, Espagne. J'ai également organisé une session spéciale d'une journée à la conférence SPIE à San Diego en juillet 1994.

J'ai été l'instigateur du working group "ATHOS", action Esprit BRA notifiée en juillet 1992. Ce consortium a permis de fédérer certains efforts du GDR-TDSI au delà des frontières de l'hexagone, dans le domaine des statistiques d'ordre élevé.

Expertises J'expertise très régulièrement des articles soumis à des revues scientifiques telles que: *IEEE Transactions on Signal Processing*, la revue européenne *Signal Processing*, ou la revue française *Traitement du Signal*. A titre plus occasionnel, j'expertise aussi des articles pour les revues *IEEE Transactions on Information Theory*, *IEEE Transactions on Circuits and Systems*, *Neural Computation*, ou *SIAM Journal on Matrix Analysis*.

Ces analyses représentent un volume de travail non négligeable dans l'activité de recherche, surtout si elles sont nombreuses (au moins douze par an). A titre occasionnel, j'ai été également convié à donner des avis scientifiques sur des soumissions de projets à la CEE (Esprit BRA, Esprit LTR). Par ailleurs, je suis amené de temps à autre à expertiser des soumissions pour des conférences que je n'ai pas co-organisées (notamment pour Eusipco92, Gretsi93, Eusipco94, et Gretsi95).

Invitations Outre quelques invitations à des séminaires privés en France (e.g. séminaire annuel du Campus Thomson en mars 1990 et avril 95) ou à l'étranger (e.g. séminaires EPFL à Lausanne en mai 1991 et novembre 1994), j'ai été invité à présenter mes travaux à plusieurs reprises, notamment à la conférence SPIE qui s'est tenue à San Diego en juillet 1990, au congrès WHOS de Chamrousse en juillet 1991, et plus récemment à la conférence ESANN en avril 1995 à Bruxelles. J'ai été invité pendant 1 mois par l'IMA (Institute for Mathematics and its Applications) à Minneapolis pour le séminaire "Linear algebra for signal processing" en avril 1992. J'ai aussi quelquefois participé à des "sessions invitées" de conférences internationales.

Chapitre 3

Introduction aux SOE

Les Statistiques d'Ordre Elevé (SOE), autrement dit, les moments et cumulants d'ordre supérieur à 2, sont utilisées essentiellement en complément aux statistiques d'ordre 2, afin de permettre la résolution de problèmes restés insolubles jusqu'alors. L'identification de modèles MA multivariables fait partie de ces problèmes [194]. En outre, les SOE ont été ensuite (et plus récemment) exploitées pour améliorer les solutions (conditionnement, identifiabilité..) déjà apportées par les techniques classiques.

Ce chapitre est surtout destiné au néophyte. Son but est de donner les définitions et propriétés nécessaires à l'introduction et à l'estimation des SOE.

3.1 Variables aléatoires réelles scalaires

Soit X une variable aléatoire à valeurs dans $\mathbb{R}$ (le cas de variables complexes sera abordé plus loin dans la section 3.3). On notera $F_x(u)$ sa fonction de répartition et on supposera généralement que X admet une densité de probabilité $p_x(u)$. Autrement dit, nous aurons $dF_x(u) = p_x(u)\,du$. Rappelons que $p_x(u)$ est positive et a pour somme l'unité. Si $F_x(u)$ est une fonction en escalier, elle n'admet pas de densité (la densité n'existe qu'au sens des distributions). Les moments généralisés de X sont définis pour toute application réelle g par:

$$E\{g(X)\} = \int_{-\infty}^{+\infty} g(u)\, p_x(u)\, du. \tag{3.1}$$

Dans la pratique, on utilise surtout des fonctions polynômiales $g(u)$, conduisant aux moments "classiques" d'ordre n, tels que la moyenne ou la vari-

ance de X, mais également des fonctions exponentielles. C'est ainsi que l'on associe "des fonctions caractéristiques" aux variables aléatoires.

La *première fonction caractéristique* de X est:

$$\Phi_x(v) = E\{\, e^{\jmath v X} \,\}, \tag{3.2}$$

où $\jmath$ désigne la racine de -1. La fonction caractéristique $\Phi(v)$ est toujours continue et vaut 1 à l'origine. Elle est donc non nulle dans un voisinage de l'origine, sur lequel on pourra définir son logarithme néperien:

$$\Psi_x(v) = \log(\, \Phi_x(v) \,). \tag{3.3}$$

Cette nouvelle fonction est communément appelée *seconde fonction caractéristique*. Lorsque X admet une densité, $p_x(u)$, on peut remarquer que $\Phi_x(v)$ n'est autre que sa transformée de Fourier:

$$\Phi_x(v) = \int_{-\infty}^{+\infty} e^{\jmath v z}\, p_x(z)\, dz. \tag{3.4}$$

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

$$p_x(z) = \int_{-\infty}^{+\infty} e^{-\jmath v z}\, \Phi_x(v)\, dv. \tag{3.5}$$

Notons $\mu'_{(r)}\{X\}$ les moments d'ordre r de X, lorsqu'ils existent:

$$\mu'_{(r)}\{X\} = E\{\, X^r \,\}, \tag{3.6}$$

et $\mu_{(r)}\{X\}$ ses moments centrés:

$$\mu_{(r)}\{X\} = E\{\, (\, X - \mu'_1 \,)^r \,\}, \tag{3.7}$$

Les fonctions caractéristiques décrivent complètement la variable aléatoire à laquelle elles sont associées. En particulier, ses moments peuvent être obtenus à partir des dérivées successives de $\Phi_x(v)$ à l'origine:

$$\mu'_{(r)}\{X\} = (-\jmath)^r \left.\frac{d^r \Phi_x(v)}{dv^r}\right|_{v=0}. \tag{3.8}$$

Les dérivées de la seconde fonction caractéristique prises à l'origine donnent les *cumulants*:

$$\mathcal{C}_{(r)}\{X\} = (-\jmath)^r \left.\frac{d^r \Psi_x(v)}{dv^r}\right|_{v=0}. \tag{3.9}$$

On montre [136] que les cumulants d'ordre n peuvent être calculés à partir des moments d'ordre inférieur ou égal à n:

$$\begin{aligned}
\mathcal{C}_{(1)}\{X\} &= \mu'_{(1)}, && (3.10)\\
\mathcal{C}_{(2)}\{X\} &= \mu_{(2)} = \mu'_{(2)} - \mu'^2_{(1)}, && (3.11)\\
\mathcal{C}_{(3)}\{X\} &= \mu_{(3)} = \mu'_{(3)} - 3\,\mu'_{(1)}\,\mu'_{(2)} + 2\,\mu'^3_{(1)}, && (3.12)\\
\mathcal{C}_{(4)}\{X\} &= \mu'_{(4)} - 4\,\mu'_{(3)}\,\mu'_{(1)} - 3\,\mu'^2_{(2)} + 12\,\mu'_{(2)}\,\mu'^2_{(1)} - 6\,\mu'^4_{(1)} && (3.13)
\end{aligned}$$

Dans le cas de variables centrées ($\mu'_1 = 0$), les expressions se simplifient:

$$\begin{aligned}
\mathcal{C}_{(1)}\{X\} &= 0, && (3.14)\\
\mathcal{C}_{(2)}\{X\} &= E\{X^2\}, && (3.15)\\
\mathcal{C}_{(3)}\{X\} &= E\{X^3\}, && (3.16)\\
\mathcal{C}_{(4)}\{X\} &= E\{X^4\} - 3\,E\{X^2\}^2. && (3.17)
\end{aligned}$$

La relation (3.9) montre que les cumulants sont les coefficients du développement en série de Taylor de la seconde fonction caractéristique. Lorsque la variable X est gaussienne, sa seconde fonction caractéristique est

$$\Psi_x(v) = \jmath\mu'_{(1)}\,v - \frac{1}{2}\,\mu_{(2)}\,v^2, \qquad (3.18)$$

ce qui montre que ses cumulants d'ordre supérieur à 2 sont tous nuls. Inversement cette propriété caractérise la loi gaussienne [134]. On peut donc en déduire que les variables gaussiennes sont entièrement décrites par leurs propriétés au second ordre. Ceci explique pourquoi les chercheurs en traitement du signal se sont longtemps limités au second ordre. En "invoquant" le théorème de la limite centrale, on peut penser que la plupart des signaux ont tendance à être gaussiens, mais ce point de vue est erroné. Nous aurons l'occasion d'y revenir.

La variance de X, $\mathcal{C}_{(2)}\{X\}$ caractérise la *puissance* de X. Les quantités $\mathcal{C}_{(3)}\{X\}$ et $\mathcal{C}_{(4)}\{X\}$ caractérisent respectivement *l'asymétrie* et *l'aplatissement* de la loi, en prenant la loi gaussienne comme référence. Afin de rendre ces mesures indépendantes de la variance, on a coutume d'utiliser des grandeurs *standardisées* parfois appelées facteur d'asymétie (*skewness* en anglais) et facteur d'aplatissement (ou *kurtosis*, mot de racine grecque[1]

[1] $\kappa\acute{\upsilon}\rho\tau\omega\sigma\iota\varsigma$: action de courber, convexité.

dont l'utilisation est autorisée en français), définies de la façon suivante:

$$\mathcal{K}_{(3)}\{X\} = \mathcal{C}_{(3)}\{X/\sqrt{\mu_2}\} \quad (3.19)$$
$$\mathcal{K}_{(4)}\{X\} = \mathcal{C}_{(4)}\{X/\sqrt{\mu_2}\}. \quad (3.20)$$

Pour une variable centrée, les facteurs d'asymétrie et d'aplatissement s'écrivent:

$$\mathcal{K}_{(3)}\{X\} = \frac{E\{X^3\}}{E\{X^2\}^{3/2}}, \quad (3.21)$$
$$\mathcal{K}_{(4)}\{X\} = \frac{E\{X^4\}}{E\{X^2\}^2} - 3. \quad (3.22)$$

Exemple: La variable aléatoire uniformément répartie dans $[-aa]$ a pour fonction caractéristique $\Phi(u) = \frac{\sin au}{au}$, pour moments d'ordre pair $\mu'_{(r)} = \frac{a^r}{r+1}$, et pour kurtosis $\mathcal{K}_{(4)} = -\frac{6}{5}$. On peut trouver dans [169] les coefficients d'asymétrie et d'aplatissement obtenus pour quelques distributions standard.

3.2 Cas vectoriel, multicorrélations

Sauf mention contraire, on supposera dorénavant que les variables aléatoires sont **centrées**. On pourra représenter les variables aléatoires à plusieurs dimensions par un vecteur colonne:

$$X^T = (X_1, X_2 \dots X_n)\,. \quad (3.23)$$

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

$$\Phi_x(v) \stackrel{\text{def}}{=} E\{e^{j\Sigma v_n x_n}\} = E\{e^{jv^T x}\}. \quad (3.24)$$

Si les composantes x_n du vecteur aléatoire x admettent une densité conjointe $p_x(u)$, alors la fonction caractéristique de x est donnée par la Transformée de Fourier de sa densité:

$$\Phi_x(v) = \int_{\boldsymbol{R}^N} e^{jv^T u}\, p_x(u)\, du. \quad (3.25)$$

La seconde fonction caractéristique garde la même définition: $\Psi_x(v) = \log \Phi_x(v)$. Ces fonctions peuvent de nouveau servir à générer les moments et les cumulants.

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

$$C_{ij} = \mathcal{C}_{(2)}\{X_i, X_j\} = E\{X_i\, X_j\}.$$

Lorsqu'on manipule des données à plusieurs dimensions, nous voyons qu'il peut devenir inutile de préciser l'ordre du moment ou du cumulant considéré. Ainsi, la matrice de covariance peut être écrite:

$$\mathcal{C}_{ij}\{X\} = \mathcal{C}\{X_i, X_j\}. \tag{3.26}$$

Il n'y aura jamais ambiguité sur la signification d'un indice puisque lorsqu'il indique l'ordre, il figure entre parenthèses. De même, les cumulants d'ordre plus élevé pourront être souvent notés de façon plus compacte, en omettant d'indiquer l'ordre lorsque ce dernier correspond au nombre de variables entre accollades. Par exemple, on notera:

$$\mathcal{C}_{ijk}\{X\} = \mathcal{C}\{X_i, X_j, X_k\}, \tag{3.27}$$
$$\mathcal{C}_{iii}\{X\} = \mathcal{C}_{(3)}\{X_i\}, \tag{3.28}$$
$$\mathcal{C}_{hijk}\{X\} = \mathcal{C}\{X_h, X_i, X_j, X_k\}, \tag{3.29}$$
$$\mathcal{C}_{iiii}\{X\} = \mathcal{C}_{(4)}\{X_i\}. \tag{3.30}$$

Ces notations étant précisées, il est facile de voir, en développant l'exponentielle $e^{\jmath v^T x}$ en série autour de $v = 0$, que les coefficients des termes de degré r, $v_i v_j..v_k$, sont $\jmath^r\, \mu_{ij..k}/r!$ [136] [160], ce qui implique que:

$$\mu'_{i_1 i_2..i_r}\{x\} = (-\jmath)^r \left.\frac{\partial^r \Phi_x(v)}{\partial v_{i_1} \partial v_{i_2}..\partial v_{i_r}}\right|_{v=0}. \tag{3.31}$$

Il est inutile de réécrire cette relation pour les cumulants car elle se déduit de celle-ci en remplaçant Φ par Ψ.

Comme dans le cas scalaire, il est possible d'établir des égalités liant moments et cumulants en développant la fonction log en série entière. On obtient par exemple:

$$\mathcal{C}_{ij}\{x\} = \mu'_{ij}\{X\} - \mu'_i\{X\}\mu'_j\{X\}. \tag{3.32}$$

Pour décrire l'ensemble de ces relations de façon plus complète, il est nécessaire d'introduire des conventions d'écriture, sans quoi nous aurions vite fait d'aboutir à des pages de termes, d'ailleurs très semblables les uns aux autres.

On conviendra d'écrire une somme de k termes se déduisant les uns des autres par permutation d'indices par une *notation crochet*. Quelques bons exemples valent mieux qu'un long discours:

$$[3]\,\delta_{ij}\,\delta_{kl} = \delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}, \tag{3.33}$$
$$[3]\,a_{ij}\,b_k\,c_{ijk} = a_{ij}b_kc_{ijk} + a_{ik}b_jc_{ijk} + a_{jk}b_ic_{ijk}. \tag{3.34}$$

La présence du crochet entraine donc une sommation implicite. On suppose toujours que les termes à r indices (qui sont des tenseurs d'ordre r) sont complètement symétriques. Pour que la notaton soit correcte, il faut que le nombre de monômes distincts que l'on puisse obtenir par permutation soit égal à l'entier figurant entre crochets. Ainsi, les écritures:

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

sont correctes. Les cumulants d'ordre 3 et 4 sont alors donnés en fonction des moments par les expressions compactes:

$$\mathcal{C}_{ijk} = \mu'_{ijk} - [3]\mu'_i\mu'_{jk} + 2\mu'_i\mu'_j\mu'_k, \tag{3.35}$$
$$\mathcal{C}_{ijkl} = \mu'_{ijkl} - [4]\mu'_i\mu'_{jkl} - [3]\mu'_{ij}\mu'_{kl} + 2\,[6]\mu'_i\mu'_j\mu'_{kl} - 6\mu'_i\mu'_j\mu'_k\mu'_l. \tag{3.36}$$

Dans le cas centré, ces expressions se simplifient et on a:

$$\mathcal{C}_{ij} = \mu_{ij}, \tag{3.37}$$
$$\mathcal{C}_{ijk} = \mu_{ijk}, \tag{3.38}$$
$$\mathcal{C}_{ijkl} = \mu_{ijkl} - [3]\mu_{ij}\mu_{kl}. \tag{3.39}$$

Il est intéressant de comparer ces expressions avec celles que l'on a obtenu dans le cas scalaire.

De façon plus générale, les cumulants sont liés aux moments par la formule de Leonov et Shiryaev (donnée ici à l'ordre r):

$$\mathcal{C}\{X_1,...,X_r\} = \sum(-1)^{k-1}(k-1)!\,E\{\prod_{i\in v_1} X_i\}\cdot E\{\prod_{j\in v_2} X_j\}\cdots E\{\prod_{k\in v_p} X_k\}, \tag{3.40}$$

où la sommation s'étend sur tous les ensembles $\{v_1, v_2, \cdots, v_p; 1 \leq p \leq r\}$ formant une partition de $\{1, 2, \cdots, r\}$. Cette expression s'étend au cas où les $\{v_i\}$ ne décrivent plus nécessairement toutes les partitions; on parle alors de cumulants généralisés [160, page 60].

Multicorrélations

A l'instar de la fonction de corrélation d'ordre 2, on peut définir des fonctions de multicorrélation d'ordre supérieur. Aux ordres 2 et 3, les moments centrés et les cumulants sont confondus, de sorte qu'il n'existe pas d'ambiguité. En revanche aux ordres plus grands, il faudra prendre garde à préciser s'il s'agit de "multicorrélations cumulantes" ou non.

Lorsque ce n'est pas spécifié, on considère en général qu'il s'agit multicorrélations cumulantes, par défaut. A l'ordre r on définit par exemple:

$$\mathcal{C}_{X,i_1 i_2..i_r}(t;\tau_2,..,\tau_r) = \mathcal{C}\{X_{i_1}(t), X_{i_2}(t+\tau_2), ..., X_{i_r}(t+\tau_r)\}. \qquad (3.41)$$

Un processus réel à temps discret de dimension N, $X(t)$, $t \in \mathbb{Z}$, est dit usuellement *stationnaire* (ou *fortement stationnaire*, ou *stationnaire au sens strict*) si et seulement si l'ensemble des propriétés statistiques conjointes des vecteurs $[X_{a_1}(t_1+t), ..X_{a_k}(t_k+t)]$ ne dépend pas de la date t, et ce pour tout $k \in \mathbb{N}$, tout k-uplet $(a_1, ..a_k)$, $1 \leq a_j \leq N$, et tout k-uplet $(t_1, .., t_k)$, $t_j \in \mathbb{Z}$. Cette définition est très forte et n'est pas toujours requise. Une définition bien connue est celle de la stationnarité jusqu'au second ordre (ou stationnarité au sens large, dite faible), qui requiert que la moyenne $\mu_X = E\{X(t)\}$ et la fonction de corrélation $\mathcal{C}_{X,ij}(\tau) = cum\{X_i(t), X_j(t+\tau)\}$ soient finies et qu'elles ne dépendent pas de la date t [175].

De la même manière, on peut définir la stationnarité à l'ordre r [85] [180] [205] [89]:

Définition 3.2.1 *Un processus réel à temps discret de dimension N, $X(t)$, $t \in \mathbb{Z}$, est dit stationnaire à l'ordre r si et seulement si ses multicorrélations (corrélations cumulantes) $\mathcal{C}_{X,i_1 i_2..i_r}(\tau_2,..,\tau_r) = cum\{X_{i_1}(t), X_{i_2}(t+\tau_2), ..., X_{i_r}(t+\tau_r)\}$ sont finies et ne dépendent pas de la date t.*

Il est clair qu'un processus stationnaire (au sens strict) est stationnaire à tous les ordres jusqu'à r si ses moments sont finis jusqu'à l'ordre r.

3.3 Cas complexe, multispectres

3.3.1 Définition et circularité

Une variable aléatoire complexe, comme l'a très justment souligné Fortet [112], n'est rien d'autre qu'une variable aléatoire réelle de dimension 2. Ainsi,

une variable aléatoire complexe z admet une densité si et seulement si ses parties réelle et imaginaire admettent une densité conjointe. On pourra convenir de noter cette densité de façon compacte par $p_z(u)$, où $u \in \mathbb{C}$.

De la même façon, on peut définir la fonction caractéristique d'une variable complexe z. Si $z = x + \jmath y$, $x \in \mathbb{R}^N, y \in \mathbb{R}^N$, alors

$$\Phi_z(u) \stackrel{\text{def}}{=} E\{e^{\jmath[x^T v + y^T w]}\} = E\{e^{\jmath\, Re[z^\dagger u]}\}, \tag{3.42}$$

si $u = v + \jmath w$. Une propriété immédiate de cette notation compacte est que

$$\Phi_{aZ}(u) = \Phi_Z(a^* u), \tag{3.43}$$

pour tout scalaire complexe a. Nous avons par conséquent à notre disposition les mêmes outils que dans le cas de variables réelles. Cependant, les variables aléatoires complexes sont la plupart du temps obtenues par Transformée de Fourier (TF) de données réelles, ce qui leur confère une structure très particulière. Les variables aléatoires complexes obtenues de cette façon ne sont donc pas de simples variables aléatoires à 2 composantes réelles, mais des contraintes lient ces 2 composantes. C'est pourquoi il est pertinent d'introduire les variables aléatoires dites *circulaires*.

Définition 3.3.1 *Nous dirons qu'un vecteur aléatoire complexe de dimension N, Z, est circulaire (ou circulaire au sens fort), si et seulement si*

$$\Phi_Z(au) = \Phi_Z(u), \forall a \in \mathbb{C}, |a| = 1. \tag{3.44}$$

En particulier, si Z admet une densité, Z est circulaire si $Z\, e^{\jmath\theta}$ a même densité de probabilité que Z.

Cette définition, proposée à l'origine dans *[68]* [17], est compatible avec les définitions proposées dans le passé. En effet, elle entraine la proposition suivante:

Proposition 3.3.2 *Soit Z un vecteur aléatoire complexe, dont les moments existent à tous les ordres. Alors Z est circulaire si et seulement si tous ses moments de la forme:*

$$\mu_{pq} = E\{ \prod_{\Sigma a_i = p} Z_i^{a_i} \prod_{\Sigma b_j = q} Z_j^{*b_j} \}$$

sont nuls dès que $p \neq q$.

Démonstration. Si Z est circulaire, alors les moments de Z et de $Z\,e^{\jmath\alpha}$ sont égaux, puisque toutes deux ont même loi. En particulier, l'égalité $\mu_{pq}\{Z\} = \mu_{pq}\{Z\,e^{\jmath\alpha}\}$ entraine que

$$\mu_{pq}\{Z\} = \mu_{pq}\{Z\}\,e^{\jmath\alpha(p-q)},$$

ce qui prouve la proposition. □

La proposition 3.3.2 montre par exemple qu'une variable aléatoire scalaire complexe circulaire vérifie $E\{Z\} = 0$, $E\{Z^2\} = 0$, $E\{Z^2\,Z^*\} = 0$...

En outre, d'après la proposition 3.3.2, les fonctions caractéristiques (et la densité de probabilité quand elle existe) d'un vecteur aléatoire complexe circulaire sont fonction uniquement de la variable matricielle $u\,u^\dagger$:

$$\exists f \,/\, \Phi_Z(u) = f(uu^\dagger). \tag{3.45}$$

Cette propriété peut être comparée à la définition des variables dites *sphériquement invariantes*. D'après [174], de telles variables sont telles que:

$$\exists f \,/\, \Phi_Z(u) = f(u^\dagger C u)$$

où C est une matrice hermitienne définie positive. Autrement dit, avec ces définitions, toute variable sphériquement invariante est circulaire, mais la réciproque n'est pas vraie.

Dans la suite, nous aurons besoin de la définition restrictive suivante:

Définition 3.3.3 *On dira qu'un vecteur aléatoire complexe Z est circulaire à l'ordre r s'il verifie*

$$E\{\prod_{\Sigma a_i = p} Z_i^{a_i} \prod_{\Sigma b_j = q} Z_j^{*b_j}\} = 0 \tag{3.46}$$

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

Notons que cette définition ne suppose pas nécessairement que les moments sont finis pour $p = q$.

Dans le cas gaussien, la circularité à l'ordre 2 entraine la circularité à tous les ordres, et est caractérisée par deux propriétés liant les parties réelle et imaginaire. En effet, posons $Z = A + \jmath B$. Si Z est circulaire, alors $E\{ZZ^T\} = 0$ implique que $E\{AA^T - BB^T\} = 0$ et que $E\{AB^T + BA^T\} = 0$. Autrement dit, A et B ont même matrice de covariance, et leur covariance

croisée est antisymétrique. C'est ainsi qu'ont été définies les variables gaussiennes complexes circulaires [201] [122].

Différentes définitions possibles de circularité ont été récemment passées en revue, et analysées plus en profondeur dans [176]. On remarquera notamment que la définition de la circularité conjointe des composantes d'un vecteur aléatoire est une notion bien plus forte que la circularité marginale de chacune de ses composantes.

3.3.2 Densités multispectrales

Dans cette section, nous rappelons quelques résultats classiques de théorie du signal. Nous renvoyons aux ouvrages [85, ch. IX] [86, ch. VIII] [180, ch. I] [175, ch. 6] [90, ch. E.II.2] [160] pour les démonstrations.

Il est connu (Herghotz, Cramér) que si $X(t)$ est un processus à temps discret de dimension N faiblement stationnaire, alors il existe une fonction matricielle unique $G(\lambda)$ à accroissements non négatifs telle que:

$$G(-1/2) = 0, \text{ et } \mathcal{C}(\tau) = \int_{-1/2}^{1/2} e^{2\pi\jmath\tau\lambda}\, dG(\lambda). \tag{3.47}$$

On convient d'appeler cette fonction $G(\lambda)$ la *répartition spectrale* de puissance de $X(t)$, et $dG(\lambda)$ la *mesure spectrale* associée (d'après le théorème de Bochner, ce résultat s'applique d'ailleurs aussi aux processus à temps continu s'ils sont continus en moyenne quadratique).

De même, si $X(t)$ est stationnaire jusqu'à l'ordre r (cf définition 3.2.1), alors il existe une fonction tensorielle $G(\lambda_2, .., \lambda_r)$ telle que:

$$G(-1/2, .., -1/2) = 0, \tag{3.48}$$

$$\mathcal{C}(\tau_2, .., \tau_r) = \int_{-1/2}^{1/2} .. \int_{-1/2}^{1/2} e^{2\pi\jmath \sum_{k=2}^{r} \tau_k \lambda_k}\, dG(\lambda_2, .., \lambda_r). \tag{3.49}$$

La quantité $dG(\lambda_2, .., \lambda_r)$ est la mesure multispectrale de $X(t)$. Remarquons que cette écriture n'est autorisée que si dG est une distribution tempérée, ce qui devrait être vérifé par ailleurs. Il se trouve que ce problème n'a été curieusement jamais abordé dans la littérature. A. Blanc-Lapierre introduit par exemple une condition d'appartenance du processus à une classe baptisée $\Phi(\infty)$, qu'il est difficile de vérifier [85, ch.X].

Le signal $X(t)$ n'admet pas toujours de *densité* multispectrale d'ordre r. Une condition suffisante pour qu'il en admette une est qu'il soit *sommable à*

l'ordre r, c'est à dire que $X(t)$ soit stationnaire d'ordre r et que ses multicorrélations d'ordre r soient absolument sommables:

$$\sum_{(u_2,..,u_r)\in\mathbb{Z}^{r-1}} |\mathcal{C}_{a_1..a_r}(u_2,..,u_r)| < \infty. \tag{3.50}$$

Cette propriété assure que les multicorrélations d'ordre r tendent suffisamment vite vers zéro pour justifier l'existence de leur transformée de Fourier, les densités multispectrales $f_{a_1..a_r}(\lambda_2,..,\lambda_r)$, qui sont alors continues. Ces dernières sont alors les différentielles d'ordre r de la fonction de répartition spectrale. En outre, le processus $X(t)$ sera dit *mélangeant* à l'ordre r: plus les échantillons sont éloignés les uns des autres, plus ils sont décorrélés à l'ordre r.

Les signaux aléatoires (faiblement stationnaires) eux-mêmes admettent une représentation spectrale (représentation dite de Cramér) qui sera notée:

$$X(t) = \int_{-1/2}^{1/2} e^{2\pi\jmath t\lambda}\, dZ(\lambda), \tag{3.51}$$

où le processus $Z(\lambda)$ est un processus à accroissements orthogonaux, défini par les relations:

$$Z(\lambda) = \lim_{T\to\infty} Z(T;\lambda),\ \lambda\in[-1/2,1/2] \tag{3.52}$$

$$Z(T;\lambda) = \int_{-1/2}^{\lambda} \sum_{t=-T}^{T} X(t)e^{-2\pi\jmath ty} dy. \tag{3.53}$$

L'ensemble des définitions et propriétés essentielles que nous avons précisées jusqu'à présent vont maintenant servir à l'établissement de la circularité des variables spectrales.

3.3.3 Circularité des variables spectrales

Nous allons voir maintenant que les variables aléatoires complexes obtenues par TF de signaux aléatoires à temps discret stationnaires sont circulaires. Cependant, les signaux stationnaires n'admettent pas de transformée de Fourier, même lorsqu'ils admettent une représentation spectrale [175, ch. 6.2]. Une façon classique de contourner le problème est de raisonner sur le processus intégral $Z(\lambda)$ introduit plus haut, ou sur ses accroissements $dZ(\lambda)$.

Proposition 3.3.4 *Soient $X(t)$ un processus réel stationnaire jusqu'à l'ordre r, $r \geq 2$, et $Z(\lambda)$ sa répartition spectrale. Alors pour tout k, $1 \leq k \leq r$, et pour tout k-uplet $(\lambda_1, .., \lambda_k)$ de fréquences, $\lambda_j \in [-1/2, 1/2]$, les cumulants des accroissements spectraux s'écrivent:*

$$cum\{dZ_{a_1}(\lambda_1), .., dZ_{a_k}(\lambda_k)\} = \delta_1(\sum \lambda_j)\, dG_{a_1..a_k}(\lambda_2, .., \lambda_k), \tag{3.54}$$

où δ_1 désigne la distribution "peigne de Dirac" de période 1, et dG la mesure multispectrale de $X(t)$.

Démonstration. Par définition de $Z(\lambda)$, il vient que:

$$\alpha \stackrel{\text{def}}{=} cum\{Z_{a_1}(\lambda_1), .., Z_{a_r}(\lambda_r)\} \tag{3.55}$$

$$= \sum_{t_1} .. \sum_{t_r} \int_{-1/2}^{\lambda_1} .. \int_{-1/2}^{\lambda_r} \mathcal{C}_{12..r}(t_2 - t_1, .., t_r - t_1)$$

$$e^{-2\pi\jmath \sum_{k=1}^{r} y_k t_k}\, dy_1..dy_r. \tag{3.56}$$

Puisque $X(t)$ est stationnaire à l'ordre r, on peut d'après (3.49) exprimer ses multicorrélations en fonction de ses mesures multispectrales correspondantes. D'où, en posant $t_1 = t$:

$$\alpha = \sum_{t_2..t_r} \int_{-1/2}^{\lambda_1} .. \int_{-1/2}^{\lambda_r} \int .. \int \sum_{t} exp\{-2\pi\jmath t(y_1 + \Sigma\, u_k)\}$$

$$exp\{2\pi\jmath \sum_{k=2}^{r} (u_k - y_k)t_k\}\, dG(u_2, .., u_r) dy_1..dy_r. \tag{3.57}$$

Or, la somme sur $t \in \mathbb{Z}$ de $e^{2\pi\jmath t\beta}$ est égale à la distribution tempérée *peigne de Dirac* de période 1, noté ici $\delta_1(\beta)$. Donc les $r-1$ premières sommes dans (3.57) valent $\delta_1(u_k - y_k)$. Par conséquent, $u_k = y_k$ et le cumulant calculé devient:

$$\alpha = \int_{-1/2}^{\lambda_1} .. \int_{-1/2}^{\lambda_r} dG(y_2, .., y_r)\, \delta_1(\sum_{k=1}^{r} y_k). \tag{3.58}$$

La proposition s'obtient alors par différentiation, grâce à la multilinéarité des cumulants [160]. □

En corollaire, on peut établir la propriété de circularité suivante:

Proposition 3.3.5 *Si en outre $X(t)$ est sommable à tous les ordres jusqu'à r, alors pour toute fréquence λ telle que $|\lambda| < \frac{1}{r}$, les vecteurs $dZ(\lambda)$ sont circulaires à l'ordre $(p+q) = r$. Autrement dit:*

$$E\{dZ_{n1}(\lambda)..dZ_{n_p}(\lambda)\, dZ^*_{m_1}(\lambda)..dZ^*_{m_q}(\lambda)\} = 0$$

dès que $p \neq q$, $1 \leq p, q \leq r$.

Notons que pour les processus à temps continu, la circularité décrite ci-dessus serait toujours assurée, pourvu que les écritures (3.47) et (3.49) soient autorisées (par exemple, lorsque les mesures multispectrales dG sont absolument sommables). On peut le vérifier en constatant que si la fréquence d'échantillonage tend vers l'infini, alors la condition sur la fréquence réduite $|\lambda| < \frac{1}{r}$ tend à être toujours vraie pour toute valeur λ finie.

Démonstration. Soit s un entier quelconque, $s \in \{1, 2, .., r\}$, et p et q deux entiers positifs tels que $p+q = s$. Appliquons la proposition 3.3.4 avec $\lambda_1 = \lambda_2 = .. = \lambda_p = \lambda$ et $\lambda_{p+1} = \lambda_{p+2} = .. = \lambda_{p+q} = -\lambda$. La somme des fréquences vaut $\sum \lambda_i = (p-q)\lambda$. Si $p \neq q$, alors une condition suffisante pour que $(p-q)\lambda$ ne soit jamais entier est que $0 < (p+q)|\lambda| < 1$. Le terme $\delta_1(\sum \lambda_i)$ est donc toujours nul sous les hypothèses de la présente proposition. Comme $X(t)$ est sommable à tous les ordres jusqu'à $r = (p+q)$, il admet une densité multispectrale d'ordre s définie par:

$$dG_{a_1 a_2 .. a_s}(\lambda_2, .., \lambda_s) = f_{a_1 a_2 .. a_s}(\lambda_2, .., \lambda_s)\, d\lambda_2 .. d\lambda_s,$$

où $f_{a_1 a_2 .. a_s}$ est finie. D'après la proposition 3.3.4, tous les cumulants de $dZ(\lambda)$ d'ordre s sont donc nuls, pour tous les ordres s inférieurs ou égaux à r. Comme les moments sont fonctions polynômiales des cumulants, ils sont par conséquent aussi tous nuls. □

Nous renvoyons le lecteur à l'article récent de B. Picinbono [176] pour une discussion plus complète, et en particulier sur les conditions de circularité conjointe.

3.4 Propriétés des moments et cumulants

Les SOS jouissent tout d'abord de deux propríetés élémentaires que nous exposons maintenant; la seconde n'est satisfaite que par les cumulants.

Proposition 3.4.1 *Les moments et cumulants satisfont la propriété dite de* multilinéarité. *Soient deux vecteurs aléatoires x et y liés par la relation*

linéaire $y = Ax$, où A est une matrice quelconque. Alors les moments et cumulants de y sont des fonctions formellement linéaires de chacune des composantes A_{ij}. Par exemple nous aurons:

$$\mathcal{C}\{y_i, y_j\} = \sum_{a,b} A_{ia} A_{jb}\, \mathcal{C}\{x_a, x_b\}, \tag{3.59}$$

$$\mathcal{C}\{y_i, y_j, y_k\} = \sum_{a,b,c} A_{ia} A_{jb} A_{kc}\, \mathcal{C}\{x_a, x_b, x_c\}, \tag{3.60}$$

$$\mu\{y_i, y_j, y_k\} = \sum_{a,b,c} A_{ia} A_{jb} A_{kc}\, \mu\{x_a, x_b, x_c\}, \tag{3.61}$$

$$\mathcal{C}_{(3)}\{y_i\} = \sum_{a,b,c} A_{ia} A_{ib} A_{ic}\, \mathcal{C}\{x_a, x_b, x_c\} \ldots \tag{3.62}$$

Démonstration. Il suffit de remarquer que $\Phi_{Ax}(u) = \Phi_x(A^\dagger u)$, d'après (3.42). En passant à la variable aléatoire réelle de taille double, on peut alors obtenir le résultat à l'aide de (3.31). □

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

$$\mathcal{C}_{(r)}\{\lambda\, x\} = \lambda^r\, \mathcal{C}_{(r)}\{x\}. \tag{3.63}$$

Proposition 3.4.2 *Les cumulants satisfont la propriété d'* additivité *suivante. Si x et y sont des vecteurs aléatoires indépendants, alors:*

$$\mathcal{C}\{x + y\} = \mathcal{C}\{x\} + \mathcal{C}\{y\}. \tag{3.64}$$

Démonstration. SI x et y sont indépendantes, alors $p_{x,y}(u, v) = p_x(u)\, p_y(v)$, d'où $\Phi_{x,y}(u, v) = \Phi_x(u)\, \Phi_y(v)$, et finalement $\Psi_{x,y}(u, v) = \Psi_x(u) + \Psi_y(v)$. Ceci prouve la proposition pour les variables réelles. Les variables complexes de dimension N peuvent être traitées comme des variables aléatoires de dimension $2N$. □

Nous avons défini dans la section 3.1 l'opération de *standardisation* pour les variables aléatoires scalaires. Cette opération peut aussi être définie dans le cas multivariable. Soit x un vecteur aléatoire de matrice de covariance C_{ij}. Si la matrice C est inversible, alors la variable standardisée est définie comme étant $\tilde{x} = R^{-1}x$, où R est une matrice telle que $RR^\dagger = C$. Noter que, la matrice R n'étant pas unique, la variable standardisée n'est pas unique,

bien qu'ayant une covariance unité: $\mathcal{C}\{\tilde{x}\} = R^{-1}CR^{-\dagger} = I$. On convient donc de choisir un procédé systématique pour calculer R, qui aura en outre le mérite de fonctionner même lorsque C ne sera pas inversible.

Définition 3.4.3 *Soient x un vecteur aléatoire de dimension N, C sa matrice de covariance, et $C = RS^2R^\dagger$ la décomposition en éléments propres correspondante, où S est une matrice diagonale $r \times r$ à éléments strictement positifs, $r \leq N$, et R une matrice vérifiant $R^\dagger R = I$. Le vecteur $\tilde{x} = S^{-1}R^\dagger x$ est le vecteur standardisé associé à x.*

Le vecteur $\tilde{x}$ est maintenant défini en général (càd si toutes ses valeurs propres non nulles de C sont distinctes) à une matrice multiplicative près de la forme ΔP, où Δ est diagonale $r \times r$ et constituée d'éléments de module 1, et P est une permutation. Le vecteur aléatoire $\tilde{x}$ a toujours une covariance unité.

Les moments et cumulants satisfont un certain nombre d'inégalités remarquables qu'il est difficile de répertorier de façon exhaustive. La proposition ci-dessous en donne quelques-unes.

Proposition 3.4.4 *Soit X un vecteur aléatoire réel d'ordre 4 et de dimension N. Alors ses cumulants standardisés $\gamma_{ijkl} = \mathcal{K}_{ijkl}\{X\}$ satisfont les relations suivantes:*

$$\gamma_{iiii} \geq -2, \tag{3.65}$$

$$\gamma_{iijj} \geq -1, \tag{3.66}$$

$$\gamma_{iiii} + 2\gamma_{iiij} \geq -2, \tag{3.67}$$

$$\gamma_{iij}^2 \leq \gamma_{iiii} + 2, \tag{3.68}$$

$$\gamma_{iij}^2 + \gamma_{ijj}^2 \leq \gamma_{iijj} + 1. \tag{3.69}$$

Démonstration. Si μ_{ijkl} désignent les moments centrés de X, alors les cumulants standardisés satisfont

$$\gamma_{iiii} = \mu_{iiii} - 3, \qquad \gamma_{iijj} = \mu_{iijj} - 1 \qquad \text{si} \qquad i \neq j$$

et $\gamma_{iiij} = \mu_{iiij}$ si $i \neq j$. D'après l'inégalité de Cauchy-Schwarz, toute variance est positive, et en particulier $var\{X_iX_j + aX_i + bX_j\} \geq 0$ quels que soient les paramètres a et b. Si nous calculons cette variance en fonction des moments centrés de X, nous obtenons donc un polynome de degré 2 en a et en b. Pour a fixé, son discriminant est donc négatif, ce qui conduit à conclure

qu'un polynome en b est à son tour positif. La négativité de son discriminant conduit finalement à la relation $\gamma_{iij}^2 + \gamma_{ijj}^2 \leq \gamma_{iijj} + 1$. De même, en étudiant le signe de $var\{aX_i^2 + bX_j\}$, on obtiendrait la relation $\gamma_{iij}^2 \leq \gamma_{iiii} + 2$.

Pour terminer, en étudiant le signe du polynome $var\{aX_i^2 + bX_j^2 + X_iX_j\}$, pour $(a, b) \in I\!R \cup \{-\infty, \infty\}$, on obtient l'inégalité suivante

$$a^2(\gamma_{1111} + 2) + b^2(\gamma_{2222} + 2) + (1 + 2ab)\gamma_{1122} + 2a\gamma_{1112} + 2b\gamma_{1222} + 1 \geq 0.$$

dont les trois premières relations de la proposition sont des cas particuliers. □

Exemple: Si $p_z(u) = \frac{1}{2}\delta(u-1) + \frac{1}{2}\delta(u+1)$, alors $\mu_{2r} = 1$ et $\mu_{2r+1} = 0$. Donc $\gamma_{(4)} = -2$ et la borne est atteinte.

Théorème de la limite centrale par les SOE

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

Proposition 3.4.5 *Soient $X(n), 1 \leq n \leq N$, N variables aléatoires scalaires indépendantes, chacune de cumuant d'ordre r borné, noté $\kappa_r(n)$. On pose*

$$\bar{\kappa}_r = \frac{1}{N}\sum_{n=1}^{N} \kappa_r(n) \text{ et } Y = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}(X(n) - \bar{\kappa}_1).$$

Alors la variable aléatoire Y tend en loi vers une variable aléatoire gaussienne. Plus précisément, ses cumulants d'ordre r, notés λ_r, sont donnés par:

$$\lambda_1 = 0, \tag{3.70}$$

$$\lambda_2 = \bar{\kappa}_2, \tag{3.71}$$

$$\lambda_r = \frac{1}{N^{r/2-1}}\bar{\kappa}_r, \ \forall r \geq 2. \tag{3.72}$$

Démonstration. En vertu de la propriété d'additivité (proposition 3.4.2), les cumulants de la variable Y s'écrivent comme la somme:

$$\mathcal{C}_{(r)}\{Y\} = N^{-r/2}\sum_{n=1}^{N}\mathcal{C}_{(r)}\{X(n)\},$$

ce qui prouve que $\lambda_r = N^{1-r/2}\bar{\kappa}_r$ par définition de $\bar{\kappa}_r$. □

3.4.1 Liens entre SOE et densité de probabilité

a) Problème des moments

La première fonction caractéristique est la transformée de Fourier de la densité de probabilité, éventuellement au sens des distributions si cette dernière n'existe pas. Ceci suffit à montrer, au moins intuitivement, que les contraintes qui vont lier les moments sont très compliquées, et qu'elles ne peuvent certainement pas être décrites simplement comme on l'a exposé dans la section 3.4, à l'aide d'inégalités de Schwarz. Ces contraintes découlent essentiellement de la positivité de la mesure. Déjà, nous avons vu que l'existence des moments n'est pas liée à celle de la densité, mais à la différentiabilité de la fonction caractéristique au voisinage de l'origine (on sait déjà qu'elle est continue partout). Plus précisément, si une fonction caractéristique admet une différentielle d'ordre r à l'origine, alors tous les moments jusqu'à l'ordre r existent si r est pair, mais jusqu'à l'ordre $r-1$ si r est impair [152, p. 29].

Le premier théorème, dû à Marcinkiewicz (1940), dont Dugué donna une démonstration plus simple en 1951, établit que si une variable aléatoire possède un cumulant non nul d'ordre $r > 2$, alors elle en possède une infinité *[2]* [178] [134] [152]:

Théorème 3.4.6 *Si une fonction caractéristique est de la forme* $\Phi(u) = exp\{P(u)\}$*, où* $P(u)$ *est un polynôme, alors ce polynôme est de degré au plus 2.*

En d'autres termes, la variable aléatoire est soit réduite à une constante, soit gaussienne.

Le problème des moments est le problème inverse, en quelque sorte. Etant donnée une suite de nombres, existe-il une densité qui les admette pour cumulants (ou moments) ?

Sous certaines conditions, la suite infinie des moments peut définir la fonction caractéristique de façon unique [134]. Mais on peut trouver des exemples de densités ayant la même suite infinie de moments [152, p. 20]. Lorsqu'une suite finie de moments est donnée, s'il existe une solution, il y en a en général plusieurs. On peut alors sélectionner une solution en maximisant par exemple l'information de Fisher [204].

Notons que le problème des moments a également été étudié à l'ordre 2, pour les processus stationnaires [146]. Il s'agit alors de compléter une suite de valeurs de la fonction de corrélation, connue en un nombre fini de valeurs.

Connaissant les 4 premiers cumulants, il peut être intéressant sur le plan pratique de connaitre une loi réaliste pouvant approximer celle des observations. Le système de lois de Pearson permet de répondre à cette préoccupation; en effet, il effectue une partition de l'ensemble des densités, dont les 4 premiers moments sont finis, en différentes familles, suivant les valeurs du couple asymétrie-kurtosis [138, vol.6, p.655–657] [138, vol.3, p.216–219] [131, ch. 12]. Evidemment, le choix de cette solution a principalement un intérêt pratique, mais n'est pas justifié par un critère d'optimalité, contrairement à l'approche du problème des moments.

b) Queues de distribution

Une idée fausse consiste à croire qu'une densité ayant des queues de distribution en-dessous de la gaussienne aura nécessairement un kurtosis négatif, et un kurtosis positif dans le cas contraire. En outre, la définition des lois sous- et sur-gaussiennes est très versatile, suivant les articles techniques, comme nous l'expliquons maintenant.

Benveniste propose notamment [83, page 390] une définition faisant intervenir la monoticité de

$$f(u) = -\frac{1}{u}\,\frac{d\,\log p_x(u)}{du}.$$

Lorsque $f(u)$ est strictement croissante (resp. décroissante), $p_x(u)$ est dite sur-gaussienne (resp. sous-gaussienne). Il est clair que certaines densités ne seront ni l'une, ni l'autre.

En revanche, de nombreux auteurs qualifient de sur-gaussiennes les densités ayant des queues de distribution supérieures à la densité gaussienne à l'infini [204], et de sous-gaussiennes les autres. En réalité, A. Mansour a montré par une simple application d'un théorème de la moyenne que cette dernière définition est équivalente au signe du kurtosis (négatif pour les densités sous-gaussiennes) si la partie paire de la densité coupe deux fois (càd une fois sur $[0 + \infty[$) la densité gaussienne de mêmes moyenne et variance. En revanche, des contre-exemples des deux types ont été donnés lorsque le nombre d'intersections est différent de deux.

On retiendra donc qu'il existe au moins trois définitions du caractère de sous- ou sur- gaussianité, et que ces dernières ne sont pas toujours équivalentes.

3.5 Estimation des moments et cumulants

3.5.1 Les κ-statistiques

Si X est une variable aléatoire scalaire, et si $x(n), 1 \leq n \leq N$, sont N réalisations de X identiquement distribuées, il est naturel d'estimer sa moyenne statistique de X par la moyenne arithmétique de ses réalisations:

$$k_{(1)} = \frac{1}{N} \sum_{n=1}^{N} x(n). \tag{3.73}$$

Il est facile de vérifier que $k_{(1)}$ est un estimateur non biaisé de $\mu'_{(1)}$. On pourrait être tenté de poursuivre aux ordres supérieurs en utilisant les moyennes empiriques suivantes

$$m_{(r)} = \frac{1}{N} \sum_{n=1}^{N} (x(n) - k_{(1)})^r, \tag{3.74}$$

mais il s'avère que ces estimateurs sont en général biaisés. En effet, nous avons par exemple, si les réalisations $x(n)$ sont indépendantes:

$$E\{m_{(2)}\} = \frac{N-1}{N} \mu_{(2)}.$$

Un estimateur non biaisé de la variance de X est donc:

$$k_{(2)} = \frac{N}{N-1} m_{(2)}. \tag{3.75}$$

Ce procédé peut être poursuivi aux ordres supérieurs à 2 en cherchant les coefficients $\alpha_{i,r}$ tels que l'expression

$$k_{(r)} = \sum_{i=1}^{r} \alpha_{i,r} \prod_{\Sigma q_i = r} m_{(q_i)} \tag{3.76}$$

soit un estimateur non biaisé de $\mathcal{C}_{(r)}\{X\}$. Ainsi on trouverait:

$$k_{(3)} = \frac{N^2}{(N-1)(N-2)} m_{(3)} \tag{3.77}$$

$$k_{(4)} = \frac{N^2}{(N-1)(N-2)(N-3)} [\,(N+1)\, m_{(4)} - 3\,(N-1)\, m_{(2)}^2\,] \tag{3.78}$$

Les quantités définies de cette façon sont communément appelées $\kappa-$ *statistiques* [136]. En ce qui concerne les cumulants standardisés, il n'existe pas d'estimateur non biaisé qui soit indépendant de la distribution de X. L'asymétrie et l'aplatissement, qui sont essentiellement les seules grandeurs standardisées qui retiendront notre intérêt, seront estimées par les grandeurs biaisées suivantes:

$$\mathcal{K}_{(3)}\{X\}: \; g_{(3)} = k_{(3)} \,/\, k_{(2)}^{3/2}, \tag{3.79}$$

$$\mathcal{K}_{(4)}\{X\}: \; g_{(4)} = k_{(4)} \,/\, k_{(2)}^{2}. \tag{3.80}$$

3.5.2 Premiers cumulants des κ-statistiques

Les cumulants successifs des estimateurs $k_{(r)}$ sont maintenant bien connus, et leurs moments et cumulants successifs peuvent être calculés de façon exacte [136] § 12.16. Pour alléger les écritures, notons les cumulants $\kappa_{(r)} = \mathcal{C}_{(r)}\{X\}$, les moments standardisés $\beta_{(r)} = \mu_{(r)}/\mu_{(2)}^{r/2}$, et les cumulants standardisés $\gamma_{(r)} = \mathcal{K}_{(r)}\{X\}$. On notera en particulier que (tooujours sous l'hypothèse que les échantillons sont i.i.d.):

$$\mu_{(2)}\{k_{(2)}\} = \frac{\kappa_4}{N} + \frac{2\,\kappa_2^2}{N-1}, \tag{3.81}$$

$$\mu_{(2)}\{k_{(3)}\} = \frac{\kappa_6}{N} + \frac{9\,(\kappa_4\kappa_2 + \kappa_3^2)}{N-1} + \frac{6N\,\kappa_2^3}{(N-1)(N-2)}, \tag{3.82}$$

$$\mu_{(2)}\{k_{(4)}\} = \frac{\kappa_8}{N} + \frac{(16\,\kappa_6\kappa_2 + 48\,\kappa_5\kappa_3 + 34\,\kappa_4^2)}{N-1} + \frac{8N(9\,\kappa_4\kappa_2^2 + 18\,\kappa_3^2\kappa_2)}{(N-1)(N-2)} + \frac{24N(N+1)\,\kappa_2^4}{(N-1)(N-2)(N-3)}; \tag{3.83}$$

et que, pour de grandes valeurs de N:

$$\mathcal{C}_{(3)}\{k_{(2)}\} \approx \frac{1}{N^2}\,[\kappa_6 + 12\,\kappa_4\kappa_2 + 4\,\kappa_3^2 + 8\,\kappa_2^3], \tag{3.84}$$

$$\mathcal{C}_{(3)}\{k_{(3)}\} \approx \frac{1}{N^2}\,[\kappa_9 + 27\,(\kappa_7\kappa_2 + 3\,\kappa_6\kappa_3 + 4\,\kappa_5\kappa_4) + 18\,(12\,\kappa_5\kappa_2^2 + 45\,\kappa_4\kappa_3\kappa_2 + 14\,\kappa_3^3 + 30\,\kappa_3\kappa_2^3)\,], \tag{3.85}$$

$$\mathcal{C}_{(4)}\{k_{(2)}\} \approx \frac{1}{N^3}[\kappa_8 + 24\,\kappa_6\kappa_2 + 32\,\kappa_5\kappa_3 + 32\,\kappa_4^2 + 144\,\kappa_4\kappa_2^2 + 96\,\kappa_3^2\kappa_2 + 48\,\kappa_2^4]. \tag{3.86}$$

En raison de leur longueur, les expressions des cumulants de $k_{(4)}$ ne sont pas rapportées ici. Plus généralement, nous avons:

$$\mathcal{C}_{(q)}\{k_{(r)}\} = O(\frac{1}{N^{q-1}}). \tag{3.87}$$

Les estimateurs $k_{(r)}$ sont par conséquent asymptotiquement gaussiens. Mais si l'approximation gaussienne est assez vite valable pour $\kappa_{(2)}$, il faudra vraisemblablement atteindre des valeurs nettement plus grandes de N pour qu'elle soit valable pour $\kappa_{(3)}$ et a fortiori $\kappa_{(4)}$; pour s'en assurer, il suffit de consulter [136] pour constater que le coefficient du terme en $1/N^{q-1}$ est généralement de plus en plus grand lorsque r augmente. Pourtant, certaines distributions échappent à cette règle heuristique [87].

Par ailleurs, les propriétés statistiques des estimateurs standardisés $g_{(r)}$ ont été étudiées seulement de façon approchée pour de grandes valeurs de N [160] [138], en raison de leur complexité. On notera en particulier qu'ils sont biaisés au premier ordre, et qu'ils sont corrélés (leur biais dépend d'ailleurs des cumulants d'ordre plus élevé) [136, ex.10.26-27]. En effectuamt un développement limité de la fonction de deux variables $w(x, y) = x/y^{3/2}$, il est possible d'obtenir les expressions approchées (3.91) et (3.92):

$$\begin{aligned}
E\{g_{(3)}\} &= \beta_{(3)} + O(\frac{1}{N}), && (3.88)\\
E\{g_{(4)}\} &= \beta_{(4)} + O(\frac{1}{N}), && (3.89)\\
\mu_{(2)}\{g_{(3)}\} &= \frac{1}{N}\left[\beta_{(6)} - 6\,\beta_{(4)} + 9 + \frac{1}{4}\,\beta_{(3)}\,(9\,\beta_{(4)} + 35) - 3\,\beta_{(5)}\beta_{(3)}\right] \\
&\quad + O(\frac{1}{N^2}), && (3.90)\\
\mu_{(2)}\{g_{(3)}\} &= \frac{1}{N}\left[\gamma_{(6)} - 3\,\gamma_{(3)}\gamma_{(5)} + 9\,\gamma_{(4)}\,(1 + \frac{1}{4}\,\gamma_{(3)}^2) - \frac{11}{2}\,\gamma_{(3)}^2 + 6\right] \\
&\quad + O(\frac{1}{N^2}), && (3.91)\\
\mu_{(2)}\{g_{(4)}\} &= \frac{1}{N}\left[\beta_{(8)} - 4\,\beta_{(6)}\beta_{(4)} + 4\,\beta_{(4)}^3 - \beta_{(4)}^2 + 16\,\beta_{(4)}\beta_{(3)}\right. \\
&\quad \left. -8\,\beta_{(5)}\beta_{(3)} + 16\,\beta_{(3)}\right] + O(\frac{1}{N^2}). && (3.92)
\end{aligned}$$

Exemple: Si $X(n)$ sont des variables indépendantes uniformément distribuées dans $[-aa]$, alors pour de grandes valeurs de N, la variance relative

du moment empirique d'ordre r vaut pour r pair:

$$\frac{Var\{k_{(r)}\}}{\mu_{(r)}^2} = \frac{1}{N}\frac{r^2}{2r+1}.$$

3.5.3 Statistiques dans le cas gaussien

Dans le cas gaussien, un certain nombre de simplifications sont possibles car tous les cumulants d'ordre supérieur à deux apparaissant dans les expressions générales sont nuls. De plus, $k_{(2)}$ et $k_{(r)}/k_{(2)}^{r/2}$ sont indépendantes. On obtient notamment [136, §12.16, 12.18]:

$$\mu_{(2)}\{k_{(2)}\} = \frac{2}{N-1}\kappa_2^2, \tag{3.93}$$

$$\mu_{(2)}\{k_{(3)}\} = \frac{6N}{(N-1)(N-2)}\kappa_2^3, \tag{3.94}$$

$$\mu_{(2)}\{k_{(4)}\} = \frac{24N(N+1)}{(N-1)(N-2)(N-3)}\kappa_2^4. \tag{3.95}$$

Cette dernière relation montre par exemple que la variance du cumulant d'ordre 4 est en $O(\frac{24}{N})$. Dans le cas complexe circulaire, on trouverait $O(\frac{4}{N})$.

En ce qui concerne les estimateurs de l'asymétrie et de l'aplatissement, nous avons dans le cas gaussien des résultats *exacts* [136] ex. 12.9, 12.10, et 12.22, [160] p 108-109, [157]:

$$\begin{aligned} E\{g_{(3)}\} &= \frac{E\{k_3\}}{E\{k_2^{3/2}\}} \\ &= \gamma_{(3)} = 0, \end{aligned} \tag{3.96}$$

$$\begin{aligned} E\{g_{(4)}\} &= \frac{E\{k_4\}}{E\{k_2^2\}} \\ &= \gamma_{(4)} = 0, \end{aligned} \tag{3.97}$$

$$\begin{aligned} \mu_{(2)}\{g_{(3)}\} &= \frac{6N(N-1)}{(N-2)(N+1)(N+3)} \\ &\approx \frac{6}{N} + O\{\frac{1}{N^2}\}, \end{aligned} \tag{3.98}$$

$$\begin{aligned} \mu_{(2)}\{g_{(4)}\} &= \frac{24N(N-1)^2}{(N-3)(N-2)(N+3)(N+5)} \\ &\approx \frac{24}{N} + O\{\frac{1}{N^2}\}. \end{aligned} \tag{3.99}$$

Autrement dit, la variance du kurtosis (aplatissement) est du même ordre que celle du cumulant d'odre 4 non standardisé.

On vérifie que les variances (3.98) et (3.99) peuvent être obtenues en annulant les cumulants standardisés dans les expressions (3.91) et (3.92). Il est aussi possible de calculer les cumulants standardisés des estimateurs standardisés $g_{(r)}$. En effet, on peut déduire de [157] que:

$$\mathcal{K}_{(3)}\{g_{(3)}\} = 0, \tag{3.100}$$

$$\mathcal{K}_{(3)}\{g_{(4)}\} = \sqrt{\frac{216}{N}} - \frac{213}{N\sqrt{N}} + o\{\frac{1}{N^2}\}, \tag{3.101}$$

$$\mathcal{K}_{(4)}\{g_{(3)}\} = \frac{36}{N} - \frac{864}{N^2} + o\{\frac{1}{N^2}\}, \tag{3.102}$$

$$\mathcal{K}_{(4)}\{g_{(4)}\} = \frac{540}{N} - \frac{20\,196}{N^2} + o\{\frac{1}{N^2}\}. \tag{3.103}$$

Il est clair que pour N de l'ordre de 300 ou plus, l'aplatissement de $g_{(3)}$ devient négligeable; en revanche, il faut atteindre des valeurs de $N > 5000$ pour avoir une approximation gaussienne acceptable pour $g_{(4)}$. La distribution exacte de $g_{(3)}$ et $g_{(4)}$ a été tabulée par Pearson et Hartley dans les années 70. Par ailleurs, D'Agostino et Pearson [102] ainsi que Anscombe et Glynn [78] donnent ces distributions pour des valeurs de N inférieures à 200.

Exemple: Si $X(n)$ sont des variables gaussiennes indépendantes, toutes de moyenne nulle et de variance σ^2, alors le moment centré d'ordre $2r$ est donné par:

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

et la variance relative du moment empirique correspondant est, pour de grandes valeurs de N:

$$\frac{Var\{k_{(2r)}\}}{\mu^2_{(2r)}} = \frac{1}{N}\left[\frac{4r!\, r!^2}{2r!^3} - 1\right].$$

Un estimateur du cumulant d'ordre 4 a été proposé récemment par Amblard [76] , et n'a pas recours de manière explicite aux moments d'ordre 4, ce qui peut être avantageux dans une implantation récursive. Cet estimateur s'écrit, à la date t, si on dispose de l'estimation à la date précédente et des observations $x(t)$ et $x(t-1)$:

$$k_{(4),t} = (1-\alpha)\, k_{(4),t-1} + \alpha\left(x(t)^4 - 3x(t)^2 x(t-1)^2\right). \tag{3.104}$$

3.5.4 Cas multivariable

Dans le cas de variables à plusieurs composantes, le principe mis en œuvre est le même, bien que les notations soient nettement plus compliquées. Il faut notamment faire appel à la convention de sommation d'Einstein, et à la notation crochet de McCullagh. C'est pourquoi on se contente d'expressions asymptotiques (pour de larges valeurs de N). Toutefois, il existe aussi des mécanismes dans le cas multivariable pour générer les κ-statistiques [160].

On citera, simplement à titre d'exemple, le cumulant d'ordre 3 de la covariance estimée:

$$\begin{aligned}\mathcal{C}_{(3)}\{\kappa_{ij},\kappa_{kl},\kappa_{mn}\} = \frac{1}{N^2}\,\kappa_{ijklmn} + [12]\,\frac{1}{N(N-1)}\,\kappa_{ijkl}\,\kappa_{mn} \\ +[4]\,\frac{N-2}{N(N-1)^2}\,\kappa_{ijk}\,\kappa_{lmn} + [8]\,\frac{1}{(N-1)^2}\,\kappa_{ij}\,\kappa_{kl}\,\kappa_{mn}. \qquad (3.105)\end{aligned}$$

3.5.5 Fonctions de multicorrélation

On définit habituellement l'estimation suivante de la fonction d'autocorrélation (d'ordre 2) d'un processus $x(t)$ scalaire stationnaire au sens large:

$$\hat{\mathcal{C}}_{(2),x}(\tau) = \frac{1}{N}\sum_{t=1}^{N} x(t)\,x(t+\tau). \qquad (3.106)$$

Notons que pour N fini, cet estimateur n'utilise pas également toutes les données. En pratique, on a aussi le choix entre deux autres estimateurs; le premier est biaisé mais de type positif, et le deuxième non biaisé:

$$\hat{\mathcal{C}}_{(2),x}(\tau) = \frac{1}{N}\sum_{t=1}^{N-\tau} x(t)\,x(t+\tau), \qquad (3.107)$$

$$\hat{\mathcal{C}}_{(2),x}(\tau) = \frac{1}{N-\tau}\sum_{t=1}^{N-\tau} x(t)\,x(t+\tau). \qquad (3.108)$$

Lorsque $\tau \ll N$, les trois estimateurs sont équivalents.

La convergence (et la consistence) de ces estimateurs est une question importante, qui est liée aux propriétés d'ergodicité du processus [137, sec. 47.7] [89, p. 41-43] [86, ch.15]. En appliquant le théorème ergodique à la série chronologique $x(t)x(t+\tau)$ à τ fixé, on pourrait étudier la convergence p.s. (presque sûre) de l'estimateur de la fonction de corrélation. Cependant, il peut être utile de se contenter d'une stationnarité plus faible, d'une part,

et de chercher à obtenir une consistence en moyenne quadratique (m.q.), d'autre part. Rappelons que les convergences m.q. et p.s. ne sont pas toujours comparables.

Nous savons qu'à l'ordre 2, la variance de la fonction d'autocorrélation (3.106) est donnée par:

$$\begin{aligned} Var\{\hat{\mathcal{C}}_{(2),x}(\tau)\} &= \frac{1}{N^2}\sum_{s=0}^{N-1}(N-s)\,\mathcal{C}_{(4)x}(s,\tau,s+\tau)+(N-s)\,\mathcal{C}^2_{(2)x}(s) \\ &\quad -s\,\mathcal{C}^2_{(2)x}(\tau)+(N-s)\,\mathcal{C}_{(2)x}(s+\tau)\,\mathcal{C}_{(2)x}(s-\tau). \quad (3.109) \end{aligned}$$

Habituellement, on adopte d'ailleurs plutôt l'expression suivante, valable si les $\mathcal{C}_{(r),x}$ décroissent assez vite vers zéro:

$$Var\{\hat{\mathcal{C}}_{(2),x}(\tau)\} \approx \frac{1}{N}\sum_{s=0}^{N-1}\mathcal{C}_{(4)x}(s,\tau,s+\tau)+\mathcal{C}^2_{(2)x}(s)+\mathcal{C}_{(2)x}(s+\tau)\,\mathcal{C}_{(2)x}(s-\tau). \quad (3.110)$$

Pour que l'estimateur $\hat{\mathcal{C}}_{(2),x}(\tau)$ converge vers $\mathcal{C}_{(2),x}(\tau)$ en moyenne quadratique (consistence forte), il suffit que [123]:

1. $x(t)$ admette des moments finis jusqu'à l'ordre 4,
2. $x(t)$ soit stationnaire jusqu'à l'ordre 4,
3. $\frac{1}{N}\sum_{u=1}^{N}\mathcal{C}_{(2),x}(u)^2 \to 0$ si $N \to \infty$
4. $\frac{1}{N}\sum_{u=1}^{N}\mathcal{C}_{(4),x}(u,\tau,u+\tau) \to 0$ si $N \to \infty$

Notamment, il est suffisant que:

$$\sum_{u=1}^{N}\mathcal{C}_{(4),x}(u,\tau,u+\tau) \quad et \quad \sum_{u=1}^{N}\mathcal{C}_{(2),x}(u)^2 \quad (3.111)$$

soient bornées quand N tend vers l'infini. Le processus $X(t)$ doit donc être mélangeant dans un sens voisin de celui défini en (3.50), si ces conditions suffisantes sont adoptées.

Ceci s'étend sans mal au cas multivariable, en remplaçant les sommes précédentes par:

$$\sum_{u=1}^{N}\mathcal{C}_{x,ijij}(u,\tau,u+\tau) \quad et \quad \sum_{u=1}^{N}\mathrm{trace}\{\mathcal{C}_x(u)\,\mathcal{C}_x(u)^T\}. \quad (3.112)$$

Des résultats similaires existent également pour les fonctions de corrélation normalisées [179, p. 76] [137, ch.48]. Dans ce dernier cas, il est plus difficile de construire un estimateur non biaisé, car le biais dépend alors de la distribution.

Lorsque le processus $X(t)$ est fortement mélangeant (*i.e.* la dépendance entre le passé avant la date $t = a$ et le futur après la date $t = b$ tend vers zéro lorsque $b - a$ augmente), alors on peut montrer que les estimateurs de $\mathcal{C}_{(2)x}(\tau)$ sont asymptotiquement conjointement gaussiens [180, p. 117]. C'est le cas des processus linéaires.

Pour les processus stationnaires à l'ordre $r > 2$, les fonctions de multicorrélation à l'ordre r se définissent comme à l'ordre 2, si on admet que la durée d'intégration, T, est grande devant l'unité. Ainsi, la fonction de multicorrélation cumulante d'ordre 3 d'un signal stationnaire jusqu'à l'ordre 3, $x(t)$, peut être estimée par:

$$\hat{\mathcal{C}}_{(3),x}(\tau_1, \tau_2) = \frac{1}{T} \sum_{t=1}^{T} x(t)\, x(t+\tau_1)\, x(t+\tau_2)). \tag{3.113}$$

Dans le cas multivariable, l'expression est similaire:

$$\hat{\mathcal{C}}_{X,ijk}(\tau_1, \tau_2) = \frac{1}{T} \sum_{t=1}^{T} X_i(t)\, X_j(t+\tau_1)\, X_k(t+\tau_2)). \tag{3.114}$$

mais on devra faire appel au produit de Kronecker si on souhaite garder une formulation compacte non indexée.

Des conditions suffisantes de consistence en m.q. peuvent être également énoncées pour les multicorrélations en s'inspirant des résultats d'ergodicité à l'ordre 2 [123] [89] [180] [121]. On aurait par exemple, pour la consistence m.q. de la multicorrélation (3.113):

1. $x(t)$ admet des moments finis jusqu'à l'ordre 6,
2. $x(t)$ est stationnaire jusqu'à l'ordre 6,
3. $\frac{1}{T} \sum_{s=0}^{T-1} \mu_{(6),x}(s, \tau_1, \tau_2, s+\tau_1, s+\tau_2) \rightarrow \mathcal{C}^2_{(3),x}(\tau_1, \tau_2)$ si $T \rightarrow \infty$.

Le troisième point peut être traduit en une condition sur des séries de multicorrélations, comme pour l'ordre 2. Mais dans le cas présent, nous aurions pas moins de 40 séries distinctes, l'une en $\mathcal{C}_{(6)}$, 15 en $\mathcal{C}_{(2)}\,\mathcal{C}_{(4)}$, 9 en $\mathcal{C}_{(3)}\,\mathcal{C}_{(3)}$, et 15 en $\mathcal{C}_{(2)}\,\mathcal{C}_{(2)}\,\mathcal{C}_{(2)}$. Ces expressions ne sont pas données ici, mais on pourra les trouver dans *[63]*.

L'extension au cas complexe de certains résultats asymptotiques peut être trouvée dans [195].

Chapitre 4

Intervention des SOE dans quelques problèmes

J'ai sélectionné trois aspects des SOE dans ce chapitre, en me basant bien sûr sur des critères de convenance personnelle, mais aussi et surtout parce que ces sujets sont d'actualité. Le premier concerne les tests de normalité. Il est naturel d'aborder ce sujet en tout premier lieu puisque si les observations sont gaussiennes, il n'y a pas lieu de recourir aux SOE. Le deuxième concerne les mélanges linéaires de signaux, sujet qui a éveillé un intérêt croissant de la part de la communauté scientifique ces cinq dernières années. Et enfin, je pense qu'il est regrettable que l'aspect tensoriel des SOE ne soit que très rarement évoqué dans les approches multivariables. Ce sujet est donc abordé en dernier lieu.

4.1 Tests de gaussianité

Le test de normalité fait partie des tests d'hypothèse sans alternative. Autrement dit, si on définit l'hypothèse H_0 comme étant: "l'observation est gaussienne", nous n'avons rien d'autre à lui opposer que l'hypothèse contraire, $\bar{H}_0$. Ces tests de normalité sont parfois qualifiés d'*omnibus* [88] [101].

Dans une telle situation, un seul paramètre permet d'ajuster la détection: le *niveau* du test, ou erreur de première espèce, défini par:

$$\alpha = Prob(choisir\ \bar{H}_0/H_0\ vraie) \tag{4.1}$$

Une autre conséquence est qu'il ne peut exister de détecteur optimal au sens de la probabilité d'erreur, l'erreur de seconde espèce restant indéfinie.

Cette constatation est loin d'être anodine, car elle montre notamment que les propriétés statistiques de la variable-test n'ont besoin d'être connues que sous l'hypothèse H_0. Par exemple, si le kurtosis empirique est utilisé, il sera suffisant de connaitre ses quantiles sous hypothèse gaussienne.

Il existe d'autres tests standard sans alternative. Citons à titre d'exemple les tests de stationnarité, les tests de blancheur (plus ou moins forte) [99], les tests de réversibilité temporelle de processus [181], ou bien encore les tests de linéarité [127]. Le test de normalité est lié aux tests précédents dans le sens où:

- un processus gaussien non stationnaire peut apparaitre comme étant non gaussien s'il est supposé stationnaire;
- la plupart des tests de normalité supposent que les processus à tester sont blancs au sens fort (échantillons indépendamment et identiquement distribués), ce qui est une de leurs principales limitations;
- tout processus gaussien est réversible, mais la réciproque n'est pas vraie;
- tout processus gaussien est linéaire, mais il existe des processus linéaires non gaussiens.

On distingue principalement deux familles de tests de normalité: les tests *scalaires* et les tests *vectoriels*. Les premiers testent la normalité marginale des échantillons, et les derniers la normalité conjointe de plusieurs échantillons (par exemple consécutifs). Il est clair que la normalité conjointe entraine la normalité marginale, mais la réciproque n'est pas vraie.

Le fait de tester la normalité conjointe d'un nombre (forcément) limité d'échantillons entraine un biais dans la décision, dans le sens évidemment où la décision "gaussien" sera prise trop souvent. A contrario, l'hypothèse d'indépendance des échantillons (blancheur forte) entraine malheureusement un biais en sens inverse, de sorte qu'on ne sera pas en mesure d'obtenir des réponses fiables ni dans un sens ni dans l'autre.

C'est pourquoi un de nos travaux récents a consisté à développer un test vectoriel capable de s'affranchir de cette hypothèse de blancheur. Avant de décrire comment ce test a été construit, il est pertinent de passer en revue un certain nombre de tests connus.

4.1.1 Les tests existants

En reprenant la distinction scalaire/vectoriel précisée plus haut, on peut dresser un historique des tests les plus représentatifs. L'ensemble des tests est résumé dans le tableau 4.1.

a) Tests scalaires

1. Test du Chi-deux (1922): Si l'existence de la distribution dite "Chi-deux" remonte à 1838 avec les travaux de Bienaymé (1852 pour la loi du χ^2 à n degrés de liberté), son utilisation pour les tests d'ajustement de lois n'a pu voir le jour qu'avec la preuve de la convergence asymtotique du rapport de vraisemblance vers une variable du Chi-deux, preuve attribuée à Fisher en 1922. Dans ce rapport de vraisemblance, la densité des observations est remplacée par un histogramme calculé à partir d'intervalles de longueur prédéterminée. Notons que la convergence du rapport de vraisemblance vers une loi du Chi-deux pour des problèmes de détection plus généraux que l'ajustement de loi n'a été prouvée que plus tard par Wilks (1938) et Wald (1943).
2. Geary (1935): Geary propose comme variable test le rapport entre $E|x|$ estimée et l'écart-type empirique $\hat{\sigma}$; cette quantité vaut $\sqrt{2/\pi}$ dans le cas gaussien.
3. Kolmogorov-Smirnov (1948): Le test de Kolmogorov est basé sur la statistique d'ordre 1 de l'échantillon observé. La variable test est la distance L^∞ entre les fonctions de répartition estimées. Kolmogorov donne en 1933 l'expression analytique de la distribution asymptotique de cette variable test (sous forme d'une série); Smirnov ne la tabule qu'en 1948, date à laquelle son utilisation devient possible.
4. Pearson-Hartley (1962): Récapitulation sous forme de tables des quantiles de toutes les variables tests usuelles. En particulier, tables pour l'aplatissement (kurtosis) estimé pour divers temps d'intégration [172].
5. Shapiro-Wilk (1965): Ici la variable test est le rapport entre le carré de l'estimation linéaire de σ à partir de la statistique d'ordre d'ordre n et la variance empirique. Les coefficients de cet estimateur linéaire sont tabulés pour différentes longueurs d'échantillon et différents ordres d'estimateur [187].
6. Lilliefors (1967): Lilliefors modifie les tables de Smirnov pour permettre l'application du test de Kolmogorov au cas composite (moyenne et

variance de la loi gaussienne la plus proche inconnues).

7. Un test du même genre que celui de Shapiro et Wilk a été proposé par D'Agostino et al en 1971 [101]. L'obectif est toujours surtout les petits échantillons (de l'ordre de 50 ou 100).
8. Test du Chi-deux de Moore (1971): Le test du Chi-deux décrit plus haut a été sensiblement amélioré par Moore pour permettre son application lorsque l'histogramme des observations est calculé avec des cellules variables [161]. En outre, la solution qu'il a proposée est applicable en dimension supérieure à 1.
9. D'Agostino-Pearson (1973): Ce n'est qu'en 1973 que l'on voit proposer les cumulants standardisés comme test de normalité composites: ce sont les tests de l'asymétrie (ordre 3) et de l'aplatissement (ordre 4). La combinaison de ces deux variables en vue de la construction d'une variable test unique plus robuste est proposée en 1977 par les mêmes auteurs [102]. Voir aussi [171].
10. Stephens (1974) propose une nouvelle amélioration à la table des quantiles du test de Kolmogorov-Smirnov [193].
11. Gasser (1975) : Tests de l'asymétrie et de l'aplatissement pour des signaux colorés; variance et normalité asymptotiques [116].
12. Vasicek (1976): La densité gaussienne est celle des densités à support réel qui a la plus grande entropie. Le test est basé sur une estimation de l'entropie. Si la variable test atteint la borne gaussienne (connue) avec une tolérance acceptable, l'hypothèse gaussienne est acceptée [200].
13. Saniga-Miles (1979): En tant que tests scalaires sur des échantillons de taille supérieure à 50, l'ordre préférentiel est le suivant: Aplatissement, Asymétrie, d'Agostino, et enfin Shapiro-Wilk (peu utilisable sur des échantillons de taille supérieure à 100 d'après les auteurs) [184].
14. Mardia (1980) : L'auteur fait dans [157] une synthèse assez complète des tests de normalité scalaires, et il en recense plus de cinquante.
15. Lin-Mudholkar (1980): Parmi les nombreux tests de normalité ultérieurs, mentionnons le "Z-test". Ce dernier est basé sur un théorème prouvé par Cramér dans les années 40 disant que la moyenne empirique $\hat{m}$ et la variance empririque $\hat{\sigma}^2$ sont statistiquement indépendantes si et seulement si l'échantillon est gaussien. Lin et Mudholkar ont donc proposé un test basé sur la corrélation entre entre $\hat{\mu}$ et $\hat{\sigma}^2$ [149].

16. Moore (1982): La coloration du processus entraine une perte apparente de normalité dans les tests scalaires [162].
17. Hinich (1982): Test de normalité et de linéarité basé sur le bispectre. Permet de s'affranchir de la coloration, mais nécessite de très longs temps d'intégration [127]. Ne fonctionne que pour des signaux dissymétriquement distribués.
18. Anscombe-Glynn (1983): Autre méthode d'approximation de la densité de l'aplatissement estimé; la dernière approche en date était celle de [102].
19. Dallal-Wilkinson (1986): Nouvelle amélioration du calcul des quantiles dans le test de Kolmogorov-Smirnov [103].
20. Fukunaga-Flick (1986) : Utilisation d'une propriété indirecte des variables normales: le produit de deux densités gaussiennes est une gaussienne [113].

Remarques. D'après Stephens [193], le test du Chi-deux est moins performant que le test de Kolmogorov. De plus, il semble spécialement sensible à la dépendance des échantillons [162].

Les tests basés sur l'asymétrie ou l'aplatissement semblent très attractifs pour des durées d'intégration considérées comme élevées ($N > 1000$), alors qu'ils sont moins performants pour des échantillons petits ($N \leq 100$) [184]. D'autres tests, comme celui de D'Agostino [101], ou celui de Shapiro et Wilk [187], sont au contraire adaptés à des échantillons que nous qualifions de courts (càd N de l'ordre de 100) [171]. Le test du Chi-deux est reconnu comme étant moins puissant que le test de Kolmogorov [193], et que les autres tests d'ajustement [184].

D'autres façons de tester la gaussianité consistent à tester des propriétés (nécessaires et suffisantes) que doivent satisfaire les observations pour être gaussiennes. Nous en avons vu un exemple avec le test de Lin et Mudholkar, qui était basé sur le théorème de Cramér. Un autre test consisterait à mélanger linéairement deux voies, et à essayer de les séparer par ACI (Analyse en Composantes Indépendantes); ce test serait alors bassé sur le théorème de Darmois [43]. Ce dernier test ne présentrait pas grand intérêt, compte tenu de la difficulté que présenterait l'évaluation de son niveau.

Si le but du test de normalité est de déceler des différences décisives entre le bruit et le signal, il peut être plus approprié de mesurer l'écart entre les densités (ou fonctions de répartition) empiriques d'une voie bruit

et d'une voie signal. Le problème n'est donc plus un test de normalité mais un test différentiel de statistiques (une sorte de test d'homogéneité). Il existe d'ailleurs un test différentiel proposé par Smirnov. On peut également essayer de développer une distribution en série autour de l'autre, ce qui donnerait naissance à des cumulants centrés autour du bruit, et non autour de la loi normale. Bien que très séduisante, la faisabilité de cette seconde approche reste malheureusement encore à prouver.

Comme nous l'avons dit plus haut, une des limitations essentielles des tests est due à la dépendance des échantillons, qui introduit un biais dans les valeurs des seuils. C'est le même phénomène que celui observé pour le test du Chi-deux par Moore en 1982 [162]. L'influence de cette dépendance sur les tests d'asymétrie et d'aplatissement a été analysée par Gasser dans [116]. Ce problème se rencontre malheureusement aussi dans les tests vectoriels (cf. section b)).

b) Tests vectoriels

Il y a comparativement beaucoup moins de tests vectoriels de normalité. Nous avons relevé les tests suivants:

1. Mardia (1970) : Une première extension des tests scalaires, simple mais peu puissante, aux dimensions supérieures à 1 consiste à projeter les observations sur une droite arbitraire.

2. Mardia a proposé comme définition de l'aplatissement l'espérance mathématique du module à la puissance 4 des mesures standardisées, $E\{\rho_n^4\}$ [155]. Par construction, ce test est invariant par transformation affine. D'autres tests multivariables sont possibles [157].

3. Andrews et al (1973) : Les auteurs abordent surtout le cas de la dimension 2. Ils proposent de calculer le module carré ρ_n^2 et l'angle polaire θ_n des échantillons standardisés $y_n \stackrel{\text{def}}{=} V_x^{-1/2}\,(x_n - \bar{x}_n)$. Alors sous H_0, ρ_n^2 suit approximativement une loi du Chi-deux à deux degrés de liberté, et θ_n suit une loi uniforme [77].

   Cette idée s'étend pour le module carré en dimension quelconque $p > 2$, puisqu'alors ρ_n^2 suit approximativement une loi du Chi-deux à p degré de liberté; mais seul un des $p - 1$ angles est uniformément distribué [157, page 314].

4. Hinich (1982) : Le test de normalité est un cas particulier du test de linéarité, comme nous l'avons déjà souligné. Hinich est à l'origine du

	Temps/ Fréquence	Indépendants/ Dépendants	Scalaire/ Vectoriel
Fisher'22	T	I	S
Geary'35	T	I	S
Kolmogorov-Smirnov'48	T	I	S
Pearson-Hartley	T	I	S
Shapiro-Wolk'65	T	I	S
Lilliefors'67	T	I	S
Mardia'70	T	I	V
d'Agostino'71	T	I	S
Moore'71	T	I	S
d'Agostino-Pearson'73	T	I	S
Andrews-et al'73	T	I	V
Stephens'74	T	I	S
Gasser'75	T	D	S
Vasicek'76	T	I	S
Saniga-Miles'79	T	I	S
Mardia'80	T	I	S
Lin-Mudholkar'80	T	I	S
Rao-Gabr'80	F	D	V
Moore'82	T	D	S
Hinich'82	F	D	V
Anscombe-Glynn'83	T	I	S
Mardia-Foster'83	T	I	V
Mardia-Kanazawa'83	T	I	V
Dallal-Wilkinson'86	T	I	S
Fukunaga-Flick'86	T	I	S
Csorgo'86	T	I	V
Epps'87	T	D	V
DalleMolle-Hinich'89	F	D	V
Steinberg-Zeitouni'92	T	I	V
Moulines-et al'93	T	D	V
Giannakis-Tsatsanis'94	T	D	V
Tugnait'94	F	D	V
Comon-Deruaz'95	T	D	V

Table 4.1: Synoptique des principaux tests

premier test de linéarité de processus. Le test proposé est basé sur le bispectre des observations: il est constant si le processus est linéaire, et cette constante est nulle si le processus est gaussien.

5. Mardia - Foster (1983) : Les auteurs reprennent les propriétés aux ordres 1 et 2 de l'aplatissement vectoriel défini par Mardia en 1970, et établies en 1974. En outre, ils calculent le moment croisé entre l'asymétrie et l'aplatissement vectoriels pour des échantillons finis sous H_0. [158]. On remarquera par exemple que les estimateurs sont biaisés, mais consistants.

6. Mardia - Kanazawa (1983) : Il est proposé d'approximer la distribution de l'aplatissement vectoriel empirique par une loi du Chi-deux. Le moment d'ordre 3 de cet aplatissement est évalué analytiquement à cet effet [159].

7. Csörgö (1986) : Test de normalité basé sur la fonction caractéristique empirique. Le test suppose des échantillons indépendants [100].

8. Epps (1987) : Technique basée sur l'écart entre des statistiques empiriques et leur valeur exacte sous H_0. Comme dans [100], les hypothèses composites sont traitées (i.e. moyenne et variance inconnues). La fonction caractéristique est un des exemples possibles de statistique [106].

9. Steinberg - Zeitouni (1992) : Test basé sur l'écart entre l'entropie empirique sous hypothèse gaussienne (calculée à partir du spectre), et l'entropie du processus estimée de façon moins restrictive [192]. L'intérêt pratique de ce test reste à montrer expérimentalement, compte tenu de la complexité des expressions intervenant dans le calcul de cette entropie

10. Moulines et al (1992) adoptent d'abord une approche inspirée de Epps [106], basée sur une mesure de déviation de la fonction caractéristique [166].

    Ils proposent ensuite un autre test sans doute plus intéressant, procédant comme suit. On construit un processus $\hat{s}(t)$ à partir du processus à tester, $x(t)$. L'opération consiste à tester si la moyenne de $\hat{s}(t)$ est égale à sa moyenne sous H_0, s_0. On pensera notamment à incorporer dans $\hat{s}(t)$ des puissances de $x(t)$. La difficulté réside dans le calcul de la variance A de $\hat{s}(t) - s_0$. La variable-test utilisée sera $Q = N\,(\hat{s} - s_0)^T A^{-1} (\hat{s} - s_0)$, qui suit asymptotiquement une loi du χ^2.

11. Giannakis et Tsatsanis (1994) : Ces auteurs proposent de tester la nullité d'un ensemble de p valeurs de la fonction de bi- ou de tri-corrélation [121], rangées dans un vecteur Z. Pour évaluer le niveau du test, ils supposent que son estimation $\hat{Z}$ est approximativement gaussienne (limite asymptotique) et testent la variable standardisée $Q = N\,\hat{Z}^T A^{-1} \hat{Z}$, qui suit une loi du χ^2 à p degrés de liberté. Cependant, la variance A de $\hat{Z}$ doit être estimée, ce qui aura pour effet de diminuer le nombre de degrés de liberté de Q, du fait que $\hat{A}$ et $\hat{Z}$ seront corrélés, ce qui n'a pas été pris en compte.

12. Tugnait (1994) : Le test est basé sur une forme partiellement intégrée du bispectre ou du trispectre [199]. L'avantage par rapport au test de Hinich réside dans une réduction du coût calcul.

4.1.2 Statistiques du kurtosis multivariable

Le test qui nous semble posséder le meilleur compromis entre complexité de calcul et niveau est celui du kurtosis multivariable [157]. Cependant, ses quantiles n'ont été évalués sous H_0 que dans le cas i.i.d., comme nous allons le voir ci-après (section a)). C'est pourquoi nous avons poursuivi une étude plus approfondie de ce test, dans le cas d'échantillons colorés (section b)).

a) Cas i.i.d.

On suppose que les observations sont des vecteurs $X(n), 1 \leq n \leq N$, chacun de dimension p, et que la suite des $X(n)$ est stationnaire au second ordre, de moyenne zéro et de covariance S. En outre dans cette section, on admet qu'ils sont statistiquement indépendants, ce qui n'est évidemment quasiment jamais vérifié en pratique.

Le kurtosis multivariable de Mardia est une contraction du moment standardisé, définie par:

$$\mathcal{B}_p(N) = \frac{1}{N} \sum_{n=1}^{N} \Big(X(n)^T S^{-1} X(n) \Big)^2 , \tag{4.2}$$

Si on note $\hat{K}$ le tenseur moment centré d'ordre 4, et G l'inverse de la matrice de covariance S, alors l'écriture suivante est équivalente:

$$\mathcal{B}_p(N) = \sum_{i,j,k,l=1}^{p} \hat{K}_{ijkl}\, G_{ij}\, G_{kl}, \tag{4.3}$$

En pratique, la covariance S doit être remplacé par une estimée $\hat{S}$, corrélée avec les données $X(n)$, et la variable test devra être notée $\hat{\mathcal{B}}_p(N)$. Dans le cas i.i.d. qui nous occupe dans cette section, les trois premiers moments de ce kurtosis multivariable sont, sous l'hypothèse H_0:

$$\begin{aligned}
\mu_{1,\mathcal{B}} &= \mathrm{E}\{\hat{\mathcal{B}}_p(N)\} = p(p+2)\,\frac{N-1}{N+1}, \\
\mu_{2,\mathcal{B}} &= \mathrm{Var}\{\hat{\mathcal{B}}_p(N)\} = \frac{8p(p+2)}{N} + o(N^{-2}), \\
\mu_{3,\mathcal{B}} &= 64\,\frac{p(p+2)(p+8)}{N^2} + o(N^{-3}).
\end{aligned} \tag{4.4}$$

Ceci montre en particulier que pour N assez grand devant p (par exemple $N = 1000$ et $p = 2$), $\hat{\mathcal{B}}_p(N)$ peut être assimilée à une variable gaussienne. Evidemment, la normalité asymptotique de telles variables est connue [157] [136, ch.12], mais il est utile de savoir à partir de quelle valeur de N cette approximation est applicable sur le plan pratique.

Nous n'avons pas repris ces calculs dans le cas où $\mathcal{B}_p(N)$ serait construit sur le tenseur cumulant, mais il semble qu'en première évaluation, les résultats ne changent pas au second ordre. Dans la suite on conservera la définition construite sur le tenseur moment, afin de pouvoir effectuer des comparaisons.

b) Cas coloré

Considérons à présent notre problème original, et notons $x(t)$ le processus à tester, $1 \leq t \leq N$. Le test vectoriel portera sur la normalité conjointe d'un nombre limité p d'échantillons. Pour ce faire, on construit le vecteur $X(n)$ suivant:

$$X(n) = \begin{bmatrix} x(n\,\Delta + 1) \\ \vdots \\ x(n\,\Delta + p) \end{bmatrix}, \quad 1 \leq \Delta << N. \tag{4.5}$$

Le paramètre Δ est fixé et permet d'ajuster un recouvrement éventuel. Evidemment, les vecteurs $X(n)$ ne sont indépendants que si $\Delta - p$ est supérieur à la durée de corrélation du procesus $x(t)$. Or, il n'est pas toujours possible de faire en sorte que ce soit vrai, notamment si on désire 1000 réalisations identiquement distribuées, compte-tenu de la durée de stationnarité du processus. On se propose donc de calculer la moyenne et la

variance de la variable test:

$$\hat{\mathcal{B}}_p(N) = \sum_{i,j,k,l=1}^{p} \hat{K}_{ijkl}\, \hat{G}_{ij}\, \hat{G}_{kl}$$

sous ces nouvelles conditions, pour de grandes valeurs de N (*i.e.* N grand devant p).

Limites de l'approximation. On pose $\delta = \hat{S} - S$ et $\varepsilon = \hat{K} - K$. La variance de ces quantités est de l'ordre de N^{-1}. En effet, sous l'hypothèse H_0, on obtient en utilisant la notation de McCullagh [160]:

$$\begin{aligned}
\mathrm{Cov}\{\hat{S}_{ab}, \hat{S}_{cd}\} &= \frac{1}{N^2} \sum_{m,n=1}^{N} ([3] R_{a_n b_n} R_{c_m d_m} - R_{a_n b_n} R_{c_m d_m}) \,, \\
\mathrm{Cov}\{\hat{K}_{abcd}, \hat{K}_{efgh}\} &= \frac{1}{N} \sum_{m,n=1}^{N} ([105] R_{a_n b_n} R_{c_n d_n} R_{e_m f_m} R_{g_m h_m} \\
&\qquad - [3] R_{a_n b_n} R_{c_n d_n} \cdot [3] R_{e_m f_m} R_{g_m h_m}) \,.
\end{aligned}$$

où $R_{i_n j_m} \stackrel{\text{def}}{=} \mathrm{E}\{X_i(n) X_j(m)\} = C_{(n-m)\Delta + i - j}$ si C_τ désigne la fonction d'autocorrélation du processus $x(t)$. On peut aussi vérifier que la covariance croisée est aussi du même ordre.

Dans le cas scalaire, ces résultats se simplifient:

$$\frac{\mathrm{Var}\{\hat{S}\}}{S^2} = \frac{2}{N^2} \sum_{m,n=1}^{N} \frac{k_0^2(n-m)}{C_0^2}, \tag{4.6}$$

$$\frac{\mathrm{Var}\{\hat{K}\}}{K^2} = \frac{8}{N^2} \sum_{m,n=1}^{N} \frac{k_0^2(n-m)}{C_0^2} \left[1 + \frac{1}{3} \frac{k_0^2(n-m)}{C_0^2} \right], \tag{4.7}$$

où on a noté $k_0(s) = C_{s\Delta}$, pour alléger les écritures ultérieures. En particulier dans le cas i.i.d., on retrouve des résultats plus familiers:

$$\frac{\mathrm{Var}\{\hat{S}\}}{S^2} = \frac{2}{N}, \qquad \frac{\mathrm{Var}\{\hat{K}\}}{K^2} = \frac{32}{3N}. \tag{4.8}$$

On supposera dans la suite que ces variances sont petites, ce qui nécessite que la corrélation C_τ décroisse suffisamment vite vers zéro. En d'autres termes, il faut que le spectre de $x(t)$ soit lisse et à large support.

Développement à l'ordre 2. Développons la matrice $\hat{G}$ au second ordre en δ:

$$\hat{G} = G - G\delta G + G\delta G\delta G + O(\delta^3). \tag{4.9}$$

Le kurtosis multivariable (4.3) peut alors être approximé de la façon suivante, après quelques manipulations:

$$\begin{aligned}\hat{\mathcal{B}}_p = \mathcal{B}_p &+ \sum_{abcd}\sum_{ijkl}\frac{1}{p^4}\varepsilon_{abcd}G_{ab}G_{cd} \\ &-\frac{2}{p^2}K_{abcd}G_{ab}G_{ci}\delta_{ij}G_{jd} \\ &-\frac{2}{p^2}\varepsilon_{abcd}G_{ab}G_{ci}\delta_{ij}G_{jd} \\ &+K_{abcd}G_{ai}\delta_{ij}G_{jb}G_{ck}\delta_{kl}G_{ld} \\ &+2K_{abcd}G_{cb}G_{ai}\delta_{ij}G_{jk}\delta_{kl}G_{ld} + o(N^{-1}).\end{aligned} \tag{4.10}$$

Cette expression nous permet maintenant de calculer la moyenne et la variance de $\hat{\mathcal{B}}_p(N)$, à l'ordre deux en N^{-1}:

$$\begin{aligned}\mathrm{E}\{\hat{\mathcal{B}}_p\} &= \mathcal{B}_p + \sum_{abcd}\sum_{ijkl} -\frac{2}{p^2}G_{ab}G_{ci}G_{jd}\mathrm{E}\{\varepsilon_{abcd}\delta_{ij}\} \\ &\quad + K_{abcd}G_{ai}G_{ld}\mathrm{E}\{\delta_{ij}\delta_{kl}\}(G_{jb}G_{ck} + 2G_{cb}G_{jk}) + o(N^{-1}).\end{aligned} \tag{4.11}$$

et

$$\begin{aligned}\mathrm{Var}\{\hat{\mathcal{B}}_p\} &= + \sum_{\substack{abcd\\efgh}}\sum_{ijkl}\frac{1}{p^4}G_{ab}G_{cd}G_{ef}G_{gh}\mathrm{E}\{\varepsilon_{abcd}\varepsilon_{efgh}\} \\ &\quad -\frac{4}{p^2}K_{efgh}G_{ab}G_{cd}G_{ef}G_{gi}G_{jh}\mathrm{E}\{\varepsilon_{abcd}\delta_{ij}\} \\ &\quad +4K_{abcd}K_{efgh}G_{ab}G_{ci}G_{jd}G_{ef}G_{gk}G_{lh}\mathrm{E}\{\delta_{ij}\delta_{kl}\} + o(N^{-1}).\end{aligned} \tag{4.12}$$

A titre illustratif, on peut en déduire dans le cas scalaire (coloré) les expressions suivantes, que l'on peut vérifier par ailleurs par un calcul direct:

$$\mathrm{E}\{\hat{\mathcal{B}}_1\} \approx \mathcal{B}_1 - \frac{2}{S^3}\mathrm{E}\{\varepsilon\delta\} + \frac{3K}{S^4}\mathrm{E}\{\delta^2\}, \tag{4.13}$$

$$\mathrm{Var}\{\hat{\mathcal{B}}_1\} \approx \frac{\mathrm{E}\{\varepsilon^2\}}{S^4} + \frac{4K^2}{S^6}\mathrm{E}\{\delta^2\} - \frac{4K}{S^5}\mathrm{E}\{\varepsilon\delta\}. \tag{4.14}$$

Mais les expressions obtenues jusqu'à présent sont génériques, et il convient de les traduire en fonction uniquement des moments d'ordre 2, puisque nous sommes sous l'hypothèse H_0, et en tenant compte de la form particulière (4.5) du vecteur $X(n)$.

Statistiques avec la covariance exacte. Si la covariance S était connue, il serait inutile de développer G au second ordre comme nous l'avons fait plus haut en préliminaires. Le biais serait alors nul, la moyenne valant alors $p(p+2)$, et la variance s'obtiendrait en développant simplement $\hat{K} - K$ en fonction de C_τ. En réalité ce calcul s'est avéré extrêmement compliqué. En outre, nous savons que la variable $\mathcal{B}_p(N)$ aura une variance nécessairement plus grande que $\hat{\mathcal{B}}_p(N)$, puisque $\hat{S}$ et $\hat{K}$ sont corrélées (résultat général sur les variables-test studentisées).

Cependant, ce calcul a malgré tout été mené à terme, avec le recours à MAPLE pour de multiples vérifications. Nous avons obtenu:

$$\mathrm{Var}\{\mathcal{B}_1(N)\} = \frac{48}{N}\left[2 + \frac{1}{N}\sum_{s=1}^{N-1}(N-s)\,\frac{k_0^2(s)}{C_0^2}(3 + \frac{k_0^2(s)}{C_0^2})\right], \quad (4.15)$$

$$\mathrm{Var}\{\mathcal{B}_2(N)\} = \frac{16}{N}\left[20 + \frac{1}{N}\frac{\sum_{s=1}^{N-1}(N-s)q_0(s)}{(C_0^2 - C_1^2)^4}\right]. \quad (4.16)$$

avec, en ayant adopté la notation compacte $k_i \equiv k_i(s) \equiv C_{s\Delta+i}$:

$$q_0(s) = q_2(s) + 16(C_0^2 - C_1^2)^2\left[(2k_0^2 + k_1^2 + k_{-1}^2)C_0^2 - 4k_0(k_1 + k_{-1})C_0C_1 + 2(k_0^2 + k_1k_{-1})C_1^2\right], \quad (4.17)$$

où $q_2(s)$ est donnée en (4.23). Il n'est pas possible de présenter par écrit les résultats pour $p \geq 3$ en raison de leur longueur.

Statistiques avec la covariance estimée. En réalité, il faut considérer le cas où la covariance S est estimée par la covariance empirique $\hat{S}$. A présent, il faut exprimer non seulement G et K en fonction de C_τ, mais aussi les moments d'ordre 2 du couple (ε, δ). On garde la notation $k_i(s) \equiv C_{s\Delta+i}$ pour alléger les écritures.

Pour $p = 1$, nous obtenons après de longs calculs assités par MAPLE:

$$\mathrm{E}\{\hat{\mathcal{B}}_1\} = 3\left[1 - \frac{2}{N} - \frac{4}{N^2}\sum_{s=1}^{N-1}(N-s)\frac{k_0^2(s)}{C_0^2}\right], \quad (4.18)$$

$$\mathrm{Var}\{\hat{\mathcal{B}}_1\} \;=\; \frac{24}{N}\left[1+\frac{2}{N}\sum_{s=1}^{N-1}(N-s)\frac{k_0^4(s)}{C_0^4}\right]. \tag{4.19}$$

On peut vérifier que dans le cas i.i.d., la moyenne est bien de $-6/N$ et la variance de $24/N$, ce qui est conforme aux résultats classiques en la matière [138, vol.5, page 219], [156].

Le cas $p = 2$ est aussi intéresant à présenter, et doit sa (relative) simplicité au fait que la matice inverse G s'exprime encore assez simplement en fonction des éléments de S. Nous obtenons:

$$\mathrm{E}\{\hat{\mathcal{B}}_2\} = 4\left[2-\frac{4}{N}-\frac{1}{N^2}\frac{\sum_{s=1}^{N-1}(N-s)q_1(s)}{(C_0^2-C_1^2)^2}\right], \tag{4.20}$$

$$\mathrm{Var}\{\hat{\mathcal{B}}_2\} = \frac{16}{N}\left[4+\frac{1}{N}\frac{\sum_{s=1}^{N-1}(N-s)q_2(s)}{(C_0^2-C_1^2)^4}\right], \tag{4.21}$$

où les fonction $q_1(s)$ et $q_2(s)$ sont définies par:

$$q_1(s) = \left[(k_1+k_{-1})^2+8\,k_0^2\right]C_0^2 - 12\,k_0(k_1+k_{-1})C_1C_0 + 2\left[(k_1+k_{-1})^2+2k_0^2\right]C_1^2, \tag{4.22}$$

$$\begin{aligned} q_2(s) \;=\; & \left[8(k_0^2-k_1k_{-1})^2+3(k_1^2+k_{-1}^2)^2+12k_0^2(k_1-k_{-1})^2\right]C_0^4 \\ & +4\left[8k_0^4+3(5k_0^2+k_1k_{-1})(k_1+k_{-1})^2-4k_{-1}k_1(k_0^2+k_1k_{-1})\right]C_0^2C_1^2 \\ & +8\left(k_{-1}^2k_1^2+k_0^4+4\,k_0^2k_1\,k_{-1}\right)C_1^4 \\ & -24\,k_0\,(k_{-1}+k_1)\left[(k_1^2+k_{-1}^2+2\,k_0^2)C_0^2+2\,(k_{-1}\,k_1+k_0^2)C_1^2\right]C_0C_1 \end{aligned} \tag{4.23}$$

4.1.3 Résultats sur signaux

Le test de normalité décrit précédemment a été appliqué à des signaux synthétiques, et à des signaux réels d'acoustique sous-marine. Nous en reportons ci-après un extrait. Dans chacun des jeux d'essais, tous les tests ont utilisé la même estimatin de la fonction d'autocorrélation C_τ pour chaque p fixé, si bien que seules les estimations de la moyenne et de la variance de la variable-test sont différentes. On a choisi $\Delta = p$ dans tous les tests.

La table 4.2 presente les valeurs obtenues pour le rapport:

$$t = \frac{\mathcal{B}-\mathrm{E}\{\mathcal{B}\}}{Var\{\mathcal{B}\}^{1/2}}, \tag{4.24}$$

Test	$\hat{\mathcal{B}}_{1,iid}$	$\mathcal{B}_p$		$\hat{\mathcal{B}}_p$	
Dim. p	1	1	2	1	2
Formule	(4.4)	(4.15)	(4.16)	(4.19)	(4.21)
Gaussien					
iid	-0.165	-0.089	0.275	-0.165	0.643
MA(1)	0.248	0.115	0.251	0.249	0.598
MA(9)	-0.255	-0.062	0.043	-0.117	0.121
AR(2)	-0.106	-0.027	-0.139	-0.029	-0.279
Uniforme					
iid	-24.79	-12.39	-9.655	-24.79	-21.59
MA(1)	-22.86	-11.18	-9.191	-22.84	-20.79
MA(9)	-2.722	-0.666	-5.182	-1.620	-11.24
AR(2)	-2.252	-0.494	-1.281	-1.283	-2.859
Mer					
mer201	-1.067	-0.193	1.433	-0.630	3.197
mer202	1.661	0.355	2.009	1.339	4.613
mer204	1.482	0.281	1.591	1.154	3.666
mer205	1.866	0.340	7.074	1.316	15.59

Table 4.2: Valeurs du rapport t.

lorsque $\mathcal{B}$ désigne soit $\hat{\mathcal{B}}_p$ soit $\mathcal{B}_p$. Les formules ayant servi à estimer la moyenne et la variance sont rappelées en haut de chaque colonne.

Rappelons qu'asymptotiquement, $t = \pm 1.960$ correspond à une probabilité de détection de 95% (niveau 5%), et $t = \pm 1.645$ à 90% (niveau 10%). Les échantillons étaient de taille $N = 10000$. La fonction d'autocorrélation a été calculée à l'aide de toute la longueur de l'échantillon, mais seules les 200 premiers retards ont été pris en compte lorsque C_τ prenait une valeur significative.

Tous les calculs ont été exécutés ùa l'aide de MATLAB sur une station SUN4 SPARC5. Les simulations peuvent être reproduites en générant les séquences i.i.d. à partir des racines 12345 et 1234567, pour les bruits gaussien etuniforme, respectivement.

Conclusions. On constate que les processus colorés non gaussiens (pilotés par une innovation uniforme) sont souvent classés comme gaussiens lorsque les queues de corrélation sont assez longues (*e.g.* les MA(9) ou AR(2)), sauf pour notre test $\hat{\mathcal{B}}_2(N)$ qui s'en tire bien. Pour les bruits de mer, le test $\hat{\mathcal{B}}_2(N)$ a toujours conclu "non gaussien", contrairement aux réponses données par

le kurtosis scalaire sous hypothèse i.i.d. De même, le test non studentisé $B_2(N)$ conclue au gaussien dans ce cas, ce qui est vraisemblablement erroné.

Ce travail, très récent *[30]*, est encore incomplet. Il serait utile d'effectuer des simulations en nombre plus important. Par ailleurs, d'autres tests sont en cours sur signaux acoustiques sous-marins.

4.2 Mélanges linéaires

4.2.1 Taxinomie

Il existe en réalité un certain nombre de problèmes similaires –mais distincts– relatifs aux mélanges linéaires. Considérons le modèle d'observation suivant:

$$y(t) = [H(z)] \cdot x(t) + v(t), \tag{4.25}$$

où $y(t)$ est un processus vectoriel de dimension N, $H(z)$ une fonction de transfert de dimensions $N \times N_s$, $x(t)$ est un processus vectoriel dit "vecteur source", et $v(t)$ un bruit indépendant de $x(t)$. Si $N_s > 1$, on supposera que les sources $x_i(t)$ sont statistiquement indépendantes. En revanche, les sources $x_i(t)$ ne sont pas toujours supposées blanches (au sens fort, ou même à l'ordre 2).

Dans le modèle (4.25), il est clair que le couple (H, x) n'est pas uniquement défini en l'absence d'hypothèses supplémentaires. C'est pourquoi une contrainte est généralement admise pour conférer l'unicité à la solution. Nous allons passer en revue les options possibles page 63 et suivantes.

Suivant l'application, le problème consiste soit à identifier H (*e.g.* localisation en traitement d'antenne, identification de systèmes), soit à extraire le vecteur source (*e.g.* déconvolution, égalisation). La littérature est très abondante en ce qui concerne l'identification entrées-sorties, c'est à dire lorsque les entrées sont aussi observées [91] [124] [135]. On s'est d'ailleurs aperçu au fil des années qu'il était préférable d'admettre que les entrées aussi pouvaient être bruitées [189] [150].

On ne va cependant pas se pencher sur ce problème, mais sur celui de l'identification (ou de la déconvolution) dite "autodidacte", c'est à dire uniquement à partir de l'observation des sorties du système. Ce type d'identification a aussi reçu d'autres qualificatifs, tels que "aveugle", "myope", ou même "extralucide". Mais ces derniers semblent moins appropriés.

Lorsque $H(z)$ est une constante, on parlera de problème de séparation de sources, ou d'*Analyse en Composantes Indépendantes* si la cohérence tem-

porelle n'est pas expoitée (*e.g.* si toutes les sources ont même spectre), et de *séparation de signaux* dans le cas contraire. On parlera en revanche de *déconvolution autodidacte* ou *aveugle* (blind deconvolution en anglais) lorsque la fonction de transfert n'est pas réduite à une constante. La séparation de signaux et l'ACI sont donc des cas particuliers de la déconvolution autodidacte.

A priori deux familles d'approches sont possibles: l'approche identification consistant à estimer $H(z)$, et l'approche déconvolution, où les entrées sont estimées directement. Si on cherche au bout du compte à reconstruire les signaux-source, la première approche nécessitera alors le calcul des résidus (prédiction linéaire).

Annonçons tout de suite que:

1. Le problème de la séparation de signaux est soluble à l'ordre 2.
2. Le problème de l'ACI n'est pas soluble à l'ordre 2, sauf cas particulier (cf. sections suivantes), et le recours aux SOE est nécessaire.
3. Le problème de la déconvolution autodidacte n'est en général pas soluble à l'ordre 2, sauf dans le cas multivariable ($N > 1$) et sous certaines conditions portant sur la matrice $H(z)$, et sur le nombre de capteurs ($N > N_s$) [151].

Les méthodes de résolution actuelles ne sont pas adaptées à la présence de bruit, éventuellement non gaussien. Seules, les méthodes cherchant à maximiser un contraste peuvent supporter un bruit non gaussien, toutefois en dessous de 0dB (la mesure du rapport signal à bruit n'a de sens dans ce cas que si les sources et le bruit ont mêmes statistiques).

Notons que pour des durées d'intégration finies, ce qui est la situation pratique incontournable, les erreurs d'estimation sur les cumulants peuvent être vues au premier ordre comme du bruit additif non gaussien; d'où l'importance d'un minimum de robustesse des méthodes vis à vis du bruit non gaussien.

Pour y voir plus clair, tentons de dresser une liste des hypothèses que l'on peut faire dans les différentes approches:

Hypothèses sur les sources.

S0. *Les sources $x(t)$ sont indépendantes.* Cette hypothèse est commune à toutes les approches. Les approches à l'ordre r se contenteront bien sûr de l'indépendance à l'ordre r.

S1. *Les sources $x_i(t)$ sont chacune i.i.d.*

S2. *Les sources sont blanches à l'ordre 2.* Si les spectres des sources sont connus, on se ramène à cette hypothèse en reportant une racine du spectre de x_i dans la fonction de transfert $H_{ii}(z)$. Cette hypothèse est plus faible que S1.

S3. *Les sources sont de variance 1.* Avec l'hypothèse S2 (ou S1), cette condition entraine que la matrice spectrale des sources est constante et égale à l'identité.

S4. *Le nombre de sources N_s est strictement inférieur au nombre de capteurs N.* Cette hypothèse est nécessaire dans la séparation de sources à l'ordre 2.

S5. *Les sources ont une distribution discrète.* Cette hypothèse permet non seulement la séparation à l'ordre 2, mais permet aussi de s'affranchir d'autres hypothèses.

Hypothèses sur le filtre.

F1. *Les éléments diagonaux de $H(z)$ sont constants et valent 1.* Nous écrivons ceci: $DiagH(z) = I, \forall z$.

F2. *Les colonnes de $H(z)$ sont normalisées.* Autrement dit, $DiagH(z)^\dagger H(z) = I$, $\forall z$. Cette convention remplace F1 lorsqu'elle est plus agréable à manipuler.

F3. *La normalisation des colonnes de $H(z)$ est faite globalement sur tout le spectre: $Diag \oint H(z)^\dagger H(z) dz = I$.* Cette normalisation est plus faible que F2.

F4. *La matrice $H(z)$ est une matrice colonne de rang plein, pour tout z.* Ceci implique en particulier que, si $H(z)$ est un filtre RIF, alors il admet un inverse à gauche lui-même RIF; en d'autres termes, $H(z)$ est à minimum de phase.

F5. *$H(z)$ est un filtre RIF dont on connait exactement le degré.*

Propriétés dûes à un prétraitement de $y(t)$.

Y1. *Chaque observation $y(t)$ est de variance 1.*

Y2. *Chaque composante $y_i(t)$ est préalablement blanchie à l'ordre 2.* On supposera que cette opération est mené avec l'adoption supplémentaire des hypothèses S2 et S3, de sorte qu'on obtient alors la relation

$DiagH(z)H(z)^\dagger = I, \forall z$. Il s'agit alors d'une normalisation des lignes de $H(z)$.

Y3. *Le processus vectoriel $y(t)$ est préalablement blanchi à l'ordre 2 dans son ensemble.* Autrement dit, $E\{y(z)y(z)^\dagger\} = I$. Avec S2 et S3, cette opération entraine que $H(z)H(z)^\dagger = I, \forall z$, ce que l'on admettra sous Y3. Ce prétraitement est très fort, puisqu'il exploite entièrement les statistiques d'ordre 2.

Y4. *Les lignes de $H(z)$ sont normalisées globalement sur tout le spectre.* Autrement dit, $Diag \oint H(z)H(z)^\dagger dz = I$. Cette hypothèse est plus faible que Y2, et a fortiori que Y3.

Hypothèses sur le bruit

B1. *Le bruit $v(t)$ est gaussien et blanc temporellement.*

B2. *Le bruit est de cohérence spatiale G connue:* $E\{v(t)v(t)^\dagger\} = \sigma^2 G$, où σ est inconnue.

Chaque hypothèse conduira, suivant le type de problème considéré, à des algorithmes très différents, comme nous le verrons plus loin (cf. section 4.2.2 et suivantes).

4.2.2 Tour d'horizon bibliographique

Ce tour d'horizon n'a pas la prétention d'être totalement exhaustif. En revanche, son objectif est de donner un échantillonnage assez complet des problèmes relatifs à la déconvolution aveugle des mélanges linéaires qui sont ou ont été à l'étude.

a) Déconvolution scalaire

Le problème de la déconvolution "aveugle", c'est à dire sans séquence d'apprentissage, a été beaucoup étudié depuis les travaux de Sato et Godard dans les années 75-80. Ces algorithmes minimisent de façon itérative un critère mesurant l'écart entre une statistique de la sortie de l'égaliseur et la même statistique de la source *[63]*.

Benveniste, qui a analysé en profondeur le comportement de ce type de critère, ainsi que celui des algorithmes itératifs proposés pour le minimiser [83] [82], est à l'origine du qualificatif "aveugle".

Macchi et Eweda ont analysé de leur côté la convergence d'un algorithme d'égalisation consistant à minimiser l'écart entre la sortie et la sortie désirée

(assimilée à l'élément le plus proche dans un alphabet donné) par un algorithme du gradient stochastique [154].

Une des approches les plus prometteuses est sans aucun doute celle revendiquée par Donoho [105]. Après avoir défini une relation d'ordre entre variables aléatoires par leur écart à la normalité, il propose une famille de critères restant compatibles avec cette relation. Pour notre part, nous appellerons ces critères des *contrastes statistiques*. L'entropie et le kurtosis sont notamment des contrastes.

Shalvi et Weinstein ont proposé dans [186] un algorithme de maximisation itérative du kurtosis. Ils en ont analysé la convergence; on peut aussi consulter les commentaires de Tugnait à ce sujet publiés dans [198].

Lorsque la source est blanche au sens fort, il est toujours possible d'emprunter une approche *identification* ARMA (pouvant faire intervenir les SOE), et de calculer les résidus. Cette approche reste d'ailleurs applicable dans le cas multivariable (cf. page 70). Compte-tenu de l'énorme quantité de travaux publiés dans ce domaine, le plus simple est de se référer à l'article de synthèse de Swami [194].

Le choix entre les techniques basées sur les SOE ou celles de type Sato reste une question ouverte. On peut par exemple mentionner le débat ouvert par Proakis dans [177], présentant notamment les avantages des approches multispectrales.

Une variante est celle proposée par Bellini [80], où l'utilisation de cumulants "généralisés" est préconisée; ces cumulants sont définis comme étant la corrélation entre les observations et une filtrée non linéaire de ces dernières.

L'intérêt du recours aux SOE dans ce type de problème a été souligné par Lacoume dans [142].

La connaissance du fait que la source est de distribution discrète peut être exploitée, que la source soit blanche ou non. La question importante est de savoir comment un algorithme basé uniquement sur cette information se comporte en présence de bruit. Par exemple dans [118], Gassiat suppose que la fonction de transfert remplit toute la bande, et que le bruit est blanc gaussien et de variance inconnue.

b) Séparation de signaux

Le problème de la séparation de signaux, tel que nous l'avons défini en section 4.2.1, exploite le fait que les signaux ont des spectres d'ordre 2 différents. Cette idée a été proposée à l'origine par Fety [110], mais a été à notre avis mal exploitée, peut-être parce que en partie mal présentée. Cependant, tous

les éléments nécessaires s'y trouvent. Un exposé technique sera développé en section 4.2.3 page 72, et donnera l'essentiel de cette approche, présenté de la façon qui s'impose à mon avis *[68]*.

Tong et Liu ont, indépendamment et bien plus tard, exposé leur approche du problème, et ont analysé les conditions d'identifiabilité de la matrice de mélange [196].

Plus récemment, VanGerven a proposé une approche applicable seulement dans le cas de 2 sources et de 2 capteurs [119], à mon avis peu intéressante. En effet, elle procède par recherche exhaustive à la résolution d'un système d'équations non linéaires. Leur approche ressemble un peu à celle de Nguyen-Thi [168], mais pour un problème bien plus simple.

En revanche, une étude plus intéressante a été publiée par Belouchrani et alterae [81]. Elle consiste à diagonaliser conjointement (de manière évidemment approximative) un ensemble de matrices de covariance (ce concept avait déjà été introduit par Souloumiac pour l'ACI [190]); en outre, une évaluation des performances asymptotiques est présentée. Cette approche est a priori meilleure que celle de Tong puisqu'elle est basée sur plusieurs matrices de covariance.

Avant de conclure, il convient de souligner l'intérêt malgré tout assez limité de ce genre de problème, vu du côté des applications. L'identification de mélanges instantanés est surtout rencontré en tant que sous-problème de mélanges réels, qui sont en très grande majorité convolutifs. Dans ce sous-problème, il s'avère justement que les signaux sont blancs, et que les conditions d'identifiabilité requises dans la *séparation de signaux* ne sont pas vérifiées. On se trouve en présence du problème d'Analyse en Composantes Indépendantes (ACI), sur lequel nous allons nous pencher maintenant.

c) Séparation de sources (ACI)

Le problème de l'ACI a été introduit à l'origine par Jutten et Hérault [132]. Un algorithme itératif de nature neuro-mimétique était proposé pour sa résolution. A l'époque, son fonctionnement était passé pour un mystère, qui a trouvé explication ultérieurement *[49] [51] [6]*. Le même problème a été abordé par Bar-Ness quelques années plus tôt, mais de manière très différente, et avec vraisemblablement une moins bonne compréhension de l'outil et de sa portée [79].

La nécessité de recourir aux SOE pour permettre l'identifiabilité a été reconnue indépendamment par Lacoume [144] et Comon *[49] [50] [48]*. Fort de cette constatation, Lacoume et Ruiz proposaient une identification basée

sur la minimisation de la somme des carrés des cumulants croisés [144]. Cependant, cet algorithme était coûteux, et pratiquement inutilisable pour plus de trois sources.

L'approche de Gaeta a été celle du maximum de vraisemblance. Il a constaté que si les cumulants d'ordre 3 sont nuls, alors la vraisemblance approchée revient au critère précédent [114]. On notera en particulier que ce sera le cas des variables aléatoires complexes circulaires. Cette analyse donne un autre éclairage du problème, mais ne propose pas de nouvel algorithme numérique. Par ailleurs, ce critère rejoint celui proposé par Cardoso et Souloumiac d'une part [97] [190], et celui proposé par Comon d'autre part *[43]*, dont nous allons parler dans un instant.

Dans le même contexte, il est possible de calculer la borne de Cramér-Rao; ces résultats sont présentés par Lacoume et Harroy, avec le formalisme nécessaire au cas complexe [143]. Il est important de souligner que dans le cas complexe, la préservation de la structure tensorielle des SOE est précieuse [142].

Souloumiac et Cardoso ont mis au point une technique de diagonalisation conjointe de plusieurs matrices (en nombre supérieur à deux, ce qui n'est possible qu'approximativement) [191] [96] [190] [97]. Il a été constaté que ses performances théoriques sont similaires à l'approche par contrastes de Comon [94]. Cette constatation a été confirmée indépendamment par Chevalier quelques années plus tard sur le plan expérimental [98].

Les critères de contrastes, déjà introduits dans le cas scalaire pour la déconvolution (cf. section a) page 65), sont des critères intéressants dans le cas de l'ACI car ils permettent de conférer une certaine optimalité au sens probabiliste aux solutions obtenues, notamment en présence de bruit, éventuellement non gaussien *[21] [2]*. Ces critères seront présentés plus en détail dans la section 4.2.5 page 83.

Comme cela a été dit plus haut, dans le cas où les cumulants d'ordre 3 sont nuls, la fonction de vraisemblance approximée coincide avec un contraste. Mais d'autres critères faisant intervenir les cumulants d'un ordre fixé r quelconque plus grand que 2 sont des contrastes *[2]*. L'information mutuelle est elle aussi un contraste, et peut être approximée par un développement d'Edgeworth de la densité. Ceci donne lieu à des critères faisant intervenir les cumulants d'ordre 3 et 4 *[2]*; les expressions correspondantes sont (4.60) (4.66).

Un algorithme d'optimisation a été également proposé par Comon, basé sur les rotations de Givens *[43] [2]*. Les solutions sont particulièrement simples dans le cas de deux sources, comme on le montre dans la section 4.2.6.

Hélas cette simplicité n'est conservée dans le cas complexe qu'en l'absence de bruit *[21] [3]*.

Plusieurs solutions récursives ont été proposées dans la littérature pour réaliser la séparation de sources. La première est décrite dans l'article en deux volets [133] [6]; cet algorithme est du type Robbins-Monro. Une autre a été proposée par Comon, mais n'est pas non plus du type gradient *[50]*. Par contre Moreau et Macchi ont proposé plusieurs algorithmes de type LMS minimisant des critères de contraste [164] [165] [163].

Cardoso et Laheld introduisent le concept de gradient relatif [145], permettant d'atteindre des performances qui ne dépendent que du niveau de bruit (qui est supposé faible) et de la distribution des sources, et pas de la matrice de mélange [94].

Par ailleurs, Cardoso a expliqué comment faire un usage optimal des cumulants d'ordre 4 dans un contexte d'ajustement de modèle [95], pour le problème de l'ACI. Il s'agit ici de l'ajustement des cumulants d'entrée ou de sortie.

Enfin, signalons qu'une fonction de contraste a été proposée par Krob pour l'identification de transformations linéaires-quadratiques [141]. Des articles longs sur ce sujet devraient paraitre prochainement.

Un résultat plus curieux est celui de Gamboa, qui a montré récemment que l' identifiabilité est possible à l'ordre 2 si sources sont de distribution discrète [115].

Très peu d'auteurs se sont penchés sur le cas où le nombre de sources est supérieur au nombre de capteurs, et ce sujet est encore très prospectif. Il semblerait que l'on doive soit utiliser un modèle de réception [93], mais alors il ne s'agit plus d'identification aveugle, soit restreindre le type de mélange à une classe paramétrée, par exemple les retards purs [31], soit avoir recours à des outils d'algèbre multilinéaire (cf. section 4.3), qui sont hélas d'une grande complexité [24]. Quoiqu'il en soit, il est clair que si l'identification est parfois possible, la séparation des sources, elle, ne l'est pas, du moins de façon exacte, même asymptotiquement.

d) Déconvolution vectorielle à l'ordre 2

Comme cela a été annoncé page 63, le problème de déconvolution aveugle peut être soluble à l'ordre 2 sous certaines conditions dans le cas multivariable ($N > 1$). Cette constatation, qui peut paraite un peu surprenante, a été bien résumée par Loubaton [151].

Si le filtre $H(z)$ est causal et d'inverse causal et stable, alors il est identifiable à l'ordre 2 seulement, à une ACI près. Si de plus il est FIR de rang plein pour tout z, alors le spectre des sources peut être aussi identifié. Tong et Xu [197] sont à l'origine de l'idée de départ, qui a été ensuite améliorée par Moulines et d'autres co-auteurs [167]. Cette solution fonctionne en présence d'une seule source.

Gesbert a proposé récemment une implantation adaptative de cette solution dans [120], toujours pour une source. Par ailleurs, Abedmeraim et alterae ont généralisé la résolution au cas de plusieurs sources [75]. Une autre direction de généralisation est celle de la coloration des sources; Fijalkow et Loubaton ont proposé récemment une technique pour traiter le cas d'une source ARMA, en présence de bruit corrélé spatialement [111].

L'idée d'identifier un modèle ARMA monique suivi d'une ACI avait été proposé par Comon dans [49] ou [21]. Mais l'identification MA monique faisait appel aux SOE, ce que l'on peut désormais éviter.

e) Déconvolution vectorielle avec les SOE

L'idée la plus ancienne consiste à supposer que les sources sont blanches temporellement (au sens fort) et spatiallement (indépendantes), et à identifier un modèle linéaire, ARMA par exemple. L'estimation des sources se fait alors par le calcul des résidus (prédiction linéaire), comme dans le cas scalaire. Même dans le cas vectoriel, il y a aujourd'hui un très grand nombre de techniques qui ont été proposées. On peut se référer aux travaux de Swami et al [194], ainsi qu'aux synthèses dressées par l'équipe de Favier [104] [108] [109]. La différence fondamentale entre les cas scalaire et vectoriel est que le recours à l'ACI est nécessaire pour terminer l'identification si le modèle n'est pas monique (*i.e.* $B_0 \neq I$) *[49] [5]*.

L'idée la plus simple qui vient ensuite à l'esprit est celle de l'extension au cas convolutif des premiers algorithmes itératifs proposés pour l'ACI, comme celui de Hérault et Jutten [125]. C'est ce qu'on fait Jutten et Nguyen-Thi dans [168], par annulation itérative des cumulants de type 31 et 13, mais sans grand succès. On connait les problèmes de convergence dont souffre déjà cet algorithme dans le cas instantané *[6]*. Plus précisément, ces problèmes sont d'une part la vitesse de convergence (très variable, et parfois extrêmement lente), mais aussi l'absence de solutions parmi les points stationnaires de l'algorithme, suivant les circonstances. Ces problèmes risquent d'être encore plus insurmontables dans le cas convolutif.

Comon s'est penché sur la faisabilité des approches multispectrales bande

étroite. L'idée est très simple: si on se place en bande étroite, le mélange devient complexe instantané, et peut être identifié par ACI. Toutefois, cette approche est peu satisfaisante. En effet, si le rapport signal à bruit est faible, il peut être indispensable d'utiliser toute la bande du signal utile. Il se trouve qu'il est très difficile de fusionner les résultats obtenus dans chacune des bandes, même dans le cas de mélanges très simples comme les retards purs amortis *[21]*, à cause de l'indétermination présente dans le modèle (permutation).

Il est clair que la solution du problème des mélanges convolutifs est indéterminée à une matrice diagonale de filtres près, et à une permutation près. On peut se débarrasser de la première indétermination en imposant (éventuellement provisoirement) une contrainte telle que S1 ou F1. En revanche, on ne peut contourner efficacement la seconde, qui n'est d'ailleurs généralement pas gênante, sauf précisément dans le cas présent. En effet, il est important de remarquer qu'en scindant un problème à bande large en plusieurs problèmes à bande étroite, on multiplie artificiellement l'indétermination en autant de permuations que de bandes. C'est là que se situe la maladresse.

Capdevielle propose une technique de détermination des permutations basée sur les liens statistiques pouvant exister entre différents canaux fréquentiels [92]. Cette approche est inspiré de [185], où un prétraitement avait été nécessaire pour rehausser la corrélation et améliorer éventuellement le conditionnement; le problème était celui d'un filtrage RIF avec accès à une réponse désirée. Cette technique risque vraisemblablement de ne pas marcher lorque les hypothèses du théorème de la limite centrale s'appliquent. En revanche, on peut espérer qu'elle fonctionne lorsque les sources contiennent des raies spectrales corrélées (par exemple, pour des machines tournantes). Certains liens doivent exister avec les travaux de Krob [139].

Dans le cas de $N_s = 2$ sources, Yellin et Weinstein ont proposé un algorithme itératif basé sur une propriété que doivent vérifier les multispectres croisés entre l'observation et la sortie du filtre déconvolueur. Il s'agit d'une approche large bande, qui peut être formulée dans le domaine temps. Présenté à l'origine à l'ordre 3 [202], l'algorithme a été étendu à l'ordre 4 [203] (ce qui était d'ailleurs évident). Cependant, cet algorithme possède de fortes limitations; en effet, on ne dispose pas de preuve de sa convergence, et il n'a été justifié et expérimenté qu'en l'absence totale de bruit.

Inouye a étendu récemment l'algorithme de Shalvi basé sur la maximisation du kurtosis au cas multivariable [130]. Ceci semble intéressant, et rejoindrait les axes de recherche que nous proposons, liés aux fonctions de

contraste. Quelques outils sont suggérés dans la section 4.2.5 à cette intention, et devront être confrontés aux travaux de Inouye (ce qui n'a pas encore été fait).

Enfin, reste le cas des mélanges particuliers, comme celui des retards purs (non multiples de la période d'échantillonnage) amortis. Il a été montré par Emile et Comon qu'il était possible d'identifier directement (non itérativement) la fonction de transfert. En outre, l'identification du mélange reste possible dans le cas où le nombre de sources est supérieur au nombre de capteurs, ce qui est une sorte de curiosité [31].

On va à présent exposer la philosophie de la séparation de signaux. On abordera ensuite les outils relatifs aux fonctions de contrastes (dédiés aux mélanges instantanés ou convolutifs), pour enfin se pencher sur l'ACI.

4.2.3 Séparation de signaux

a) Mélanges instantanés inversibles de signaux

Supposons que l'on observe un signal aléatoire $y(t)$ à N composantes, et que ce dernier satisfasse le modèle linéaire suivant:

$$y(t) = Hx(t), \tag{4.26}$$

où H est une matrice carrée inversible, et $x(t)$ un signal dont les N composantes $x_n(t)$ sont statistiquement indépendantes et non identiquement nulles. Ce mélange sera dit *instantané* car la réponse impulsionnelle du filtre dont les entrées sont $x_n(t)$ et les sorties $y_n(t)$ est une constante. La question que l'on se pose est de savoir s'il est possible d'identifier la matrice H uniquement à partir de l'observation des sorties $y_n(t)$.

Proposition 4.2.1 *S'il existe une solution particulière $(H_o, x_o(t))$, alors il existe toute une classe de solutions $(H, x(t))$ de la forme $H = H_o\Lambda P$, $x(t) = P^T\Lambda^{-1}x_o(t)$, se déduisant de la solution particulière par un changement d'échelle Λ (matrice diagonale régulière) et une permutation P.*

Dans cette mesure, on peut dire que le problème est mal posé. On peut soit chercher un représentant canonique de la classe d'équivalence des solutions, soit une solution particulière quelconque, sachant que la seconde donnera accès à la première, et que toutes deux pourront générer l'ensemble des solutions. Nous décrivons donc maintenant une méthode permettant

d'obtenir une solution particulière, dont le principe a été proposé à l'origine par Fety [110, p.109].

La matrice de covariance de l'observation, $\Gamma_{ij}(\tau) = \mathcal{C}\{y_i(t)\, y_j(t+\tau)\}$, s'écrit en fonction de la matrice de covariance de x, notée $C_{ij}(\tau)$:

$$\Gamma(\tau) = H\, C(\tau)\, H, \tag{4.27}$$

où la matrice $C(\tau)$ est diagonale quelle que soit la valeur de τ, puisque les composantes de $x(t)$ sont indépendantes. Une façon d'aborder ce problème est de construire les deux matrices suivantes:

$$\Gamma_1 = \sum_\tau \alpha_\tau\, \Gamma(\tau), \text{ et } \Gamma_2 = \sum_\tau \beta_\tau\, \Gamma(\tau), \tag{4.28}$$

où α_τ et β_τ sont des coefficients scalaires arbitraires qu'il faudra choisir de façon à satisfaire les conditions d'identifiabilité (que nous allons aborder ci-après).

Il existe au moins un jeu de coefficients $\{\alpha_\tau, \beta_\tau\}$ tel que Γ_2 soit inversible. En effet, $\Gamma(0)$ est par exemple inversible car H et $C(0)$ le sont. On peut donc dorénavant supposer que Γ_2 est inversible sans restreindre la généralité. Les matrices Γ_1 et Γ_2 vérifient:

$$\Gamma_1 = H\, K_1\, H^T, \text{ et } \Gamma_2 = H\, K_2\, H^T, \tag{4.29}$$

où K_1 et K_2 sont des matrices diagonales. Soit la décomposition en *éléments propres généralisés* suivante:

$$\Gamma_1\, U = \Gamma_2\, U\, \Lambda, \tag{4.30}$$

où U est une matrice inversible et Λ une matrice diagonale. Les colonnes de U sont les vecteurs propres du faisceau $\{\Gamma_1, \Gamma_2\}$ et les éléments de Λ les valeurs propres associées (valeurs propres d'une matrice $\Gamma_2^{-1/2}\Gamma_1\Gamma_2^{-T/2}$).

Proposition 4.2.2 *La matrice H peut être identifiée à une permutation et un facteur d'echelle près si et seulement si les valeurs propres Λ_{nn} du faisceau $\{\Gamma_1, \Gamma_2\}$ sont toutes distinctes.*

Pour démontrer cette proposition, il est plus clair d'introduire les deux lemmes suivants *[68]*.

Lemme 4.2.3 *Si une matrice W inversible satisfait une relation $KW = W\Lambda$, où K et Λ sont diagonales, alors il existe (au moins) une permutation P telle que $K = P^T \Lambda P$.*

Démonstration. Nous avons $K_{ii}W_{ij} = W_{ij}\Lambda_{jj}$, pour tout couple (i,j). Comme W est de rang plein, il existe au moins un élément non nul W_{aj} dans chaque colonne j, ce qui montre que pour tout j, il existe un a tel que $\Lambda_{jj} = K_{aa}$. Ce résultat peut être aussi vu comme une conséquence de l'unicité de la décomposition spectrale. □

Lemme 4.2.4 *Les seules matrices A satisfaisant $A\Lambda = \Lambda A$, où Λ est diagonale de composantes toutes distinctes, sont les matrices diagonales.*

Démonstration. La matrice A doit vérifier pour tout couple (i,j) la relation $A_{ij}(\Lambda_{ii} - \Lambda_{jj}) = 0$. Il est alors évident que si les composantes de Λ sont toutes distinctes, $A_{ij} = 0$ pour $i \neq j$. □

Démonstration. (proposition 4.2.2). Supposons Γ_2 inversible, sans restreindre la généralité. Les matrices H et Γ_2 étant inversibles, la relation $HK_1H^T\,U = HK_2H^T\,U\Lambda$ implique que $K_2^{-1}K_1\,H^TU = H^TU\,\Lambda$. Or d'après le lemme 4.2.3, nous avons nécessairement $K_2^{-1}K_1 = P^T\Lambda P$, où P est une permutation. Il vient donc que $\Lambda\,PH^TU = PH^TU\,\Lambda$. Maintenant d'après de lemme 4.2.4, la matrice PH^TU doit être diagonale. Appelons-la Δ. On a donc finalement $H = U^{-T}\Delta P$, ce qui montre que H a été identifiée aux matrices multiplicatives Δ et P près. □

b) Mélanges instantanés singuliers

Si Γ_1 et Γ_2 sont singulières, les conditions d'identifiabilité de la procédure précédente ne sont plus forcément valables. Il en est de même si H est rectangulaire ou carrée singulière. Nous qualifions ces modèles de *singuliers*.

En pratique, si on s'attaque à un problème d'identification *aveugle*, on ne sait en général pas grand chose sur la matrice H, et il pourrait bien se faire notamment qu'elle soit singulière. Cela ne peut et ne doit se détecter dans le cadre de l'identificaton aveugle, qu'à partir des observations $y(t)$. Dans tous les cas, le modèle

$$y(t) = H\,x(t) + v(t) \tag{4.31}$$

est sans doute plus réaliste, où la matrice H est éventuellement singulière, et où $v(t)$ est un signal de "nuisance", indiquant l'écart entre l'observation réelle et le modèle idéal. Pour un système ayant moins d'entrées que de sorties, H aura pour rang au plus le nombre d'entrées s'il n'y a pas de bruit. Inversement, pour un système ayant plus d'entrées que de sorties, N entrées

seront prises en compte dans la partie $Hx(t)$, et les autres devront figurer dans le terme $v(t)$ au titre de nuisances.

Pour tenter de discerner les différentes singularités, nous proposons la procédure suivante.

1. Si $\Gamma_y(0)$ est singulière, alors la matrice H n'est pas inversible. Notons la décomposition spectrale de $\Gamma_y(0)$ comme suit:

$$\Gamma_y(0) = R\,S\,R^T, \tag{4.32}$$

   où S est diagonale inversible de dimension $r \times r$. Les matrices H et $\Gamma_y(0)$ ayant le même espace image (de dimension r), il existe une matrice $\bar{H}$ de rang plein est de taille $r \times N$ telle que $H = R\bar{H}$. On a calculé R, voyons maintenant comment identifier $\bar{H}$. On peut poser

$$\bar{y}(t) = R^T\,y(t), \tag{4.33}$$

   et considérer le modèle $\bar{y}(t) = \bar{H}\,x(t) + \bar{v}(t)$, où $\Gamma_{\bar{y}}$ est inversible. On est alors ramené au point suivant.

2. Si $\Gamma_y(0)$ est régulière, on peut appliquer la procédure décrite dans la section a) pour identifier le mélange. Si les valeurs propres Λ_{nn} sont distinctes, on peut estimer les signaux source par la relation

$$\hat{x}_i(t) = U^T y(t). \tag{4.34}$$

   De deux choses l'une. Ou bien certains signaux $\hat{x}_i(t)$ obtenus sont suffisamment décorrélés entre eux, et le modèle (4.26) est satisfait pour ces composantes, ou bien il reste une corrélation importante entre toutes les composantes de $\hat{x}(t)$, et on peut conclure par la présence d'une nuisance $v(t)$ importante. Cette dernière peut être dûe à du bruit de mesure, ou bien à la présence d'autres sources que les $x_i(t), 1 \leq i \leq N$. Dans ce dernier cas, il est nécessaire d'avoir plus d'informations pour pouvoir identifier le mélange. On pourra notamment recourir à une méthode spécifique utilisant les statistiques d'ordre supérieur (cf. section 4.2.5).

3. Si certaines valeurs propres Λ_{nn} sont confondues, cela veut dire que la diversité des fonctions $\Gamma_{ij}(\tau)$ n'est pas assez riche pour permettre de conclure, et il faudra recourir à des statistiques d'ordre supérieur (i.e. Analyse en Composantes Indépendantes).

On pourra également utiliser les techniques d'ordre supérieur pour confirmer ou affiner un résultat obtenu avec les statistiques d'ordre 2.

Cette section n'a exposé que schématiquement le principe de la résolution du problème de la séparation de signaux *[68]* [110]. Une étude plus approfondie du problème peut être trouvée dans [81].

L'hypothèse S0 d'indépendance des sources est dans tous les cas au centre de toutes les approches. C'est pourquoi il convient dans un premier temps d'introduire diverses mesures d'indépendance statistique. En outre, ces éléments permettent de donner des fondements théoriques aux approches par maximisation de contrastes *[2]*.

4.2.4 Indépendance statistique

Dans cette section, on passe en revue quelques moyens de mesurer l'indépendance statistique, sur les plans théorique et pratique.

a) Information mutuelle

Définition 4.2.5 *Soit x un vecteur aléatoire de dimension N admettant une densité $p_x(u)$. Les composantes x_i de x sont dites indépendantes si et seulement si la distribution conjointe des x_i est égale au produit de leurs distibutions marginales:*

$$p_x(u) = \prod_{i=1}^{N} p_{x_i}(u_i). \tag{4.35}$$

Une façon naturelle de mesurer l'indépendance des variables x_i est donc de mesurer la distance $\delta(p_x, \prod_i p_{x_i})$ entre ces deux densités. Parmi toutes les mesures de distance disponibles, l'une est particulièrement usitée. Il s'agit de la divergence de Kullback:

$$\delta(p_x, p_y) \stackrel{\text{def}}{=} \int p_x(u) \, \log \frac{p_x(u)}{p_y(u)} \, du. \tag{4.36}$$

Noter le vocabulaire: le mot *divergence* a été utilisé, car cette mesure d'écart n'est pas une fonction symétrique de ses arguments, et ne mérite donc pas le titre de *distance*.

Proposition 4.2.6 *La divergence de Kullback est toujours positive et vérifie:*

$$\delta(p_x, p_y) = 0, \text{ si et seulement si } p_x(u) \stackrel{\text{pp}}{=} p_y(u). \tag{4.37}$$

Démonstration. Pour tout réel positif w, on a l'inégalité de convexité $\log w \leq w - 1$, avec égalité si et seulement si $w = 1$. En appliquant cette inégalité au rapport $p_y(u)/p_x(u)$, on obtient:

$$-\delta(p_x, p_y) \leq \int p_x(u) \, [\frac{p_y(u)}{p_x(u)} - 1] \, du.$$

Or le second membre est toujours nul puisque $\int p(u) = 1$ pour toute densité de probabilité. Par ailleurs comme la fonction $\log w$ est tangente à $w - 1$ en $w = 1$, l'égalité n'a lieu que si $p_y(u)/p_x(u) = 1$ pour presque tout u. □

Proposition 4.2.7 *La divergence de Kullback est invariante par transformation inversible.*

Démonstration. Posons $Y = Ay$ et $X = Ax$, où A est une matrice inversible. Alors $p_X(v) = p_x(A^{-1}v)/|det(A)|$, et $P_Y(v) = p_y(A^{-1}v)/|det(A)|$. Il vient par conséquent

$$\delta(p_X, p_Y) = \int p_x(A^{-1}v) \, \log \frac{p_x(A^{-1}v)}{p_y(A^{-1}v)} \, \frac{dv}{|\det A|}.$$

On pose $u = A^{-1}v$, c.à.d. $dv = |\det A| du$. Alors

$$\delta(p_X, p_Y) = \int p_x(u) \, \log \frac{p_x(u)}{p_y(u)} \, du,$$

ce qui termine la démonstration. □

La divergence de Kullback appliquée à $p_y(u) = \prod p_{x_i}(u_i)$ conduit à la mesure d'indépendance suivante:

$$I(p_x) = \int p_x(u) \, \log \frac{p_x(u)}{\prod_{i=1}^{N} p_{x_i}(u_i)} \, du. \tag{4.38}$$

Cette quantité n'est autre que l'information mutuelle moyenne, bien connue en codage et en télécommunications. En vertu de la proposition 4.2.6, l'information mutuelle est toujours positive, et s'annule si et seulement si les variables x_i sont indépendantes.

Contrairement à ce que l'on pourrait croire, l'information mutuelle n'est pas invariante par changement de base, bien que la divergence de Kullback le

soit. Pour s'en convaincre, il suffit de considérer le contre-exemple suivant. Prenons une variable gaussienne de dimension N, de densité

$$\Phi_x(u) = [2\pi]^{-N/2} \, [\det V]^{-1/2} \, e^{-\frac{1}{2}u^T V^{-1} u}, \tag{4.39}$$

où V est une matrice de covariance inversible. Alors son information mutuelle est donnée par:

$$I(\Phi_x) = \frac{1}{2} \log \frac{\prod V_{ii}}{\det V}. \tag{4.40}$$

Prenons comme changement de base une matrice A telle que $AA^T = V^{-1}$. Alors l'information mutuelle après changement de base est $I(\Phi_{Ax}) = 0$. Il faudrait que la covariance V soit diagonale, ou que la matrice A soit de la forme ΛP, Λ diagonale et P permutation, pour que l'information mutuelle ne change pas. En d'autres termes, si $X = Ax$, nous n'avons en général pas $\prod p_{x_i}(u_i) = \prod p_{X_j}(v_j)$. En revanche nous aurons invariance par changement d'échelle, comme cela sera précisé avec la proposition 4.2.13.

b) Néguentropie

L'entropie différentielle, ou plus simplement l'entropie, d'une variable aléatoire admettant $p_x(u)$ pour densité de probabilité est définie par:

$$S(p_x) \stackrel{\text{def}}{=} -\int p_x(u) \, \log p_x(u) \, du. \tag{4.41}$$

On pourra notamment remarquer que l'information est une différence d'entropies: $I(p_x) = \sum S(p_{x_i}) - S(p_x)$. L'entropie joue un rôle tout à fait particulier en statistiques. En effet, il est possible de montrer qu'il n'existe pas d'autre fonctionnelle satisfaisant quatre axiomes de base, découlant du principe fondamental selon lequel *si un problème peut être résolu de plusieurs façons, alors les solutions obtenues doivent être les mêmes* [188, page 27]. Mais ceci nous éloigne de notre propos. Nous avons introduit pour mémoire l'entropie, mais c'est en réalité la *néguentropie* qui va surtout présenter un intérêt pour notre propos.

Définition 4.2.8 *Soit x un vecteur aléatoire centré admettant $p_x(u)$ pour densité. Notons $\varphi(x)$ la variable aléatoire gaussienne centrée de même covariance que x, et $\Phi_x(u)$ sa densité. Alors la néguentropie associée à p_x est*

$$J(p_x) = \int p_x(u) \, \log \frac{p_x(u)}{\Phi_x(u)} \, du. \tag{4.42}$$

Il est facile de remarquer que la néguentropie est une mesure d'**écart à la distribution gaussienne**, puisqu'elle est égale à la divergence $\delta(p_x, \Phi_x)$. Nous avons par conséquent la propriété:

Proposition 4.2.9 *La néguentropie d'une distribution p_x est toujours positive, et s'annule si et seulement si p_x est presque partout gaussienne.*

De façon encore plus explicite, nous avons:

Proposition 4.2.10 *La néguentropie est la différence des entropies suivantes:*

$$J(p_x) = S(\Phi_x) - S(p_x). \tag{4.43}$$

Démonstration. Par définition de l'entropie, nous avons:

$$\begin{aligned} S(\Phi_x) - S(p_x) &= \int p_x(u) \log p_x(u)\, du - \int p_x(u) \log \Phi_x(u)\, du \\ &\quad + \int p_x(u) \log \Phi_x(u)\, du - \int \Phi_x(u) \log \Phi_x(u)\, du. \end{aligned}$$

D'où on déduit immédiatement $S(\Phi_x) - S(p_x) = J(p_x) + \int \log \Phi_x(u)\, [p_x(u) - \Phi_x(u)]\, du$. Or ce dernier terme est nul puisque, par définition, x et $\varphi(x)$ ont même variance, et $\log \Phi_x(u)$ est un polynôme de degré 2. □

Proposition 4.2.11 *L'entropie et la néguentropie sont invariantes par changement de base orthonormé.*

Démonstration. Considérons deux vecteurs aléatoires, x et $y = Qx$, où Q est une matrice inversible. Alors L'entropie de y s'écrit

$$S(p_x) = -\int p_y(Qu)\ \log[|\det Q|\, p_y(Qu)]\, |detQ|\, du,$$

ce qui donne la règle de transformation de l'entropie par changement de base:

$$S(p_x) = S(p_y) - \log|\det Q|. \tag{4.44}$$

Il est clair que l'entropie est invariante par toute transformation dont le déterminant est de module 1, et en particulier par transformation orthogonale. Par ailleurs, la néguentropie est invariante au moins sur le même ensemble de transformations, d'après la proposition 4.2.10. □

Proposition 4.2.12 *La néguentropie est invariante par changement de base (inversible).*

Démonstration. On applique simplement (4.44), qui est valable pour toute matrice inversible, à p_x et ϕ_x. Par différence, $\log|\det Q|$ disparait et $J(p_x) = J(\phi_x)$. □

Proposition 4.2.13 *L'information mutuelle est invariante par changement d'échelle, et peut s'écrire:*

$$I(p_x) = J(p_x) - \sum_i J(p_{x_i}) + I(\Phi_x). \tag{4.45}$$

Démonstration. Par définition de l'information, nous avons

$$I(p_x) = \sum S(p_{x_i}) - S(p_x). \tag{4.46}$$

Soit Λ une matrice diagonale régulière. Le vecteur Λx a pour entropie $I(p_{\Lambda x}) = \sum S(\Phi_{x_i}) - \log \Lambda_{ii} - S(p_x) + \log\det\Lambda$, ce qui prouve que $I(p_{\Lambda x}) = I(p_x)$. En outre, en utilisant la propriété 4.2.10, la relation (4.46) donne:

$$I(p_x) = \sum S(\Phi_{x_i}) - \sum J(p_{x_i}) - S(\Phi_x) + J(p_x).$$

Enfin, un nouveau recours à $I(\Phi_x) = \sum S(\Phi_{x_i}) - S(\Phi_x)$ permet de conclure. Notons que dans ce résultat, l'information gaussienne peut être remplacée par sa valeur donnée en (4.40). □

Cette dernière propriété met en relief les termes suscepibles d'entrainer une dépendance statistique entre les composantes x_i. Tout d'abord $I(\Phi_x)$, qui est une contribution d'ordre 2, peut être facilement éliminée en standardisant les données. Après standardisation, il ne reste que les deux néguentropies de la formule (4.45), qui sont des termes d'ordre supérieur. Si nous ne voulons pas détruire la décorrélation d'ordre 2, les seules transformations linéaires que l'on a le droit de faire subir à un vecteur aléatoire x sont les transformations diagonales, qui n'ont aucun effet sur l'information comme on l'a montré avec la proposition 4.2.13, et les transformations orthogonales. Or d'après la proposition 4.2.11, la première néguentropie est invariante par transformation orthogonale; il ne reste donc plus que le second terme de (4.45), qui soit susceptible de mesurer la dépendance statistique entre les composantes d'un vecteur standardisé.

Malheureusement, les densités sont en général inconnues, de sorte qu'il faudra approximer les néguentropies par des estimations. Le but de la section d) sera précisément de proposer un moyen pratique de mettre en œuvre cette estimation à partir des moments ou cumulants d'ordre supérieur.

c) Développement en série d'Edgeworth

Soit une variable aléatoire scalaire x de seconde fonction caractéristique $\Psi_x(u)$, supposée être voisine d'une fonction $\Psi_o(u)$. Par définition, $\Psi_x(u)$ génère les cumulants dans son développement en série entière:

$$\Psi_x(u) = \kappa_1 u + \frac{1}{2!}\kappa_2 u^2 + \frac{1}{3!}\kappa_3 u^3 + \ldots, \tag{4.47}$$

où κ_r désigne le cumulant d'ordre r, $\mathcal{C}_{(r)}\{x\}$. Posons λ_r le cumulant d'ordre r dans le développement en série de $\Psi_o(u)$, et $\eta_r = \kappa_r - \lambda_r$. Alors la différence des fonctions caractéristiques s'écrit:

$$\Psi_x(u) - \Psi_o(u) = \sum_{r=1}^{\infty} \frac{1}{r!}\eta_r u^r. \tag{4.48}$$

Notons qu'il n'existe pas forcément de variable aléatoire dont les cumulants d'ordre r sont égaux à η_r. Mais on peut tout de même noter les "moments" μ_r définis par:

$$exp[\sum_{r=1}^{\infty} \frac{1}{r!}\,\eta_r u^r] = \sum_{j=0}^{\infty} \frac{1}{k!}\,\mu_k u^k. \tag{4.49}$$

A partir de cette relation, il est possible de développer $p_x(v)$ autour de $p_o(u)$ comme suit:

$$p_x(v) = p_o(v) \sum_{k=0}^{\infty} \frac{1}{k!}\,\mu_k h_k(v), \tag{4.50}$$

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

$$h_k(v) = \frac{(-1)^k}{p_o(v)}\,\frac{d^k p_o}{dv^k}(v). \tag{4.51}$$

Le développement (4.50) ne revêt une forme simple que pour certaines densités $p_o(v)$ particulières, notamment celles pour lesquelles les fonctions $h_k(v)$ sont des polynomes.

Le développement en série d'Edgeworth de type A permet d'approximer une densité lorsque $p_o(v)$ est gaussienne. Dans un souci de consistence des notations, on notera alors $p_o(v) = \Phi_x(v)$. Pour simplifier les expressions, et sans restreindre la généralité, on se placera dans le cas gaussien standardisé. Dans ce cas, les fonctions $h_k(v)$ sont les polynomes de Hermite définis par

la récurrence:

$$h_0(v) = 1, \tag{4.52}$$
$$h_1(v) = v, \tag{4.53}$$
$$h_{k+1}(v) = v\,h_k(v) - \frac{d}{dv}\,h_k(v). \tag{4.54}$$

Par exemple, $h_2(v) = v^2 - 1$ et $h_3(v) = v^3 - 3v$. En outre, le développement de Edgeworth se distingue de celui de Gram-Charlier par le fait que les termes sont ordonnés non pas par degré croissant, mais par ordre de grandeur décroissant sous les hypothèses du théorème de la limite centrale (page 36). Le classement des termes, s'il n'a aucune importance dans une série infinie convergente, en a beaucoup lorsqu'il s'agit de tronquer la série. Le théorème de la limite centrale nous dit que, si x est la somme de m variables aléatoires indépendantes de cumulants bornés, alors le cumulant d'ordre r de x est de l'ordre de $m^{1-r/2}$. Ceci conduit au classement suivant:

Ordre							
$m^{-1/2}$	κ_3						
m^{-1}	κ_4	κ_3^2					
$m^{-3/2}$	κ_5	$\kappa_3\kappa_4$	κ_3^3				
m^{-2}	κ_6	$\kappa_3\kappa_5$	$\kappa_3^2\kappa_4$	κ_4^2	κ_3^4		
$m^{-5/2}$	κ_7	$\kappa_3\kappa_6$	$\kappa_3^2\kappa_5$	$\kappa_4^2\kappa_3$	κ_3^5	$\kappa_4\kappa_5$	$\kappa_3^3\kappa_4$

Ainsi le développement en série de Edgeworth de la densité $p_x(v)$ autour de $\Phi_x(v)$ s'écrit [136, formule 6.49]:

$$\begin{aligned} p_x(v)/\Phi_x(v) &= 1 \\ &+\frac{1}{3!}\,\kappa_3 h_3(v) \\ &+\frac{1}{4!}\,\kappa_4 h_4(v) + \frac{10}{6!}\,\kappa_3^2 h_6(v) \\ &+\frac{1}{5!}\,\kappa_5 h_5(v) + \frac{35}{7!}\,\kappa_3\kappa_4 h_7(v) + \frac{280}{9!}\kappa_3^3 h_9(v) \\ &+\frac{1}{6!}\,\kappa_6 h_6(v) + \frac{56}{8!}\kappa_3\kappa_5 h_8(v) + \frac{35}{8!}\kappa_4^2 h_8(v) + \frac{2100}{10!}\kappa_3^2\kappa_4 h_{10}(v) \\ &\qquad + \frac{15400}{12!}\kappa_3^4 h_{12}(v) \\ &+O(m^{-2}). \end{aligned} \tag{4.55}$$

d) Approximation de la néguentropie

Dans cette secton, nous allons utiliser le développement de Edgeworth pour approximer la néguentropie que nous avons définie précdemment en (4.42). La relation (4.40) a montré que $I(\Phi_x) = 0$ si et seulement si la matrice de covariance est diagonale. Pour des distributions non gaussiennes, la décorrélation à l'ordre 2 est insuffisante pour assurer l'indépendance. En revanche, la néguentropie sera suffisante pour assurer l'indépendance statistique. En général, à l'instar de la densité de probabilité, la néguentropie est en général inconnue. On se propose ici de l'approximer à l'aide des cumulants d'ordre croissant.

Posons $p_x(u) = \Phi_x(u)[1+f(u)]$, où $f(u)$ est donnée par le développement de Edgeworth. On adopte le développement en série du logarithme suivant: $(1+f)\log(1+f) = f+f^2/2-f^3/6+f^4/12+o(f^4)$. En reportant cette approximation dans l'expression de la néguentropie (4.42), et en replaçant $f(u)$ par sa valeur, on peut obtenir l'approximation escomptée. L'expression finale de la néguentropie nécessite les propriétés intégrales suivantes des polynomes de Hermite:

$$\int \Phi(v)\, h_p(v)\, h_q(v)\, dv = p!\, \delta_{pq}, \tag{4.56}$$

$$\int \Phi(v)\, h_3^2(v)\, h_4(v)\, dv = 3!^3, \tag{4.57}$$

$$\int \Phi(v)\, h_3^2(v)\, h_6(v)\, dv = 6!, \tag{4.58}$$

$$\int \Phi(v)\, h_3^4(v)\, dv = 93\, 3!^2. \tag{4.59}$$

On obtient alors après calcul, si z est une variable aléatoire scalaire standardisée:

$$J(p_z) = \frac{1}{12}\kappa_3^2 + \frac{1}{48}\kappa_4^2 + \frac{7}{48}\kappa_3^4 - \frac{1}{8}\kappa_3^2\kappa_4 + o(m^{-2}). \tag{4.60}$$

4.2.5 Contrastes statistiques

a) Généralités

Les ingrédients introduits dans cette section sont communs aux trois sections qui vont suivre. Le premier ingrédient est la notion de filtre trivial.

Définition 4.2.14 *La suite de matrices $\{A(k)\}$ est dite triviale si et seulement si pour tout indice i fixé, il existe un seul couple d'indices (j,k) tel que: $A_{ij}(k) \neq 0$.*

On admet que le nombre de sources est égal au nombre de capteurs, puisque ceci n'est pas restrictif lorsque $N \geq N_s$ dans le modèle d'observation de départ (4.25), comme on l'a déjà souligné plus haut.

On admet que le processus observé se modélise comme suit:

$$y(t) = \sum_k H(k)\, x(t-k) + v(t), \tag{4.61}$$

où $y(t)$ et $x(t)$ sont de dimension N, et où les matrices $H(k)$ sont cette fois carrées. On désigne par $H(z)$ la transformée en z de la suite $H(k)$. Les autres notations restent celles du modèle (4.25). Il sera en outre nécessaire d'imposer des contraintes supplémentaires pour assurer l'unicité de la modélisation.

Soit $\mathcal{H}$ un sous-ensemble des filtres $H(z)$ de norme L^2 finie, et $\mathcal{P} = \{y(t)\}$ un ensemble de processus de dimension N. Pour alléger les écritures, on adoptera parfois la notation compacte: $H \cdot x \equiv H(z) \cdot x(t)$. De même, on notera $\mathcal{H} \cdot \mathcal{P}$ l'ensemble image de $\mathcal{P}$ par les filtres de $\mathcal{H}$.

Définition 4.2.15 *Une application Υ associant la densité de probabilité d'un élément $y(t) \in \mathcal{H} \cdot \mathcal{P}$ à un nombre réel positif, noté $\Upsilon(y)$, sera dite contraste probabiliste discriminant, ou plus simplement* **contraste** *sur $(\mathcal{P}, \mathcal{H})$, si elle vérifie les trois conditions suivantes:*

C1. Υ est invariante par changement d'échelle; c'est à dire que $\Upsilon(\Lambda y) = \Upsilon(y)$, $\forall y \in \mathcal{H} \cdot \mathcal{P}$, et $\forall \Lambda$, matrice constante diagonale régulière de $\mathcal{H}$.

C2. Si les composantes $x_i(t)$ d'un processus $x(t) \in \mathcal{P}$ sont indépendantes, et chacune blanche au sens fort, alors $\Upsilon(H \cdot x) \leq \Upsilon(x)$, $\forall H(z) \in \mathcal{H}$.

C3. Il y a égalité dans C2 si et seulement si $H(k)$ est triviale. C'est cette condition qui assure le caractère **discriminant** *du contraste.*

Un contraste ne vérifiant pas la propriété C3 sera peu utile car, en l'absence de bruit, $\Upsilon(H \cdot x)$ pourrait atteindre son maximum, $\Upsilon(x)$, sans pour autant que le filtre H soit triviale.

Il faut remarquer que la notation $\Upsilon(y)$ constitue un abus, puisque Υ est construite sur la densité de probabilité de y (éventuellement au sens des distributions). La notation correcte serait $\Upsilon(p_y)$, mais cela alourdirait considérablement les écritures. L'abus est donc admis, mais il faut en être conscient.

Cette définition étend le concept proposé dans *[43] [2]*, tout en assurant la compatibilité avec le concept introduit par Donoho pour la déconvolution scalaire [105] [117], comme nous allons le préciser dans ce qui va suivre.

Il existe des relations d'équivalence entre les couples $(\mathcal{P}, \mathcal{H})$. Par exemple, $\mathcal{P}$ =$\{$*processus temporellement blancs à l'ordre 2 et de variance 1*$\}$ et $\mathcal{H}$ =$\{$*matrices rationnelles*$\}$, peut être remplacé par $\mathcal{P}$ =$\{$*processus à spectre rationnel*$\}$ et $\mathcal{H}$ =$\{$*filtres rationnels* $H(z)$ *tels que* $DiagH(z) = I$, $\forall z\}$.

b) Déconvolution scalaire

Voyons à présent comment les définitions précédentes se particularisent au cas de la déconvolution scalaire.

Corollaire 4.2.16 *Dans le cas scalaire* $N = 1$*, les filtres triviaux sont ceux dont la réponse impulsionnelle est nulle partout sauf en un point; autrement dit, ce sont les retards purs multiples de la période d'échantillonnage, suivis d'un facteurs d'échelle.*

Proposition 4.2.17 *Le module du cumulant standardisé d'ordre* $r > 2$ *à l'origine est un contraste sur l'ensemble* $\mathcal{P}$ *des processus non gaussiens admettant des moments finis jusqu'à l'ordre* r*, et l'ensemble* $\mathcal{H}$ *des filtres non nuls.*

Démonstration. La condition C1 est assurée par la standardisation des cumulants. Par ailleurs, $y(t) = \sum_k H(k)\, x(t-k)$. Grâce à la propriété de multilinéarité des cumulants (3.4.1), nous avons:

$$\mathcal{K}_{(r),y} = \mathcal{K}_{(r),x}\, \frac{\sum_k H(k)^r}{[\sum_k H(k)^2]^{r/2}}\,.$$

Or, par inégalité entre les normes L^p, dès que $m \geq 2$, $[\sum H(k)^r]^{1/r} \leq [\sum H(k)^2]^{1/2}$, ce qui entraine bien $|\mathcal{K}_{(r),y}| \leq |\mathcal{K}_{(r),x}|$, donc C2.

Enfin, l'égalité $|\mathcal{K}_{(r),y}| = |\mathcal{K}_{(r),x}|$ entraine $[\sum H(k)^r]^{1/r} = [\sum H(k)^2]^{1/2}$, ce qui n'est possible que lorsque $H(k)$ ne contient qu'une seule valeur non nulle lorsque $r > 2$. Ceci prouve C3. □

Ces propriétés ont éte prouvées à l'origine par C.W. Granger vers 1976 [105]. Une étude générale rigoureuse des contrastes dans le cas scalaire peut être trouvée dans [117]. Les cumulants standardisés d'ordre 3 ou 4, appelés asymétrie et aplatissement (kurtosis), ont été également utilisés

indépendamment en analyse de données [128]. On voit aussi que la fonctionnelle $\Upsilon(y) = |\mathcal{K}_{(r),y}|^\alpha$ est également un contraste, pour tout $\alpha > 0$ et tout $r > 2$.

Proposition 4.2.18 *L'opposé de l'entropie moyenne de Shannon, $\Upsilon_0(y) = -S(y) = \int \log p_y(u)\, p_y(u)\, du$, est un contraste sur l'ensemble $\mathcal{P}$ des processus non gaussiens de variance finie, et l'ensemble $\mathcal{H}$ des filtres conservant la variance, i.e. , satisfaisant $\sum_k H(k)^2 = 1$.*

Démonstration. Ce résultat est énoncé dans [105] sans preuve détaillée. En réalité, la démonstration est un peu plus pénible qu'on pourrait le croire, et fait appel à la propriété: $S(\sum H(k)\, x(k)) - S(x) \geq \log(\sum H(k)^2)/2$ [84]. Cette propriété est satisfaite dès que les variables $x(k)$ sont i.i.d. Elle n'est pas donnée ici à cause de sa longueur. On se réfèrera à *[63].* □

c) Mélange instantané vectoriel

Si les composantes z_i d'un vecteur aléatoire sont statistiquement indépendantes, alors celles du vecteur $\Lambda P z$ le sont aussi si Λ est une matrice diagonale et P une permutation. Une première exigeance que l'on est en droit d'imposer est donc qu'une fonction de contraste soit insensible à des transformations du type ΛP. En conséquence la matrice F ne peut être définie qu'à cette indétermination près. C'est d'ailleurs la même chose dans le problème de séparation de signaux. On retrouve cette indétermination dans la définition des filtres triviaux:

Corollaire 4.2.19 *Dans le cas de mélanges instantanés, les filtres triviaux sont de la forme $H = \Lambda\, P$, où Λ est diagonale et P est une permutation.*

Les matrices triviales orthogonales sont les "permutations signées". Dans le cas complexe, les matrices unitaires triviales seront appelées de la même façon, sachant qu'elles sont en fait le produit d'une permutation par une matrice diagonale formée d'éléments de module 1.

La définition générale 4.2.15 tient compte de cette indétermination par l'insensibilité au facteur d'échelle. L'insensibilité à la permutation découle de celle de la densité de probabilité. La définition 4.2.15 se particularise dans le cas des mélanges instantanés à la définition suivante:

Corollaire 4.2.20 *Un contraste sur $(\mathcal{P}, \mathcal{H})$ est une application Υ de $\mathcal{P}$ dans $\mathbb{R}$ telle que les trois conditions suivantes soient satisfaites:*

C1. $\Upsilon(\Lambda y) = \Upsilon(y)$, *pour toute matrice diagonale* $\Lambda \in \mathcal{H}$;

C2. Si x *a des composantes indépendantes, alors* $\Upsilon(Ax) \leq \Upsilon(x)$ *pour toute matrice* $A \in \mathcal{H}$.

C3. $\Upsilon(Ax) = \Upsilon(x)$ *n'est vérifiée pour tout* $x \in \mathcal{P}$ *de composantes indépendantes que si* A *est de la forme* ΛP.

Sauf mention contraire, $\mathcal{H}$ sera, dans le contexte des mélanges instantanés, l'ensemble des matrices carrées inversibles. On va donner dans la suite quatre exemples de contrastes.

Proposition 4.2.21 *L'application* $\Upsilon_o(z) \stackrel{\text{def}}{=} -I(p_{\tilde{z}})$, *où* $\tilde{z}$ *est le vecteur standardisé associé à* z, *conformément à la définition 3.4.3, est un contraste sur l'ensemble* $\mathcal{P}$ *des vecteurs aléatoires de covariance finie et inversible. En outre, il est discriminant sur le sous-ensemble des vecteurs aléatoires ayant au plus une composante gaussienne.*

Démonstration. La démonstration découle directement des propriétés 4.2.6 page 76 et 4.2.13 page 80 *[2]*. □

Comme nous l'avons déjà souligné, l'information mutuelle est en général difficilement utilisable dans la pratique car les densités sont inconnues, même si des tentatives ont été faites dans ce sens [173]. Il est donc utile de se tourner vers des contrastes plus "pratiques".

Proposition 4.2.22 *L'application* $\Upsilon_{2,r}(z) \stackrel{\text{def}}{=} \sum_{i=1}^{N} \mathcal{K}^2_{(r),y_i}$, *est un contraste sur* $(\mathcal{P}, \mathcal{H})$, *où* $\mathcal{P}$ *désigne le sous-ensemble des vecteurs aléatoires ayant des moments finis jusqu'à l'ordre* r, *pour* $r > 2$, *et ayant au plus un cumulant marginal d'ordre* r *nul.* $\mathcal{H}$ *désigne l'ensemble des matrices carrées inversibles.*

Démonstration. Dans cette démonstration, il est légitime de poser $H = L\,Q$, où Q est orthogonale, et $\tilde{y}(t) = Q\,x(t)$. En effet, puisque $\Upsilon_{2,r}$ est construit sur les cumulants standardisés de $y(t)$, nous savons que $\Upsilon_{2,r}(y) = \Upsilon_{2,r}(\tilde{y}) = \Upsilon_{2,r}(Q\,x)$. Tout se passe donc comme si $\mathcal{H}$ était l'ensemble des matrices orthogonales.

Condition C1. La condition C1 découle de la standardisation.

Condition C2. On note $\Upsilon(x) = \sum_p \mathcal{K}(r), {x_p}^2$, en omettant provisoirement l'indexage $(_{2,r})$ pour alléger, et:

$$\Omega(y) = \sum_{i_1 \dots i_r} \mathcal{K}^2_{i_1 \dots i_r, y}.$$

Alors $\Upsilon(y) \leq \Omega(y)$ puisque tous les termes sont positifs. Par ailleurs, $\Omega(y)$ s'écrit, par multilinéarité des cumulants:

$$\Omega(y) = \sum_{pq} \sum_{i_1 \dots i_r} Q_{i_1 p} \, Q_{i_1 q} \cdots Q_{i_r p} \, Q_{i_r q} \, \mathcal{K}_{(r),x_p} \, \mathcal{K}_{(r),x_q}.$$

Or, comme $Q^T Q = I$ par hypothèse, il vient que $\Omega(y) = \Upsilon(x)$. En conclusion, nous avons:

$$\Upsilon_{2,r}(y) \leq \Omega(y) = \Upsilon_{2,r}(x). \tag{4.62}$$

Cette relation, qui prouve C2, nous sera utile pour la suite de la démonstration.

Condition C3, première démonstration. C3 a été prouvée dans le cas général par Comon *[42] [2]*. Toutefois, une démonstration plus simple a été donnée par M. Krob dans le cas $r = 4$ [139]. Le principe de cette démonstration reste valable lorsque r est multiple de 4, comme nous allons l'expliquer maintenant.

Posons $r = 4s$, s entier. Notons $\Upsilon(\bar{y})$ et $\Omega(\bar{y})$ les quantités obtenues en remplaçant les cumulants $\mathcal{K}_{(r),x_p}$ par leur module dans les expressions développées de $\Upsilon(y)$ et $\Omega(y)$, respectivement. Alors, par le même raisonnement que pour obtenir (4.62), on obtiendrait:

$$\Upsilon_{2,r}(\bar{y}) \leq \Omega(\bar{y}) = \Upsilon_{2,r}(x).$$

Mais on peut observer que, par inégalité triangulaire, $\Upsilon(y) \leq \Upsilon(\bar{y})$. Par conséquent, si on a l'égalité $\Upsilon(y) = \Upsilon(x)$, on a nécessairement égalité de tous ces termes entre eux, et en particulier $\Upsilon(\bar{y}) = \Omega(\bar{y})$. Ceci se traduit par les égalités:

$$\sum_p Q_{i_1 p} \cdots Q_{i_r p} \, |\mathcal{K}_{(r),x_p}| = 0, \forall (i_1 \cdots i_r) \neq (i_1 \cdots i_1). \tag{4.63}$$

En particulier, si le r-uplet $(i_1 \cdots i_r)$ ne contient que deux indices distincts, i et j, en nombre égal (ce qui est possible car r est pair):

$$\sum_p Q^{2s}_{ip} Q^{2s}_{jp} \, |\mathcal{K}_{(r),x_p}| = 0, \forall (i,j), i \neq j.$$

Pour tout p tel que $\mathcal{K}_{(r),x_p} \neq 0$, ceci entraine la nullité du produit $Q_{ip}\,Q_{jp}$, puisque tous les termes sont positifs dans la somme. Si le vecteur x a au plus un cumulant d'ordre r non nul, alors le produit $Q_{ip}\,Q_{jp}$ est nul pour $N-1$ valeurs de p. En conséquence, $N-1$ colonnes de Q ne contiennent qu'un seul élement non nul. Comme Q est orthogonale, ses lignes sont normées, et elle est nécessairement une permutation signée.

Condition C3, seconde démonstration. Lorsque r n'est plus un multiple de 4, la démonstation précédente s'effondre. La démonstration qui suit reproduit celle donnée dans *[42] [2]*. On utilise le lemme suivant:

Lemme 4.2.23 *Soit ${}^rQ$ la matrice dont les éléments sont $|Q_{ij}|^r$. Alors, si Q est orthogonale, ${}^2Q$ vérifie: $||{}^2Q\,u|| \leq ||u||$, $\forall u$, pour la norme L^2.*

Démonstration du lemme. Notons $\bar{Q}$ (resp $\bar{u}$) la matrice (resp. le vecteur) dont les éléments sont les modules de ceux de Q (resp. ceux de u). Si Q est unitaire, alors ${}^2Q$ est bistochastique, c'est à dire que la somme de ses composantes au sein d'une même ligne ou d'une même colonne vaut 1. D'après un théorème de Birkhoff, l'ensemble des matrices bistochastiques est un polyèdre convexe dont les sommets sont des permutations; ceci veut dire que:

$$ {}^2Q = \sum_s \alpha_s\,P_s, \alpha_s \geq 0, \sum_s \alpha_s = 1. $$

Par inégalité triangulaire, il vient finalement:

$$ ||{}^2Q\,u|| \leq ||{}^2\bar{Q}\,\bar{u}|| \leq \sum_s \alpha_s ||P_s\,u|| = ||u||. $$

Démonstration de C3. Comme Q est unitaire, ses composantes sont de module plus petit que 1, et $|{}^rQ_{ij}| \leq |{}^2Q_{ij}|$ dès que $r \leq 2$. Par inégalité triangulaire, on en tire que, pour tout vecteur u:

$$ \sum_{i,j,k} Q^r_{ki}Q^r_{kj}\,u_i\,u_j \leq \sum_{i,j,k} \bar{Q}^r_{ki}\bar{Q}^r_{kj}\,\bar{u}_i\,\bar{u}_j \leq \sum_{i,j,k} \bar{Q}^2_{ki}\bar{Q}^2_{kj}\,\bar{u}_i\,\bar{u}_j. $$

En appliquant le lemme, on obtient que:

$$ ||{}^rQ\,u||^2 \leq ||{}^2\bar{Q}\,\bar{u}||^2 \leq ||\bar{u}||^2 = ||u||^2. $$

Appliquons ce résultat au vecteur u formé des cumulants marginaux d'ordre r de la variable standardisée $\tilde{y}$. On obtient alors simplement:

$$ \Upsilon(Q\,\tilde{y}) \leq \Upsilon(\tilde{y}), $$

qui reste valable pour le vecteur y. Ceci prouve C2 en passant. Si on a égalité, cela veut dire que:

$$||^2\bar{Q}\,\bar{u}||^2 - ||^r\bar{Q}\,\bar{u}||^2 = 0,$$

pour un certain vecteur u ayant au plus une composante nulle. Comme elle est toujours positive ou nulle, la quantité suivante doit donc être nulle:

$$[^2\bar{Q}\,\bar{u}]_i^2 - [^r\bar{Q}\,\bar{u}]_i^2 = 0, \forall i.$$

Cette égalité reste vraie en enlevant les carrés, à cause le la positivité des termes. Tous les termes étant de nouveau positifs, on doit avoir

$$[^2\bar{Q}_{ij} - {}^r\bar{Q}_{ij}]\bar{u}_j = 0, \forall (i,j).$$

La chute est similaire à celle de la première démonstration. On remarque que, comme $\bar{u}_j$ est strictement positif pour au moins $N-1$ valeurs de j, l'égalité ${}^2\bar{Q}_{ij} = {}^r\bar{Q}_{ij}$ pour $r > 2$ entraine que $N-1$ colonnes de la matrice Q ne contiennent qu'un seul élément non nul. L'orthogonalité de Q implique enfin que Q est une permutation signée.

La démonstration s'étend au cas complexe lorsque l'ordre r est pair, et que les cumulants sont définis de façon à ce que la moitié des termes soit conjugués *[2]*. □

Un autre contraste a été proposé récemment par Moreau et Macchi [163] [164] [165] et semble particulièrement intéressant. En effet, une analyse précise de leur démonstration montre que le fait que Q soit orthogonale est incomplètement utilisé, et qu'une condition bien plus faible suffit, comme nous allons le voir plus loin.

Proposition 4.2.24 *L'application* $\Upsilon_{1,r}(z) \stackrel{\text{def}}{=} \sum_{i=1}^{N} |\mathcal{K}_{(r),y_i}|$, *est un contraste sur* $(\mathcal{P}, \mathcal{H})$, *où* $\mathcal{P}$ *est l'ensemble des vecteurs aléatoires ayant des moment finis jusqu'à l'ordre* r, *pour* $r > 2$, *et ayant au plus un cumulant marginal d'ordre* r *nul.*

Démonstration. Commençons par donner la démonstration lorsque Q est orthogonale, calquée sur celle de E. Moreau. La condition C1 ne pose pas de problème grâce à la standardisation des cumulants. Pour démontrer C2, remarquons que:

$$\Upsilon_1(y) = \sum_j \left| \sum_p Q_{jp}^r \mathcal{K}_{(r),x_p} \right| \leq \sum_{j,p} |Q_{jp}|^r \cdot |\mathcal{K}_{(r),x_p}|.$$

Or, comme $\sum_j |Q_{jp}|^2 = 1$ et $r \geq 2$, alors $\sum_j |Q_{jp}|^r \leq 1$, il vient:

$$\Upsilon_1(y) \leq \sum_p |\mathcal{K}_{(r),x_p}| = \Upsilon_1(x),$$

qui prouve C2.

Pour prouver C3, supposons que $\Upsilon_1(y) = \Upsilon_1(x)$. Alors:

$$\sum_p \left[1 - \sum_j |Q_{jp}|^r \right] \cdot |\mathcal{K}_{(r),x_p}| = 0.$$

Comme tous les termes sont positifs, on en déduit que $\sum_j |Q_{jp}|^r = 1$ pour tout p tel que $\mathcal{K}_{(r),x_p} \neq 0$. Donc la $p$ième colonne de Q ne contient qu'un élément non nul. Enfin, comme Q est orthogonale, si $N-1$ de ses colonnes n'ont qu'un élément non nul, elle est une permutation signée. □

On remarquera que, jusqu'à la dernière étape de la démonstration, seules la propriété de normalisation des colonnes a été utilisée pour prouver C3. Ceci nous conduit à considérer le contraste suivant:

Proposition 4.2.25 *L'application définie par* $\Upsilon(y) = \sum_i |\mathcal{C}_{(r),y_i}|$ *est un contraste sur* $(\mathcal{P}, \mathcal{H})$*, si* $\mathcal{P}$ *désigne l'ensemble des vecteurs aléatoires de dimension* N *ayant des moments finis jusqu'à l'ordre* r *et ayant au plus un cumulant marginal d'ordre* r *nul, et si* $\mathcal{H}$ *désigne l'ensemble des matrices satisfaisant les relations suivantes:*

Y2. $\mathrm{Diag}\, H\, H^T = I$ *(chaque ligne est normée),*

F2. $\mathrm{Diag}\, H^T\, H = I$ *(chaque colonne est normée).*

Démonstration. La preuve est la même que celle de la proposition 4.2.24. Les différences sont les suivantes: Y2 est utilisée pour prouver la condition C1, F2 est utilisée pour C2, et F2 et Y2 sont nécessaires pour C3, sans nécessiter l'orthogonalité de Q. □

On peut se demander si ces contraintes sont légitimes. Nous savons que le modèle d'observation $y = Hx$ souffre d'une indétermination à une matrice multiplicative près, de la forme ΛP. Il est donc utile (et ici nécessaire) d'imposer N contraintes permettant de lever cette indétermination. En standardisant les données, on avait réduit cette indétermination à une permutation signée, mais en même temps on se débarrassait des corrélations d'ordre 2, ce qui n'était pas indispensable.

Pourtant, il peut être désagréable d'être obligé de recourir à cette procédure, notamment en présence de bruit gaussien. Ceci peut se comprendre en remarquant que la standardisation devrait faire intervenir une racine de la matrice de covariance du signal seul, et non de celle de l'observation.

Une contrainte possible est: $Diag\, H = I$. Mais cette contrainte manque de souplesse dans le contexte qui nous occupe. Une autre contrainte consiste à imposer une variance unité sur chacune des sorties. C'est précisément ce que traduit la contrainte Y2. Comme nous l'avons précisé page 64, la contrainte F2 peut être vue comme une normalisation des colonnes de H.

Enfin, on ne manquera pas de remarquer que F2 et Y2 réunies sont des contraintes bien plus faibles que l'orthogonalité. Il est facile de le vérifier en remarquant que si H est orthogonale, alors $H^T H = H\, H^T = I$, tandis que F2 et Y2 ne correspondent qu'aux diagonales de ces deux dernières égalités. Par exemple si $N = 2$, les rotations hyperboliques satisfont F2 et Y2.

Le contraste 4.2.25 ne fait plus intervenir de cumulants standardisés, et peut donc potentiellement prétendre à une insensibilité au bruit gaussien. Un autre constraste proposé par E. Moreau avait également cette faculté, mais faisait intervenir le module d'un cumulant croisé [164].

d) Mélanges convolutifs vectoriels

L'objectif de cette section est assez modeste. Il consiste à présenter une ébauche de projet de recherche orienté vers la mise au point d'algorithmes de déconvolution multivariable autodidacte en présence de bruit, éventuellement non gaussien.

L'intérêt des fonctions de contraste a déjà été souligné et il n'est pas nécessaire d'y revenir. En revanche, la définition 4.2.15 donnée en page 84 ne présenterait pas beaucoup d'intérêt s'il n'était pas possible d'exhiber au moins une fonction de contraste applicable dans cette situation. C'est ce qui va être fait maintenant.

Proposition 4.2.26 *La fonction* $\Upsilon_1(y) = \sum_j |\mathcal{C}_{(r),y_j}|$ *est un contraste sur* $(\mathcal{P}, \mathcal{H})$*, si* $\mathcal{P}$ *est l'ensemble des processus stationnaires à l'ordre* r*, ayant au plus une composante gaussienne, et si* $\mathcal{H}$ *est l'ensemble des filtres* $H(z)$ *vérifiant les conditions suivantes, notées F3 et Y4 page 64:*

$$\text{[F3.]}\ Diag \oint H(z)\, H(z)^\dagger\, dz \;=\; I, \tag{4.64}$$

$$\text{[Y4.]}\ Diag \oint H(z)^\dagger\, H(z)\, dz \;=\; I. \tag{4.65}$$

Par exemple, si le filtre $H(z)$ préserve l'énergie sur chaque composante, et si l'entrée $x(t)$ est blanche au second ordre, alors la normalisation F3 sera satisfaite (cf. page 64).

Démonstration. La condition C1 demandée dans la définition 4.2.15 est assurée par F3.

Pour la condition C2, on remarque que le contraste se développe comme suit:

$$\Upsilon_1(y) = \sum_i \left| \sum_{p,k} H_{jp}^r(k)\, \mathcal{C}_{(r),x_p} \right| \leq \sum_{i,p,k} |H_{ip}(k)|^r \cdot |\mathcal{C}_{(r),x_p}|.$$

Mais Y4 implique que $\sum_{i,k} |H_{ip}(k)|^2 = 1$, $\forall i$. Donc pour $r \geq 2$, on a bien:

$$\Upsilon_1(y) \leq \sum_p |\mathcal{C}_{(r),x_p}| = \Upsilon_1(x).$$

Pour prouver C3, supposons que $\Upsilon_1(y) = \Upsilon_1(x)$. Alors:

$$\sum_p \left[1 - \sum_{i,k} |H_{ip}(k)|^r \right] \cdot |\mathcal{C}_{(r),x_p}| = 0.$$

Donc pour tout p tel que $\mathcal{C}_{(r),x_p} \neq 0$, il vient que:

$$\sum_{ik} |H_{ip}(k)|^r = 1.$$

On rencontre alors une complication qui n'apparaissait pas dans le cas instantané. En effet, on voudrait que la somme sur i et p soit égale à 1 pour pouvoir conclure. Pour ce faire, on remarque simplement que la dernière égalité entraine que $\sum_{i,p,k} |H_{ip}(k)|^r = N$. Mais comme $\sum_{p,k} |H_{ip}(k)|^r \leq 1$ d'après F3, on doit avoir aussi:

$$\sum_{pk} |H_{ip}(k)|^r = 1, \forall i.$$

Ceci se constate immédiatement, par exemple par l'absurde. On en déduit finalement que $\{H(k)\}$ est triviale, dès que $r > 2$, en utilisant une nouvelle fois F3. □

La "complication" est également surmontable dans le cas de la fonctionnelle Υ_2, dans des conditions similaires. Ceci est heureux car Υ_2 présente l'intérêt d'être différentiable, contrairement à Υ_1, qui fait intervenir une valeur absolue. L'absence de différentiabilité est un obstacle supplémentaire à une mise en œuvre en ligne.

4.2.6 Un algorithme pour l'ACI

Dans cette section, on présente un algorithme hors ligne, proposé pour calculer l'ACI (solution dans le cas des mélanges instantanés). On pourra trouver diverses solutions en ligne dans [163], par exemple.

a) Approche en deux étapes

Nous avons vu avec la proposition 4.2.13 que l'information mutuelle s'écrivait:

$$I(p_x) = J(p_x) - \sum_{i=1}^{N} J(p_{x_i}) + I(\Phi_x).$$

L'approche a priori la plus précise consisterait à maximiser cette information mutuelle, éventuellement en remplaçant les néguentropies par des approximations. Afin d'éviter à avoir à résoudre un problème d'optimisation multimodal de grande dimension, on propose dans cette section une approche en deux étapes.

Si x est un vecteur aléatoire standardisé, il est clair que la composante $I(\Phi_x)$ est nulle. La nullité de ce terme est préservée par transformation orthogonale. On peut donc décomposer la matrice cherchée F en deux facteurs: $F = QL^-$, où L^- assure la standardisation, et où Q est orthogonale. Or, d'après (4.2.11), la quantité $J(p_x)$ est invariante par transformation orthogonale. Il ne reste donc que le deuxième terme, $\sum_i J(p_{x_i})$, à maximiser dans l'ensemble des matrices orthogonales.

D'après l'équation (4.60), nous voyons que le critère Υ_o peut être approximé par Υ_3 si les cumulants d'ordre 3 ne sont pas nuls. S'ils sont nuls, alors Υ_o peut être approximé par Υ_4. Malheureusement, les conditions dans lesquelles l'expression

$$\Upsilon_{3,4} \stackrel{\text{def}}{=} \sum_{i=1}^{N} 4\mathcal{K}_{3,i}^2 + \mathcal{K}_{4,i}^2 + 7\mathcal{K}_{3,i}^4 - 6\mathcal{K}_{3,i}^2\mathcal{K}_{4,i} \tag{4.66}$$

pourrait être un contraste discriminant n'ont pas été obtenues à ce jour. On utilisera donc ce dernier critère avec prudence.

b) Algorithme Contraste–Maximisation (CM)

Dans cette section, on s'intéressera à la recherche de la matrice Q maximisant le contraste $\Upsilon_r(p_z)$, $r \in \{3, 4\}$, $z = Q\tilde{y}$. La matrice Q a $N(N-1)/2$ paramètres libres, si elle est de dimension N. Il n'est pas aisé de mener une

telle optimisation, même s'il est possible de calculer la différentielle de Υ_r, et d'obtenir analytiquement l'équation des valeurs stationnaires *[36]*. On peut par contre décomposer la matrice Q en un produit de $N(N-1)/2$ rotations planes de la forme:

$$Q_{(i,j)} = \frac{1}{\sqrt{1+\theta^2}} \begin{pmatrix} 1 & \theta \\ -\theta & 1 \end{pmatrix}, \tag{4.67}$$

où θ désigne la tangente de l'angle de la rotation plane $Q_{(i,j)}$, opérant dans le plan défini par les i^{ieme} et j^{ieme} coordonnées. Cette décomposition n'est pas unique.

La recherche d'une rotation plane maximisant le contraste $\Upsilon_r(p_z),\ r \in \{3,4\}$ se fait de façon purement analytique en résolvant des polynômes de degré inférieur ou égal à 4 (c'est donc possible par radicaux). L'ensemble de l'algorithme CM est décrit ci-dessous.

Définition 4.2.27 Algorithme CM

1. *Calculer la SVD de la matrice de données: $Y = VSU^\dagger$, et construire la matrice de données standardisées $Z = \sqrt{T}\,U^\dagger$ et la matrice de passage $L = VS/\sqrt{T}$. Si Y est $N \times T$ et de rang ρ, alors Z est $\rho \times T$ et L est $N \times \rho$.*
2. *Initialiser $F \leftarrow L$.*
3. *Commencer la boucle sur les balayages: $k = 1, 2, \ldots, kmax$; On fixe $kmax \leq 1 + \sqrt{\rho}$.*
4. *balayer les $\rho(\rho-1)/2$ paires (i,j), conformément à schéma de balayage fixé (par exemple cyclique par lignes). Pour chaque paire, faire:*
   (a) *Estimer les $r+1$ cumulants d'ordre r des lignes i et j de la matrice Z.*
   (b) *Calculer l'angle α maximisant le contraste Υ_r, dans l'intervalle:*
   $$]-\pi/4,\ \pi/4].$$
   (c) *Accumuler la matrice de passage: $F \leftarrow F\,Q^\dagger_{(i,j)}$.*
   (d) *Mettre à jour la matrice de données: $Z \leftarrow Q_{(i,j)}\,Z$.*
5. *Arrêter la boucle si $k = kmax$ ou bien si tous les angles estimés dans le dernier balayage sont petits devant $1/T$.*

6. *Calculer la norme des colonnes de F:* $\Delta_{ii} = ||F_{i:}||$.
7. *Ordonner les composantes de Δ par ordre décroissant:* $\Delta \leftarrow P\,\Delta\,P^\dagger$ *et changer l'ordre des colonne de F en conséquence:* $F \leftarrow F\,P^\dagger$.
8. *Normaliser la matrice F par:* $F \leftarrow F\,\Delta^{-1}$.
9. *Fixer la phase (signe) de chaque colonne de F de façon à ce que l'élément de plus grand module soit réel positif:* $F \leftarrow F\,D$.

Les quatre dernières étapes de l'algorithme décrit ci-dessus sont facultatives, dans le sens où elles ne servent qu'à déterminer de façon unique un représentant de la classe d'équivalence des solutions *[21] [2]*. L'algorithme CM est sous-optimal à deux niveaux: d'abord, il n'utilise pas les néguentropies pour identifier la meilleure standardisation, puisque seuls les moments d'ordre 2 sont utilisés dans cette étape. Ensuite, l'optimisation sur Q ne se fait pas non plus globalement mais est décomposée en plusieurs problèmes d'optimisation monodimensionnels, un peu à la manière d'une relaxation. Ce n'est pas exactement une relaxation, car la décomposition se fait dans un groupe et non dans un espace vectoriel. L'algorithme ci-dessus est décrit comme si les données étaient complexes, mais pour alléger, on s'est limité ci-après à la description du cas réel.

c) Obtention de la rotation plane dans l'algorithme CM

L'objectif de cette section est de décrire en détail comment l'angle α d'une rotation plane peut être choisi de façon à maximiser les fonctions de contrastes Υ_3 ou Υ_4. Cette description correspond à l'étape 4b de l'algorithme CM défini en 4.2.27.

Supposons que la matrice de données standardisées, notée ici Y, soit de dimension $2 \times T$, et que nous cherchions la rotation Q maximisant la fonctions de contraste $\Upsilon_r(p_z)$. Nous avons:

$$Q = \frac{1}{\sqrt{1+\theta^2}} \begin{pmatrix} 1 & \theta \\ -\theta & 1 \end{pmatrix}, \text{ et } Z = QY. \tag{4.68}$$

La fonction de contraste $\Upsilon_r(p_z)$, ne dépendant que des cumulants marginaux d'ordre r de Z, par construction, est fonction implicite de θ et des cumulants d'ordre r de Y, compte tenu de la propriété de multilinéarité des cumulants. Utilisons κ pour désigner les cumulants de Z et g ceux de Y. On peut donc convenir de noter –abusivement– cette fonction $\Upsilon_r(\theta; g)$. A priori cette

fonction est une fraction rationelle de θ, ce qui explique qu'il soit faisable de trouver son maximum *absolu* en un nombre réduit d'opérations élémentaires. La valeur de l'angle α se déduit directement de celle de sa tangente θ; il suffit de prendre par exemple l'angle se trouvant dans l'intervalle $]-\pi/2, \pi/2]$, mais cela n'est pas nécessaire. En effet, cette indétermination n'affecte la matrice Q que par multiplication par une matrice de la forme ΛP. Voyons donc comment obtenir la tangente θ.

d) Maximisation de Υ_3

Si $r = 3$, nous avons l'expression suivante pour le contraste:

$$\Upsilon_3(\theta; g) = (\theta + \frac{1}{\theta})^{-3} \sum_{i=1}^{3} a_i \left(\theta^i - (-\theta)^{-i}\right), \tag{4.69}$$

où les coefficients a_i sont donnés par

$$\begin{aligned} a_3 &= g_{111}^2 + g_{222}^2, && (4.70)\\ a_2 &= 6\,(g_{122}\,g_{222} - g_{111}\,g_{112}), && (4.71)\\ a_1 &= 9\,(g_{122}^2 + g_{112}^2) + 6\,(g_{112}\,g_{222} + g_{111}\,g_{122}). && (4.72) \end{aligned}$$

Les points stationnaires correspondant à $d\Upsilon_3/d\theta = 0$ sont les racines du polynome:

$$\omega_3(\xi; g) = d_2\,\xi^2 + d_1\,\xi - 4\,d_2, \text{ si } d_2 \neq 0, \tag{4.73}$$

où nous avons utilisé la variable auxiliaire $\xi = \theta - 1/\theta$, et où les coefficients d_i sont donnés par:

$$\begin{aligned} d_2 &= a_2/6 \;=\; g_{122}\,g_{222} - g_{111}\,g_{112}, && (4.74)\\ d_1 &= a_1/3 - a_3. && (4.75) \end{aligned}$$

Il suffit donc de calculer toutes les racines réelles de $\omega_3(\xi; g)$, et de calculer ensuite pour chacune d'elles la solution θ correspondante grâce à la relation $\theta^2 - \xi\theta - 1 = 0$, qui n'admet toujours qu'une seule racine dans l'intervalle $]-1, 1]$. Enfin, s'il y a plus d'une racine réelle, on sélectionnera celle des solutions donnant la plus grande valeur du contraste.

e) Maximisation de Υ_4

Si $r = 4$, la procédure est similaire. On préfèrera exprimer le contraste en fonction de la variable auxiliaire $\xi = \theta - 1/\theta$ dès le départ, de sorte que:

$$\Upsilon_4(\xi; g) = (\xi^2 + 4)^{-2} \sum_{i=0}^{4} b_i \, \xi^i. \tag{4.76}$$

De même, les points stationnaires de $\Upsilon_4(\xi; g)$ sont donnés par les racines d'un polynome en ξ:

$$\omega_4(\xi; g) = \sum_{i=0}^{4} c_i \, \xi^i. \tag{4.77}$$

Les valeurs des coefficients b_i et c_i en fonction des cumulants des observations sont données dans [2]. Ces racines sont faciles à calculer puisque le polynome n'est que de degré 4. Comme précédemment, après avoir calculé les racines ξ_p de $\omega_4(\xi; g)$, il suffit de les reporter dans l'expression de $\Upsilon_4(\xi; g)$ pour avoir celle correspondant au maximum absolu. La valeur de θ correspondante s'obtient en calculant la racine de $\theta^2 - \xi\theta - 1 = 0$ se trouvant dans l'intervalle $]-1, 1]$.

4.3 Décompositions tensorielles

Les moments et les cumulants de variables aléatoires sont des objets tensoriels, comme l'a bien expliqué Mc Cullagh dans son livre [160]. Pourtant, il est bien rare qu'ils soient considérés comme tels aux ordres supérieurs à 2.

Evidemment à l'ordre 2, les outils d'algèbre linéaire étant bien rodés depuis des années, on a pris l'habitude de ranger les moments dans une matrice de covariance. Mais il n'y a en principe aucune raison qu'on ne fasse pas de même pour les SOE, si ce n'est précisément à cause de l'absence d'*outils spécifiquement tensoriels*. L'objet de cette section est d'expliquer pourquoi ces outils ne sont pas disponibles aujourd'hui.

Il y a au mons deux raisons à cela. La première est que ces outils sont difficiles à mettre au point, comme il va être bientôt démontré avec l'exemple de la diagonalisation. La seconde est qu'il n'y a sans doute jamais eu véritablement de demande dans ce sens. Les tenseurs étaient utilisés essentiellement en physique, et n'étaient d'ailleurs pas symétriques. En statistiques, les moyens limités en puissance de calcul interdisaient l'utilisation pratique des SOE multivariables pour des problèmes concrets.

La situation a aujourd'hui changé, et il apparait utile de développer quelques outils supplémentaires pour le traitement du signal de demain.

4.3.1 Diagonalisation tensorielle

Considérons un tableau réel T à d indices i_k, $1 \leq k \leq d$, variant chacun dans $\{1, .., n\}$. On suppose que T est complètement symétrique, c'est à dire que l'ordre des indices ne change rien à la valeur de T; par exemple, $T_{123} = T_{231}$. On dira que d est l'ordre de T et n sa dimension.

Si $d = 2$, T est une matrice symétrique, et on sait qu'elle est diagonalisable par transformation congruente:

$$T = A\,\Lambda\,A^{\dagger}. \tag{4.78}$$

Cette décomposition n'est pas unique. Un théorème établi par Sylvester nous apprend d'ailleurs que la signature (le nombre de signes + et −) de la matrice diagonale Λ est un invariant parmi les solutions. Un de ces choix est celui de la décomposition spectrale (*i.e.* en éléments propres), où on impose à la matrice A d'être orthogonale.

La question que l'on se pose est de savoir si une telle décomposition peut s'étendre au cas de tables à plus de deux indices. Décrivons le problème dans le cas de trois indices, pour simplifier. La diagonalisation congruente s'écrirait:

$$T_{ijk} = \sum_{p=1}^{r} A_{ip}\,A_{jp}\,A_{kp}\,\Lambda_{pp}. \tag{4.79}$$

Dans cette décomposition, le nombre r joue le même rôle que celui du rang pour les matrices. Cependant, rien ne nous dit que $r \leq n$, malheureusement *[24] [34]*.

A ce stade, plusieurs familles de problèmes peuvent être identifiés. Dans la première, on ne considère que les décompositions pour lesquelles $r = n$. Cette décomposition ne sera alors exacte que dans des cas très particuliers *[35]*, ce qui apparaitra plus clairement après les explications de la section 4.3.3. Quoiqu'il en soit, on voit déjà le lien étroit existant entre de telles diagonalisations et le problème de l'ACI décrit au chapitre 4.2 si la table T désigne le tenseur cumulant d'ordre d des observations.

Dans les problèmes de la seconde famille, on considère le cas générique, c'est à dire le cas le plus souvent rencontré. Pour les matrices par exemple, le cas générique est celui du rang plein. Pour les tenseurs, le rang générique n'est pas le rang maximal possible; ceci est une première particularité (cf.

section 4.3.3). S'il s'agit de tenseurs d'ordre $d > 2$, alors la seconde famille est très différente de la première.

Dans la troisième famille, on s'intéresse aux cas non génériques. On trouve notamment le cas du rang maximal, mais aussi les cas intermédiaires entre $r = n$ et r générique. L'intérêt de considérer ces cas intermédiaires est d'ordre pratique: on espère être capable de proposer des algorithmes pour calculer la diagonalisation congruente dans ces cas-là, ou au moins prouver qu'ils existent théoriquement, ce qui n'est pas aujourd'hui possible pour toutes les valeurs du triplet (p, n, r), loin s'en faut.

4.3.2 Polynômes homogènes

On ne trouve pratiquement rien dans la littérature sur la diagonalisation de tables à plus de deux indices. Pourtant, il y avait au 19ème siècle une grosse activité de recherche sur les polynômes homogènes, qui a malheureusement été étouffée en grande partie à cause de Hilbert. En effet, ce dernier a tué un des plus gros débouchés pour ces chercheurs qu'était le théorie des invariants [170].

On peut très brièvement expliquer pourquoi les tenseurs et les polynômes homogènes sont liés. On trouvera plus de détails dans *[24]* si nécessaire. Tout tenseur symétrique G de dimension n et d'ordre d est associé de façon unique à un polynôme homogène p de degré d à n variables par la relation:

$$p(x_{i_1}, x_{i_2}, .., x_{i_d}) = \sum_{i_1, i_2, .. i_d = 1}^{n} G_{i_1 i_2 .. i_d} \, x_{i_1} \, x_{i_2} .. x_{i_d} \tag{4.80}$$

Il est évident qu'avec cet éclairage, la diagonalisation de G est équivalente à la décomposition de p en puissances de formes linéaires. Dans le cas des matrices par exemple, on sait que toute matrice symétrique est associée à une forme quadratique, et vice-versa.

C'est grâce à cette connexion qu'il a été possible de remonter aux travaux de Salmon, puis de Rota et de Reznick *[24]*. Ce sont ces travaux qui nous ont permis d'établir ce qui va être décrit maintenant dans la section 4.3.3.

4.3.3 Rang générique et nombre de solutions

A partir d'un très grand nombre de relations, rassemblées pour la plupart par Reznick et Rota, il a été possible de dresser la table 4.3. Cette dernière donne la valeur du rang générique r pour des valeurs quelconques du couple (d, n). On constatera que dans la première colonne, on a bien $r = n$, puisque

ordre dimension	2	3	4	5	6	7	8
2	2	2	3	3	4	4	5
3	3	4	6	7	10	12	15
4	4	5	10	14	22	30	42
5	5	8	15	26	42	66	99
6	6	10	22	42	77	132	215
7	7	12	30	66	132	246	429
8	8	15	42	99	215	429	805

Table 4.3: *Rang générique des tenseurs en fonction de leur dimension et de leur ordre.*

cette dernière correspond au cas matriciel $d = 2$. Il est important de noter qu'il n'y a aucune règle systématique connue pour calculer r; certaines de ces valeurs n'ont été obtenues que récemment, après plusieurs articles visant à les encadrer par des bornes.

En outre, le nombre de solutions est infini si $rn > D$, si on pose

$$D = \begin{pmatrix} n+d-1 \\ d \end{pmatrix}. \tag{4.81}$$

La dimension de la variété des solutions est donnée par la différence $nr - D$, qui est reproduite dans le tableau 4.4. On vérifie bien que dans le cas matriciel, la dimension est $n(n-1)/2$. On constate aussi que pour certains couples (n, d), il n'existe qu'un nombre fini de solutions; citons notamment les couples $(4, 3)$, $(7, 3)$, ou $(7, 4)$.

Mis à part les cas où $r \leq n$ et $d = 2$, il n'a été possible pour l'instant de diagonaliser un tenseur symétrique que pour $n = 2$, ce qui présente un intérêt très limité *[24]*. Ceci fait l'objet de recherches en cours.

ordre dimension	2	3	4	5	6	7	8
2	1	0	1	0	1	0	1
3	3	2	3	0	2	0	0
4	6	0	5	0	4	0	3
5	10	5	5	4	0	0	0
6	15	4	6	0	0	0	3
7	21	0	0	0	0	6	0
8	28	0	6	0	4	0	5

Table 4.4: *Dimension de l'ensemble des solutions.*

Chapitre 5

Orientations et perspectives

Dans ce chapitre on décrit les perspectives de travaux telles qu'elles peuvent être envisagées aujourd'hui en reprenant la partition en quatre volets proposée aux chapitres 1 et 2. Je saurai gré au lecteur intéressé par l'un de ces sujets de recherche de bien vouloir m'en avertir, pour mettre en place une coopération éventuelle, qui sera la bienvenue.

Certains des sujets décrits ci-après trouvent des applications dans le monde industriel, notamment sonar; cependant, ces applications ne sont pas précisées pour des raisons de confidentialité.

Traitement d'antenne

1 Il est non seulement discutable d'imposer l'hypothèse de stationnarité des sources dans les problèmes de mélanges linéaires, au regard des conditions expérimentales, mais aussi très certainement superflu sur le plan théorique. Il est même vraisemblable que la non stationnarité des sources permette d'atteindre, lorsque le milieu est constant, de meilleures performances.

2 Le problème de la calibration automatique sans bruiteur coopérant, révélant des problèmes d'observabilité complexes, est d'un grand intérêt opérationnel [148]. Il fait l'objet d'une étude en 1995-96.

3 Celui de la pondération optimale d'antennes de géométrie quelconque est posé encore aujourd'hui sous la forme d'un problème d'optimisation multimodal général, malgré ses nombreuses particularités.

4 L'évaluation de bornes permettant d'accéder aux performances ultimes atteignables en estimation de paramètres est également difficile à met-

tre en œuvre lorsque le modèle gaussien est irréaliste. Ceci pose le problème du traitement d'antenne ultime (lorsque les statistiques des observations et des sources sont parfaitement connues), si on cherche à construire un traitement basé sur les SOE, en complément aux statistiques d'ordre 2.

5 Le calcul de performances de certaines chaines de traitement complexes (très non linéaires) est difficile à mener à bien dans des conditions réalistes, c'est à dire sur signaux réels, en particulier lorsque les données sont peu nombreuses. Les techniques de rééchantillonnage (Bootstrap) sont bien adaptées à ce calcul.

6 Nous avons vu que la déconvolution des mélanges linéaires multivariables est possible avec le recours uniquement aux moments d'ordre 2, dans certaines circonstances. En particulier, lorsque les observations sont suréchantillonnées. D'un autre côté l'aspect cyclostationnaire a été peu utilisé pour cette catégorie de problèmes, et conduit sans doute aux mêmes conclusions.

7 Les modèles bilinéaires occupent une place toute particulière parmi les processus non linéaires. En effet, bien qu'ayant une structure plus simple, ils permettent d'approcher tout type de processus. On peut les utiliser notamment pour extraire un signal d'un bruit multiplicatif.

8 L'ACI peut être vue comme une décomposition d'un vecteur aléatoire sur une base adéquate, dans laquelle ses composantes sont statistiquement indépendantes. Elle est donc applicable aux signaux aléatoires, et conduirait à une décomposition sur une base de fonctions déterministes, à l'instar de la décomposition de Karhunen-Loeve (KL). On peut se demander si cette nouvelle démarche n'aurait pas bien plus de sens physique que la simple orthogonalité de la décomposition de KL.

9 La particularité de certains mélanges linéaires a déjà été mentionnée, et notamment ceux ne comportant que des retards purs et des amortissements. Evidemment, si les retards sont multiples de la période d'échantillonnage, le problème semble simple, mais n'a pas grand intérêt. En revanche lorsque ce n'est pas le cas, les filtres sont non causaux et de réponse impulsionnelle infinie, ce qui n'est pas très plaisant.

Cependant, le fait que les réponses aient une forme très particulière doit pouvoir être pris en compte. Un problème intéressant, mais difficile à

poser correctement, est celui d'un seul capteur et d'une seule source, en présence de trajets multiples. C'est un des points à aborder dans la thèse de B. Emile.

10 Lorsque le nombre de sources est supérieur au nombre de capteurs, bien peu de méthodes sont applicables. On a vu que si le mélange est composé de retards purs, l'identification de la fonction de transfert est *théoriquement* possible, quel que soit le nombre K de capteurs et le nombre P de sources. Elle conduit donc à une estimation asymptotiquement clairvoyante (les sources ne peuvent pas être parfaitement estimées, mais leur vecteur directionnel peut l'être). B. Emile se penche sur de tels mélanges dans sa thèse.

Cependant, les conditions sous lesquelles l'identification d'un mélange (convolutif quelconque) de P sources avec K capteurs est possible ne sont aujourd'hui pas clairement établies. Dans le cas de mélanges instantanés, on avance notamment la condition suffisante que $P < (3K-2)/2$. Mais aucun algorithme n'est disponible aujourd'hui, qui soit capable de réaliser cette identification. Ceci fait l'objet d'une coopération avec l'université de Leuven (KUL).

11 On sait qu'un mélange instantané de 2 sources sur 2 capteurs est identifiable avec un nombre fini (et très réduit) d'opérations. Néanmoins, ce résultat n'est vrai qu'en l'absence de bruit si les variables aléatoires sont complexes (même circulaires). Il est pourtant vraisemblable que la maximisation d'un contraste soit possible analytiquement dans ce cas, avec le recours à la résolution de polynômes de degré inférieur ou égal à quatre (elle l'est dans le cas de variables aléatoires réelles, en présence de bruit).

12 La fonction d'ambiguité a été définie aux ordres élevés par Porat. On peut se demander s'il ne serait pas possible de la définir tous ordres confondus (*e.g.* par une divergence de Kullback).

13 La formation de voies est une mesure d'inhomogéneité du champ basée sur les moments d'ordre 2. Rien n'empêche de construire un outil mesurant les inhomogéneités en se basant sur les variations de la fonction de répartition avec la direction. On devrait gagner en pouvoir de détection si les sources sont non gaussiennes.

Statistiques d'ordre élevé

14 Il est important de donner une préférence aux algorithmes de déconvolution de mélanges linéaires qui fonctionnent en bande large. Elles ont plus d'attrait lorsque le signal est large-bande et de faible niveau. Une formulation dans le domaine temps est notamment souhaitable.

Les fonctions de contrastes dans le domaine temps répondent à ce besoin. Toutefois, la définition des fonctions de contrastes dans le cas des mélanges convolutifs vectoriels donnée dans la section 4.2.5 aura sans doute besoin d'être parachevée. En outre, d'autres exemples de fonctions de contraste devront être exhibés, et des algorithmes de résolution mis au point. Je termine par exemple la mise au point de la factorisation "doublement diagonale", normalisant à la fois les lignes et les colonnes d'une matrice. Par ailleurs, il serait utile de comparer mon contraste basé sur l'ordre 4 seulement avec la solution préconisée par Inouye.

15 La définition des contrastes introduite dans la section 4.2.5 fait intervenir un couple $(\mathcal{H}, \mathcal{P})$ d'ensembles (Filtre, Processus). On conçoit que plus ces ensembles sont grands, plus le contraste considéré est "puissant". Il est donc possible de dresser une sorte de hiérarchie dans les fonctions de contrastes.

16 Dans le chapitre 3, on a souligné le fait que les conditions d'existence des multispectres n'ont jamais été clairement établies. Cette lacune fondamentale reste à combler.

17 Le test de normalité décrit en détail dans la section 4.1 doit être expérimenté sur un plus grand nombre de signaux-test (Monte carlo et données rélles), et comparé avec les tests de Giannakis-Tsatsanis et de Moulines et al. Il est aussi possible de le comparer avec un test de type TIU (intersection-union), revenant dans le cas présent à pré-déconvoluer, ce qui pourrait se faire avec l'algorithme de Shalvi-Weinstein par exemple.

18 Un multispectre, en tant que fonction de plusieurs variables (nécessairement discrétisées), peut être vu comme un tenseur, de même que le spectre d'ordre 2 peut être vu comme une matrice. La factorisation d'un multispectre revient à la diagonalisation de ce tenseur. Lorsque le tenseur est de rang 1, le processus considéré est linéaire.

Le lien entre factorisation multispectrale, diagonalisation tensorielle, et tests de rang, peut avoir un intérêt dont la nature n'est pas uniquement théorique. La thèse de D. Rossille [182] l'a bien mis en évidence.

Algorithmique numérique Dans ce paragraphe sont inclus les problèmes d'algèbre linéaire théoriques ainsi que les aspects plus appliqués d'algorithmique.

19 Dans le troisième volet, je propose la poursuite des travaux sur l'analyse de la stabilité des algorithmes rapides de résolution de systèmes de faible rang de déplacement. En outre, la mise au point d'algorithmes rapides de résolution de systèmes structurés singuliers au sens des moindres carrés reste un problème ouvert.

20 La caractérisation des éléments propres des matrices structurées (Töplitz, produit de Töplitz...) est un problème lié au précédent d'une certaine manière, bien que ce dernier ne nécessite que la caractérisation du noyau. L'état de l'art dans ce domaine avance assez irrégulièrement.

21 Sur un autre plan, comme nous l'avons vu, l'ACI peut être vue comme la diagonalisation approchée d'un tenseur symétrique d'ordre supérieur à deux. Nous avons pu caractériser l'ensemble des tenseurs symétriques linéairement diagonalisables. Il serait intéressant d'étendre ces résultats dans un premier temps au cas complexe (avec le problème de la définition de la symétrie pour les ordres impairs), et dans un second temps aux cas non génériques.

22 La factorisation (*e.g.* diagonalisation par transformation congruente) des tenseurs de rang plus grand que leur dimension dépasse le cadre de l'ACI. La recherche d'algorithmes spécifiques aux tenseurs est un domaine qui a été –curieusement– très peu exploré. Une coopération avec l'université KUL et avec l'INRIA est en cours.

23 J'ai mis en évidence l'intérêt que peut présenter l'Analyse en sous-espaces Indépendants (ASI), et tout spécialement pour la classification. Un sujet de recherche consiste à concevoir un algorithme numérique performant de calcul de l'ASI.

24 Un vieux sujet auquel je souhaite m'atteler depuis des années est celui de la révision des algorithmes numériques d'algèbre linéaire lorsque les matrices sont aléatoires. Par exemple, la factorisation QR est utilisée

pour résoudre des systèmes linéaires, à des fins de prédiction en traitement du signal. Mais la stratégie de pivotage ne tient pas compte de la variance des éléments de la matrice.

Apprentissage Le quatrième et dernier axe concerne la théorie et la mise en œuvre de l'apprentissage, supervisé ou non.

25 Les estimateurs à noyau de probabilité, dont les avantages ne sont plus à vanter, souffrent d'une limitation pratique: on ne sait toujours pas évaluer le paramètre de largeur minimisant le biais ou l'erreur quadratique. On peut se limiter au cas des noyaux à largeur fixe, car le passage au noyau variable a été fait correctement par Abramson.

26 Le problème de l'apprentissage non supervisé peut être vu sous l'angle de l'identification de lois mélange. Sans vouloir adresser le problème de la détermination du nombre de modes qui est mal posé, on peut se pencher sur l'identification des paramètres du modèle en supposant que le nombre de modes est connu. L'ajustement des moments conduit à la résolution d'un système polynomial.

27 L'identification d'un filtre linéaire à partir des ses entrées et sorties peut sans doute être abordée sans modèle paramétrique (tel que les modèles de Volterra). On propose la construction d'une transformation non linéaire (par exemple réseau de neurones avec temps de retard) de façon incrémentale sans rétropropagation. Il en résulterait un gain immense en temps de calcul dans l'apprentissage (comparer par exemple à l'apprentissage supervisé des réseaux stratifiés comme le PMC). Plus précisément, il est reconnu que les théorèmes théoriques de représentation (*e.g.* Sprecher) ne sont pas applicables dans la pratique.

28 Comme cela a été expliqué au chapitre 2, il est des cas où la dimension des vecteurs de la base de données est trop grande; la notion de petite taille ou de grande dimension est ici comprise conformément à la définition de *[29] [65]*. Ce sujet de recherche consiste en l'application de l'algorithme d'ASI à des bases de données, décrit au point 23, afin d'évaluer son intérêt sur des problèmes réels. On peut penser à des applications dans le domaine de la reconnaissance de la parole par exemple. D'autres applications peuvent être envisagées lorsqu'elles demandent le découplage de variables. Des coopérations sont en cours avec les universités INPG et KUL.

29 Si, après avoir procédé aux réductions de dimension classiques (*e.g.* PCA), et aux partitions de l'espace (*i.e.* ASI), la dimension reste trop élevée, on peut malgré tout être convaincu que l'apprentissage est possible dans des conditions raisonnables. Ceci n'est pas contradictoire, puisque la condition sur la dimension (donnée page 17) est une condition suffisante non nécessaire de garantie de performances.

Si cette intuition est bien fondée, alors elle signifie que les données sont cantonnées au voisinage d'une variété non linéaire de dimension plus faible. On peut s'en convaincre en calculant par exemple la dimension fractale moyenne des données. Le problème consiste ensuite à identifier, par voie homotopique, la surface en question, et à "projeter" les données sur un espace vectoriel de même dimension. J'envisage volontiers une coopération avec J. Hérault de l'INPG sur ce sujet.

30 Dans un certain nombre de situations pratiques, il peut être intéressant de bâtir un classifieur à K classes avec des briques simples; l'intérêt peut tout simplement provenir d'une question de rentabilité de production en grand nombre. Je baptise ce problème "classification modulaire". Le premier exemple est celui où ces briques sont des classifieurs binaires. Le second, déjà étudié en communications, est celui de la classification à ressources distribuées avec contraintes de communications entre processeurs. Le troisième exemple, dont j'ai déjà parlé page 16, est celui de la classification distribuée avec contraintes de mémoire totale.

La classification modulaire nécessite bien sûr la mise au point d'algorithmes, mais aussi soulève des questions d'ordre plus théoriques relevant de l'optimalité de l'approche (critère d'optimisation, bornes ultimes sur les performances indépendamment de tout algorithme...). Une coopération est en cours avec l'EPFL sur un des aspects de ce problème.

Chapitre 6

Bibliographie

6.1 Publications personnelles

Les publications sont regroupées par type, puis listées par ordre rétro-chronologique dans chacun des types. Les articles de conférence parus dans des proceedings réédités sous forme de livre sont classés dans la rubrique "conférences avec actes".

6.1.1 Articles parus dans des revues internationales ou dans des ouvrages édités en langue anglaise

[1] P. COMON, "Structured matrices and inverses", in *Linear Algebra for Signal Processing*, A. Bojanczyk, G. Cybenko, Eds., vol. 69 of *IMA Volumes in Mathematics and its Applications*, pp. 1–16. Springer Verlag, 1995.

[2] P. COMON, "Independent Component Analysis, a new concept ?", *Signal Processing, Elsevier*, vol. 36, no. 3, pp. 287–314, Apr. 1994, Special issue on Higher-Order Statistics.

[3] E. CHAUMETTE, P. COMON, D. MULLER, "An ICA-based technique for radiating sources estimation; application to airport surveillance", *IEE Proceedings - Part F*, vol. 140, no. 6, pp. 395–401, Dec. 1993, Special issue on Applications of High-Order Statistics.

[4] P. COMON, P. LAURENT, "Displacement rank of generalized inverses of persymmetric matrices", *SIAM Journal on Matrix Analysis*, vol. 14, no. 3, pp. 646–654, July 1993.

[5] P. COMON, "MA identification using fourth order cumulants", *Signal Processing, Eurasip*, vol. 26, no. 3, pp. 381–388, 1992.

[6] P. COMON, C. JUTTEN, J. HERAULT, "Separation of sources, part II: Problems statement", *Signal Processing*, vol. 24, no. 1, pp. 11–20, July 1991.

[7] P. COMON, G. H. GOLUB, "Tracking of a few extreme singular values and vectors in signal processing", *Proceedings of the IEEE*, vol. 78, no. 8, pp. 1327–1343, Aug. 1990, Published from Stanford report 78NA-89-01, feb 1989.

[8] P. COMON, Y. ROBERT, D. TRYSTRAM, "Systolic implementation of the adaptive solution to normal equations", *Computer Vision, Graphics and Image Processing*, vol. 52, pp. 402–408, 1990.

[9] J. Y. BLANC, P. COMON, D. TRYSTRAM, "Using preconditioned conjugate gradient for solving consecutive linear systems", *Communications in Applied Numerical Methods*, vol. 6, no. 3, pp. 231–240, 1990.

[10] P. COMON, D. T. PHAM, "Estimation of the order of a FIR filter for noise cancellation", *IEEE Trans. on Inf. Theory*, vol. 36, no. 2, pp. 429–434, Mar. 1990.

[11] P. COMON, D. T. PHAM, "An error bound for a noise canceller", *IEEE Trans. on ASSP*, vol. 37, no. 10, pp. 1513–1517, Oct. 1989.

[12] P. COMON, L. KOPP, "Comments on a real-time high resolution technique for angle of arrival estimation", *Proceedings of the IEEE*, vol. 77, no. 3, pp. 492–494, Mar. 1989.

[13] P. COMON, D. TRYSTRAM, "An incomplete factorization algorithm for adaptive filtering", *Signal Processing*, vol. 13, pp. 353–360, Dec. 1987.

[14] P. COMON, Y. ROBERT, "A systolic array for computing B A-1", *IEEE Trans. on ASSP*, vol. 35, no. 6, pp. 717–723, 1987.

[15] P. COMON, J. L. LACOUME, "Noise reduction for an estimated Wiener filter using noise references", *IEEE Trans. on Information Theory*, vol. 32, no. 2, pp. 310–313, Mar. 1986.

6.1.2 Articles parus dans des revues en langue française

[16] E. KAZAMARANDE, P. COMON, "Performances numériques de l'algorithme de Levinson", *RAIRO Mathematical Modeling and Numerical Analysis*, vol. 29, no. 2, pp. 123–170, June 1995.

[17] P. COMON, "Circularité et signaux aléatoires à temps discret", *Traitement du Signal*, vol. 11, no. 5, pp. 417–420, Dec 1994.

[18] P. COMON, "Classification supervisée par réseaux multicouches", *Traitement du Signal*, vol. 8, no. 6, pp. 387–407, dec 1991.

[19] P. COMON, "Performances de la régression linéaire dans le cas gaussien", *Traitement du Signal*, vol. 8, no. 4, pp. 281–282, 1991.

[20] P. COMON, "Classification bayésienne distribuée", *Revue Technique Thomson CSF*, vol. 22, no. 4, pp. 543–561, 1990.

[21] P. COMON, "Analyse en Composantes Indépendantes et identification aveugle", *Traitement du Signal*, vol. 7, no. 3, pp. 435–450, Dec. 1990, Numero special non lineaire et non gaussien.

[22] P. COMON, D. TRYSTRAM, Y. ROBERT, "Implementation systolique de systemes adaptatifs", *Traitement du Signal*, vol. 4, no. 4, pp. 73–85, 1987.

[23] P. COMON, "Estimation multivariable complexe", *Traitement du Signal*, vol. 3, no. 2, pp. 97–101, 1986.

6.1.3 Articles soumis à des revues avec comité de lecture

[24] P. COMON, B. MOURRAIN, "Decomposition of quantics in sums of powers of linear forms", *Signal Processing*, Feb. 1995, submitted, special issue on High-Order Statistics, Giannakis and Swami editors.

[25] P. COMON, G. BIENVENU, "Ultimate performance of QEM classifiers", *IEEE Trans. Neural Networks*, May 1995, submitted.

[26] B. EMILE, P. COMON, "Estimation of time delays between colored sources", *IEEE Trans. on Signal Processing*, 1995, submitted.

6.1.4 Conférences avec actes

[27] B. EMILE, P. COMON, "Estimation de temps de retard entre signaux colorés", in *XVieme Colloque Gretsi*, Juan les Pins, 18–22 Sept 1995.

[28] P. COMON, Y. CHENEVAL, "Supervised classification with variable kernel estimators", in *IWANN*, Mira Cabestany Prieto, Ed., Malaga, Spain, June 7-11 1995, pp. 1099–1106, Springer-Verlag, Lecture Notes in Computer Sciences.

[29] P. COMON, "Supervised classification, a probabilistic approach", in *ESANN-European Symposium on Artificial Neural Networks*, Verleysen, Ed., Brussels, Apr 19-21 1995, pp. 111–128, D facto Publ., invited paper.

[30] P. COMON, L. DERUAZ, "Normality tests for coloured samples", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp 217–221.

[31] B. EMILE, P. COMON, J. LE ROUX, "Estimation of time delays between wide-band sources", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp 111-115.

[32] P. COMON, B. EMILE, "Estimation of time delays in the blind mixture problem", in *EUSIPCO 94*, Edinburgh, Scotland, September 13-16 1994, pp. 482–485.

[33] P. COMON, J. L. VOZ, M. VERLEYSEN, "Estimation of performance bounds in supervised classification", in *ESANN-European Symposium on Artificial Neural Networks*, M. Verleysen, Ed., 45 rue Masui, B-1210 Brussels, Belgium, April 20-22 1994, pp. 37–42, D facto Publ.

[34] P. COMON, B. MOURRAIN, "Decomposition of quantics in sums of powers", in *SPIE conference on Advanced Signal Processing V*, San Diego, July 24–29 1994, pp. 93–104.

[35] P. COMON, "Tensor diagonalization, a useful tool in signal processing", in *IFAC-SYSID, 10th IFAC Symposium on System Identification*, M. Blanke, T. Soderstrom, Eds., Copenhagen, Denmark, July 4-6 1994, vol. 1, pp. 77–82, invited session.

[36] P. COMON, "Remarques sur la diagonalisation tensorielle par la methode de Jacobi", in *XIVeme Colloque Gretsi*, 13-16 Septembre 1993, pp. 125–128.

[37] P. COMON, G. BIENVENU, T. LEFEBVRE, "Supervised design of optimal receivers", in *Acoustic Signal Processing for Ocean Exploration Processing and Ocean Exploration*, J. M. F. Moura, I. M. G. Lourtie, Eds. 1993, pp. 547–552, Kluwer Academic Publishers, Proceedings of

the NATO Advanced Study Institute on Acoustic Signal Processing and Ocean Exploration, July 26-Aug. 7, 1992, Madeira, Portugal.

[38] E. CHAUMETTE, P. COMON, D. MULLER, "Application of ICA to airport surveillance", in *IEEE Signal Processing Workshop on High-Order Statistics*, South Lake Tahoe, California, June 7-9 1993, pp. 210–214.

[39] C. JUTTEN, P. COMON, "Neural Bayesian classifier", in *IWANN*, A. Prito J. Mira, J. Cabestany, Ed., Stiges, Spain, June 9–11 1993, pp. 119–124, Springer Verlag.

[40] P. COMON, "Independent component analysis, and the diagonalization of symmetric tensors", in *European Conference on Circuit Theory and Design ECCTD*, H. Dedieu, Ed., Davos, Aug 30-Sept 3 1993, pp. 185–190, Elsevier, invited session.

[41] P. COMON, "Displacement rank of pseudo-inverses", in *International Conference on Acoustics, Speech and Signal Processing - ICASSP*, Mar. 23-26 1992, vol. V, pp. 49–52.

[42] P. COMON, "Blind identification in presence of noise", in *Proc. European Signal Processing Conf. EUSIPCO*, Brussels, Aug 24 - 27 1992, pp. 835–838.

[43] P. COMON, "Independent component analysis", in *Proc. Int. Sig. Proc. Workshop on Higher-Order Statistics*, Chamrousse, France, July 10-12 1991, pp. 111–120, Republished in *Higher-Order Statistics*, J.L.Lacoume ed., Elsevier, 1992, pp 29–38.

[44] P. COMON, G. BIENVENU, "Detection et estimation supervisees", in *XVieme Colloque Gretsi*, Juan les Pins, 16–20 Sept 1991, pp. 277–280.

[45] J. F. CARDOSO, P. COMON, "Tensor-based independent component analysis", in *Proc. European Signal Processing Conf. EUSIPCO*, Barcelona, Spain, September 18-21 1990, pp. 673–676.

[46] P. COMON, "High-order separation, application to detection and localization", in *Proc. European Signal Processing Conf. EUSIPCO*, Barcelona, Spain, September 18-21 1990, pp. 277–280.

[47] P. COMON, J. F. CARDOSO, "Eigenvalue decomposition of a cumulant tensor with applications", in *SPIE Conference on Advanced Signal Processing Algorithms*, San Diego, California, July 10-12 1990, pp. 361–372, Architectures and Implementations, vol.1348.

[48] P. COMON, "Separation of stochastic processes", in *Proc. Workshop on Higher-Order Spectral Analysis*, Vail, Colorado, June 28-30 1989, IEEE-ONR-NSF, pp. 174–179.

[49] P. COMON, "Separation of sources using high-order cumulants", in *SPIE Conference on Advanced Algorithms and Architectures for Signal Processing*, San Diego, California, August 8-10 1989, pp. 170–181, vol. Real-time signal processing XII.

[50] P. COMON, "Separation de melanges de signaux", in *XII Colloque Gretsi*, Juan les Pins, 12 -16 juin 1989, pp. 137–140.

[51] P. COMON, "Statistical approach to the Jutten-Herault algorithm", in *NATO Workshop on Neuro-Computing*, Les Arcs, France, Feb. 27-March 3 1989, Republished in: Neurocomputing, Algorithms, Architectures and Applications, F.Fogelman and J. Herault editors, NATO ASI series, Springer Verlag, 1990, pp81–88.

[52] P. COMON, "Fast updating of a low-rank approximate to a varying hermitian matrix", in *22nd Asilomar Conference*, Pacific Grove, Nov. 2-4 1988, pp. 358–362.

[53] P. COMON, "Fast computation of a restricted subset of eigenpairs of a varying hermitian matrix", in *NATO ASI on Num. Linear Algebra, Digital Sig.Proc. and Parallel Algorithms*, Leuven, Belgium, Aug. 1988, Republished in: Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms, G.H. Golub and P. VanDooren editors, Springer Verlag, NATO ASI series vol. F70, 1991, pp457–466.

[54] P. COMON, "Adaptive computation of a few extreme eigenpairs of a positive definite hermitian matrix", in *European Signal Processing Conference EUSIPCO*, Grenoble, France, Sept. 5-8 1988, pp. 647–650.

[55] P. COMON, T. KAILATH, "An array processing technique using the first principal component", in *First International Workshop on SVD and Signal Processing*, Sept. 1987, Extended version published in: SVD and Signal Processing, E.F. Deprettere editor, North Holland, 1988, 301–316.

[56] E. MOISAN, P. COMON, "Ponderations variables pour les filtres en treillis adaptatifs", in *XIe Colloque GRETSI*, Nice, 1-5 juin 1987, pp. 309–312.

[57] P. COMON, J. L. LACOUME, "About Capon estimator optimality", in *Third Workshop on Spectrum Estimation and Modeling*, Boston, Nov. 17-18 1986.

[58] P. COMON, J. L. LACOUME, "Signal estimation using a reception model", in *International Symposium EUSIPCO*, The Hague, Netherlands, Sept. 2-5 1986.

[59] P. COMON, J. L. LACOUME, "A robust adaptive filter for noise reduction problems", in *International Conference on Acoustics, Speech and Signal Processing - ICASSP*, Tokyo, Japan, Apr. 7-11 1986, pp. 2599–2602.

[60] P. COMON, F. PLANSON, "Ground response to electromagnetic natural excitation", in *IASTED International Symposium*, Paris, June 19-21 1985, vol. Applied Signal Processing, pp. 271–274.

[61] P. COMON, G. LEJEUNE, "Extrapolation de signaux lacunaires", in *IXeme Colloque Gretsi*, Nice, 16-20 mai 1983, pp. 199–203.

6.1.5 Livres

[62] L. KOPP, P. COMON, J. P. LECADRE, *Traitement d'antenne Sonar*, livre en préparation.

[63] J. L. LACOUME, P. COMON, P. O. AMBLARD, *Statistiques d'ordre élevé en traitement du signal*, livre en préparation.

6.1.6 Autres: Brevets, Conférences sans actes, notes de cours

[64] P. COMON, B. EMILE, "Estimation de temps de retard à l'aide de cumulants", in *Journée signal de cergy*, ENSEA, Cergy, 2 fev 1995, pp. 18–20.

[65] C. JUTTEN et al., *Enhanced learning for evolutive neural architectures*, Louvain la Neuve, April 1995.

[66] P. COMON, *Procédé et Dispositif d'Estimation Aveugle de Retards Différentiels*, 1993, Brevet enregistré en aout 1994 pour Thomson-Sintra, no 59-358V(X 5991).

[67] P. COMON, C. JUTTEN, *Neural classifiers, Courses notes*, Neuro-Nimes, Oct. 1993.

[68] P. COMON, J. L. LACOUME, "Statistiques d'ordres supérieurs pour le traitement du signal", Ecole Predoctorale de Physique, Les Houches, 30 aout – 10 septembre 1993, P. Flandrin et J. L. Lacoume ed.

[69] P. COMON, "Wavefields separation: Neural networks versus batch methods", in *EAPG Workshop on Multichannel Filtering of Seismic Data*, Paris, June 1992, Abstracts only.

[70] P. COMON, "Structured matrices and their inverses", in *IMA Workshop on Linear Algebra for Signal Processing*, Minneapolis, Apr. 6-10 1992.

[71] P. COMON, "ATHOS, qu'est-ce que cela évoque pour vous ?", *Traitement du Signal*, vol.10, no. 1, 1993, Editorial.

[72] P. COMON, "Distributed detection and estimation", in *ESPRIT BRA Workshop on Neural Networks and Artificial Vision*, Chamrousse, France, Jan. 29-30 1991, Proceedings of extended abstracts.

[73] P. COMON, *Method and Device for Real-time Signals Separation*, 1989, Patent registrated for Thomson-Sintra, January 1990, no 9000436. International extension confirmed on March, 1992.

[74] P. COMON, L. KOPP, *Traitement du signal Sonar, Notes de cours*, ESSI, 1989-1994.

En outre, une quinzaine de rapports ont été rédigés en relation avec les contrats de recherche, et ne sont pas mentionnés ici.

6.2 Autres références bibliographiques

[75] K. ABEDMERAIM, P. LOUBATON, E. MOULINES, "Subspace method for blind identification of multichannel FIR filters in noise field with unknown spatial covariance", in *Asilomar conference*, Asilomar, California, 1994.

[76] P. O. AMBLARD, J. M. BROSSIER, N. CHARKANI, "New adaptive estimation of the fourth-order cumulant...", in *Proc. EUSIPCO*, Edinburgh, Sept. 1994, pp. 466–469.

[77] D. F. ANDREWS, R. GNANADESIKAN, J. L. WARNER, "Methods for assessing multivariate normality", in *Multivariate Analysis III*, P. R. Krishnaiah, Ed., pp. 95–116. Academic Press, 1973.

[78] F. J. ANSCOMBE, W. J. GLYNN, "Distribution of the kurtosis statistic b2 for normal samples", *Biometrika*, vol. 70, no. 1, pp. 227–234, 1983.

[79] Y. BAR-NESS, J. W. CARLIN, M. L. STEINBERGER, "Bootstrapping adaptive interference cancelers: Some practical limitations", in *Proc. The Globecom. Conference*, Miami, Nov. 1982, pp. 1251–1255, paper No F3.7.

[80] S. BELLINI, F. ROCCA, "Asymptotically efficient blind deconvolution", *Signal Processing, Elsevier*, vol. 20, pp. 193–209, 1990.

[81] A. BELOUCHRANI, K ABEDMERAIM, J. F. CARDOSO, E. MOULINES, "Second-order blind separation of correlated sources", in *Proc. Int. Conf. Digital Signal Processing*, Cyprus, 1993, pp. 346–351.

[82] A. BENVENISTE, M. GOURSAT, "Blind equalizers", *IEEE Trans. Communications*, vol. 32, no. 8, pp. 871–883, Aug. 1984.

[83] A. BENVENISTE, M. GOURSAT, G. RUGET, "Robust identification of a non-minimum phase system", *IEEE Trans. Auto. Control*, vol. 25, no. 3, pp. 385–399, June 1980.

[84] R. E. BLAHUT, *Principles and Practice of Information Theory*, Addison-Wesley, 1987.

[85] A. BLANC-LAPIERRE, R. FORTET, *Theorie des Fonctions Aleatoires*, Masson, 1953.

[86] A. BLANC-LAPIERRE, B. PICINBONO, *Fonctions aleatoires*, Masson, 1981.

[87] C. BOURAIN, P. BONDON, "Efficiency of high-order moment estimates", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp. 186–190.

[88] K. O. BOWMAN, L. R. SHENTON, "Omnibus contours for departures from normality based on b1 and b2", *Biometrika*, vol. 62, pp. 243–250, 1975.

[89] D. R. BRILLINGER, *Time Series, Data Analysis and Theory*, Holden-Day, 1981.

[90] D. De BRUCQ, *Theorie du Signal*, Masson, 1988.

[91] J. A. CADZOW, O. M. SOLOMON, "Algebraic approach to system identification", *IEEE Trans. ASSP*, vol. 34, pp. 462–469, 1986.

[92] V. CAPDEVIELLE, C. SERVIERE, J. L. LACOUME, "Separation of wide band sources", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp. 66–70.

[93] J. F. CARDOSO, "Localisation et identification par la quadricovariance", *Traitement du Signal*, vol. 7, no. 5, pp. 397–406, Dec. 1990.

[94] J. F. CARDOSO, "On the performance of source separation algorithms", in *Proc. EUSIPCO*, Edinburgh, Sept. 1994, pp. 776–779.

[95] J. F. CARDOSO, S. BOSE, B. FRIEDLANDER, "Output cumulant matching for source separation", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp. 44–48.

[96] J. F. CARDOSO, A. SOULOUMIAC, "Blind beamforming for non-Gaussian signals", *IEE Proceedings - Part F*, vol. 140, no. 6, pp. 362–370, Dec. 1993, Special issue on Applications of High-Order Statistics.

[97] J. F. CARDOSO, A. SOULOUMIAC, "An efficient technique for blind separation of complex sources", in *Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA*, 1993, pp. 275–279.

[98] P. CHEVALIER, "On the performance of higher order blind sources separation methods", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp. 30–34.

[99] W. J. CONOVER, *Practical NonParametric Statistics*, Wiley, 1980.

[100] S. CSÖRGÖ, "Testing for normality in arbitrary dimension", *The Annals of Statistics*, vol. 14, no. 2, pp. 708–723, 1986.

[101] R. D'AGOSTINO, "An omnibus test of normality for moderate and large size samples", *Biometrika*, vol. 58, no. 2, pp. 341–348, 1971.

[102] R. D'AGOSTINO, E. S. PEARSON, "Tests for departure from normality. empirical results for the diustribution of b2 and b1", *Biometrika*, vol. 60, no. 3, pp. 613–622, 1973.

[103] G. E. DALLAL, L. WILKINSON, "An analytic approximation to the distribution of Lilliefors's test statistic for normality", *The American Statistician*, vol. 40, no. 4, pp. 294–296, Nov. 1986.

[104] D. DEMBELE, *Identification de modèles ARMA linéaires à l'aide de statistiques d'ordre élevé, Application à l'égalisation aveugle*, Doctorat, Université de Nice Sophia-Antipolis, juillet 1995.

[105] D. DONOHO, "On minimum entropy deconvolution", in *Applied time-series analysis II*, pp. 565–609. Academic Press, 1981.

[106] T. W. EPPS, "Testing that a stationary time series is Gaussian", *The Annals of Statistics*, vol. 15, no. 4, pp. 1683–1698, 1987.

[107] G. FAVIER, *Filtrage, modélisation et identification de systèmes liéaires stochastiques à temps discret*, CNRS, 1982.

[108] G. FAVIER, D. DEMBELE, J. L. PEYRE, "Identification de modèles paramétriques AR, MA, et ARMA avec les statistiques d'ordre supérieur, et analyse des performances", in *XIVeme Colloque Gretsi*, 13-16 Septembre 1993, pp. 137–140.

[109] G. FAVIER, J. P. PUY, G. MAYNARD, "Identification de modèles ARMA", in *XIIème Colloque Gretsi*, Juan les Pins, 12 -16 juin 1989, pp. 153–156.

[110] L. FETY, *Methodes de Traitement d'Antenne Adaptees aux Radiocommunications*, Doctorat, ENST, 1988.

[111] I. FIJALKOW, P. LOUBATON, "Identification of rank one rational spectral densities from noisy observations: a stochastic realization approach", *Systems and Control Letters*, , no. 24, pp. 201–205, 1995.

[112] R. FORTET, *Elements de la theorie des probabilites*, CNRS, 1965.

[113] K. FUKUNAGA, T. E. FLICK, "A test of the Gaussian-ness of a data set using clustering", *IEEE Trans. Pattern Ana. Mach. Intel.*, vol. 8, no. 2, pp. 240–247, 1986.

[114] M. GAETA, J. L. LACOUME, "Source separation without a priori knowledge: the maximum likelihood solution", in *Proc. EUSIPCO,* Barcelona, Spain, 1990, pp. 621–624.

[115] F. GAMBOA, "Separation of sources having unknown discrete supports", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp. 56–60.

[116] T. GASSER, "Goodness-of-fit tests for correlated data", *Biometrika*, vol. 62, no. 3, pp. 563–570, 1975.

[117] E. GASSIAT, *Déconvolution aveugle*, Doctorat, Université de Paris-sud, Orsay, janvier 1988.

[118] E. GASSIAT, "Blind deconvolution of discrete linear systems perturbed with additive noise", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp. 305–309.

[119] S. Van GERVEN, D. Van COMPERNOLLE, "On the use of decorrelation in scalar signal separation", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'94), vol.III*, Adelaide, Australia, Apr. 1994, pp. 57–60.

[120] D. GESBERT, P. DUHAMEL, S. MAYRARGUE, "Subspace-based adaptive algorithms for the blind equalization of multichannel FIR filters", in *Proc. EUSIPCO*, Edinburgh, Sept. 1994, pp. 712–715.

[121] G. B. GIANNAKIS, M. K. TSATSANIS, "Time-domain tests for Gaussianity and time-reversibility", *IEEE Trans. on Signal Processing*, vol. 42, no. 12, pp. 3460–3472, Dec. 1994.

[122] N. R. GOODMAN, "Statistical analysis based on certain multivariate complex normal distributions", *Annals Math. Stat.*, vol. 34, pp. 152–177, 1963.

[123] E. J. HANNAN, *Multiple time series*, Wiley, 1970.

[124] E. J. HANNAN, M. DEISTLER, *The statistical theory of linear systems*, Wiley, 1988.

[125] J. HERAULT, C. JUTTEN, *Réseaux neuronaux et traitement du signal*, Traitement du Signal. Hermes, Paris, 1994.

[126] J. HERTZ, A. KROGH, R. G. PALMER, *Introduction to the theory of Neural Computation*, Addison Wesley, 1991.

[127] M. HINICH, "Testing for Gaussianity and linearity of a stationary time series", *Journal of Time Series Analysis*, vol. 3, no. 3, pp. 169–176, 1982.

[128] P. J. HUBER, "Projection pursuit", *The Annals of Statistics*, vol. 13, no. 2, pp. 435–475, 1985, Invited paper with discussion.

[129] Y. INOUYE, "Modeling of multichannel time series and extrapolation of matrix-valued autocorrelation sequences", *IEEE Trans ASSP*, vol. 31, no. 1, pp. 45–55, Feb. 1983.

[130] Y. INOUYE, T. HABE, "Blind equalization of multichannel linear time-invariant systems", *The Institute of Electronics Information and Communication Engineers*, , no. 24, pp. 9–16, May 1995.

[131] N. L. JOHNSON, S. KOTZ, *Distributions in statistics: Continuous Univariate Distributions-1*, Wiley, 1970.

[132] C. JUTTEN, J. HÉRAULT, "Independent component analysis versus PCA", in *Proc. EUSIPCO,* Grenoble, France, 1988, pp. 643–646.

[133] C. JUTTEN, J. HERAULT, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture", *Signal Processing, Elsevier*, vol. 24, no. 1, pp. 1–10, 1991.

[134] A. M. KAGAN, Y. V. LINNIK, C.R. RAO, *Characterization Problems in Mathematical Statistics*, Wiley, 1973.

[135] T. KAILATH, *Linear Systems*, Prentice-Hall, 1980.

[136] M. KENDALL, A. STUART, *The Advanced Theory of Statistics, Distribution Theory*, vol. 1, C. Griffin, 1977.

[137] M. KENDALL, A. STUART, *The Advanced Theory of Statistics, Design and Analysis, and Time-Series*, vol. 3, C. Griffin, 1979.

[138] S. KOTZ, N. L. JOHNSON, *Encyclopedia of Statistical Sciences*, Wiley, 1982.

[139] M. KROB, *Identification aveugle de modèles non linéaires à l'aide de statistiques d'ordre supérieur*, Doctorat de l'Université de Paris-sud, Orsay, 8 fevrier 1994.

[140] M. KROB, M. BENIDIR, "Blind identification of a linear-quadratic mixture", in *Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA*, 1993, pp. 351–355.

[141] M. KROB, M. BENIDIR, "Une fonction de contraste pour l'identification aveugle d'un modele lineaire quadratique", in *XIVeme Colloque Gretsi*, 13-16 Septembre 1993, pp. 101–104.

[142] J. L. LACOUME, M. GAETA, P. O. AMBLARD, "From order 2 to HOS: new tools and applications", in *Proc. European Signal Processing Conf. EUSIPCO*, Brussels, Aug 24 - 27 1992, pp. 91–98.

[143] J. L. LACOUME, F. HARROY, "Performances in blind sources separation", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp. 25–29.

[144] J. L. LACOUME, P. RUIZ, "Separation of independent sources from correlated inputs", *IEEE Trans. Sig. Proc.*, vol. 40, no. 12, pp. 3074–3078, Dec. 1992.

[145] B. LAHELD, J. F. CARDOSO, "Adaptive source separation without prewhitening", in *Proc. EUSIPCO*, Edinburgh, Sept. 1994, pp. 183–186.

[146] H. J. LANDAU, "Maximum entropy and the moment problem", *Bulletin of the American Math. Soc.*, vol. 16, no. 1, pp. 47–77, Jan. 1987.

[147] P. LASCAUX, R. THEODOR, *Analyse numérique matricielle appliquée à l'art de l'ingénieur*, Masson, 1986.

[148] J. P. LECADRE, "Au carrefour de nombreuses applications, la calibration d'antenne", *Traitement du signal*, vol. 10, no. 5, pp. 347, 1993, Numéro spécial calibration.

[149] C. C. LIN, "A simple test for normality against asymmetric alternatives", *Biometrika*, vol. 67, no. 2, pp. 455–461, 1980.

[150] L. LJUNG, T. SODERSTROM, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, 1983.

[151] P. LOUBATON, "Techniques du second ordre pour la déconvolution aveugle multi-sources multi-capteurs", in *Journée signal de cergy*, ENSEA, Cergy, 2 fev 1995, pp. 2–15.

[152] G. LUKACS, *Characteristic functions*, Griffin, 1960.

[153] O. MACCHI, *Adaptive processing*, Wiley, 1995.

[154] O. MACCHI, E. EWEDA, "Convergence analysis of self-adaptive equalizers", *IEEE Trans. Information theory*, vol. 30, no. 2, pp. 161–176, Mar. 1984.

[155] K. V. MARDIA, "Measures of multivariate skewness and kurtosis with applications", *Biometrika*, vol. 57, pp. 519–530, 1970.

[156] K. V. MARDIA, "Applications of some measures of multivariate skewness and kurtosis for testing normality", *Sankhya B*, vol. 36, pp. 115–128, 1974.

[157] K. V. MARDIA, "Tests of univariate and multivariate normality", in *Handbook of Statistics, Vol.1*, P. R. Krishnaiah, Ed., pp. 279–320. North-Holland, 1980.

[158] K. V. MARDIA, K. FOSTER, "Ominibus tests of multinormality based on skewness and kurtosis", *Commun. Statist. Simula. Computa.*, vol. 12, no. 2, pp. 207–221, 1983.

[159] K. V. MARDIA, M. KANAZAWA, "The null distribution of multivariate kurtosis", *Commun. Statist. Simula. Computa.*, vol. 12, no. 5, pp. 569–576, 1983.

[160] P. McCULLAGH, *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability. Chapman and Hall, 1987.

[161] D. S. MOORE, "A chi-square statistic with random cell boundaries", *The Annals of Statistics*, vol. 42, no. 1, pp. 147–156, 1971.

[162] D. S. MOORE, "The effect of dependence on chi squared tests of fit", *The Annals of Statistics*, vol. 10, no. 4, pp. 1163–1171, 1982.

[163] E. MOREAU, *Apprentissage et adaptivité, séparation auto-adaptative de sources indépendantes par un réseau de neurones*, Doctorat de l'Université de Paris-sud, Orsay, 1 fevrier 1995.

[164] E. MOREAU, O. MACCHI, "Separation de sources adaptative sans blanchiment prealable", in *XIVeme Colloque Gretsi*, 13-16 Septembre 1993.

[165] E. MOREAU, O. MACCHI, "A one stage self-adaptive algorithm for source separation", in *Proc. ICASSP*, Adelaide, Australia., 1994.

[166] E. MOULINES, K. CHOUKRI, M. CHARBIT, "Testing that a multivariate stationary time series is Gaussian", in *Sixth SSAP Workshop on Stat. Signal and Array Proc.*, Oct. 1992, pp. 185–188.

[167] E. MOULINES, P. DUHAMEL, J. F. CARDOSO, S. MAYRAGUE, "Subspace methods for the blind identification of multichannel FIR filters", *IEEE Trans. on Signal Processing*, vol. 43, no. 2, pp. 516–525, Feb. 1995.

[168] H. L. NGUYEN-THI, C. JUTTEN, "Comparaison de quelques algorithmes adaptatifs de separation de sources dans un melange convolutif", in *XIV Colloque GRETSI*, Juan les Pins, France, Sept. 13–16 1993, pp. 333–336.

[169] C. L. NIKIAS, A. P. PETROPULU, *Higher-Order Spectra Analysis*, Signal Processing Series. Prentice-Hall, Englewood Cliffs, 1993.

[170] K. V. PARSHALL, "The one-hundred anniversary of the death of invariant theory", *The Mathematical Intelligencer*, vol. 12, no. 4, pp. 10–16, 1990.

[171] E. S. PEARSON, R. B. D'AGOSTINO, K. O. BOWMAN, "Tests for departure from normality: Comparison of powers", *Biometrika*, vol. 64, no. 2, pp. 231–246, 1977.

[172] E. S. PEARSON, H. O. HARTLEY, *Biometrika Tables for Statisticians*, vol. I, Cambridge University Press, 1962.

[173] D. T. PHAM, P. GARRAT, "Separation of a mixture of independent sources through a maximum likelihood approach", in *Proc. European Signal Processing Conf. EUSIPCO*, Brussels, Aug 24 - 27 1992, pp. 771–774.

[174] B. PICINBONO, "Spherically invariant and compound stochastic processes", *IEEE Trans. Information Theory*, vol. 16, no. 1, pp. 77–79, Jan. 1970.

[175] B. PICINBONO, *Random Signals and Systems*, Prentice-Hall, 1993.

[176] B. PICINBONO, "On circularity", *IEEE Trans. Signal Processing*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.

[177] J. G. PROAKIS, C. L. NIKIAS, "Blind equalization", in *SPIE Adaptive Signal Processing*, 1991, vol. 1565, pp. 76–88.

[178] S. PROSPERI, "Décomposition de lois, fonctions caractéristiques, et caractérisation", *Traitement du Signal*, vol. 11, no. 2, pp. 117–131, février 1994.

[179] G. C. REINSEL, *Elements of multivariate time series analysis*, Springer-Verlag, 1993.

[180] M. ROSENBLATT, *Stationary Processes and Random Fields*, Birkhauser, 1985.

[181] M. ROSENBLATT, "Gaussian and nongaussian linear sequences", in *New directions in time series analysis*, D. Brillinger et al, Ed., vol. 45 of *IMA Volumes in Mathematics and its Applications*, pp. 327–333. Springer Verlag, 1992.

[182] D. ROSSILLE, *Reconstruction à partir du bispectre, Application à l'astronomie, Effets de l'échantillonnage et de la stationnarité sur les spectres d'ordre supérieur*, Doctorat, Université de Nice Sophia-Antipolis, 20 juin 1995.

[183] J. LE ROUX, D. ROSSILLE, C. HUET, "A multiresolution extension of Lohmann-Weigelt-Wirnitzer recursion for computing a Fourier transform phase from a third order spectrum phase", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp. 315–319.

[184] E. M. SANIGA, J. A. MILES, "Power of some standard goodness-of-fit tests of normality against stable asymmetric alternatives", *Jour. Am. Stat. Assoc.*, vol. 74, no. 368, pp. 861–865, Dec. 1979.

[185] C. SERVIERE, V. CAPDEVIELLE, "An identification method of FIR digital filters in frequency domain", in *Proc. EUSIPCO*, Edinburgh, Sept. 1994, pp. 1058–1061.

[186] O. SHALVI, E. WEINSTEIN, "New criteria for blind deconvolution of nonminimum phase systems", *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 312–321, Mar. 1990.

[187] S. S. SHAPIRO, M. B. WILK, H. J. CHEN, "A comparative study of various tests for normality", *American Statistical Association Journal*, vol. 63, pp. 1343–1372, Dec. 1968.

[188] J. E. SHORE, R. W. JOHNSON, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy", *IEEE Trans. Information Theory*, vol. 26, no. 1, pp. 26–37, Jan. 1980.

[189] T. SODERSTROM, P. STOICA, *System Identification*, Prentice-Hall, 1989.

[190] A. SOULOUMIAC, *Utilisation des statistiques d'ordre supérieur pour la séparation et le filtrage*, Doctorat, ENST, Février 1993.

[191] A. SOULOUMIAC, J. F. CARDOSO, "Performances en séparation de sources", in *Proc. GRETSI, Juan les Pins,France*, 1993, pp. 321–324.

[192] Y. STEINBERG, O. ZEITOUNI, "On tests for normality", *IEEE Trans. on Inf. Theory*, vol. 38, no. 6, pp. 1779–1787, Nov. 1992.

[193] M. A. STEPHENS, "Edf statistics for goodness of fit and some comparisons", *Journal of the American Statistical Association*, vol. 69, no. 347, pp. 730–737, 1974.

[194] A. SWAMI, G. GIANNAKIS, S. SHAMSUNDER, "Multichannel ARMA processes", *IEEE Trans. on Signal Processing*, vol. 42, no. 4, pp. 898–913, Apr. 1994.

[195] P. TICHAVSKY, A. SWAMI, "Statistical characterization of sample fourth-order cumulants of a noisy complex sinusoidal process", *IEEE Trans. on Signal Processing*, vol. 43, July 1995.

[196] L. TONG, R. LIU, V. C. SOON, "Indeterminacy and identifiability of blind identification", *IEEE Trans Circuits and Systems*, vol. 38, no. 5, pp. 499–509, May 1991.

[197] L. TONG, G. XU, T. KAILATH, "Blind identification and equalization based on second-order statistics: a time domain approach", *IEEE Trans. on Signal Processing*, vol. 40, no. 2, pp. 340–349, Mar. 1994.

[198] J. TUGNAIT, "Comments on 'new criteria for blind deconvolution of nonminimum phase systems"', *IEEE Trans. Inf. Theory*, vol. 38, no. 1, pp. 210–213, Jan. 1992.

[199] J. K. TUGNAIT, "Detection of non-Gaussian signals using integrated polyspectrum", *IEEE Trans. on Signal Processing*, vol. 42, no. 11, pp. 3137–3149, Nov. 1994.

[200] O. VASIECEK, "A test for normality based on sample entropy", *Jour. Roy. Statist. Soc. B*, vol. 38, pp. 54–59, 1976.

[201] R. A. WOODING, "The multivariate distribution of complex normal variables", *Biometrika*, vol. 43, pp. 212–215, 1956.

[202] D. YELLIN, E. WEINSTEIN, "Multi-channel signal separation based on cross-bispectra", in *Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA*, 1993, pp. 270–274.

[203] D. YELLIN, E. WEINSTEIN, "Criteria for multichannel signal separation", *IEEE Trans. on Signal Processing*, vol. 42, no. 8, pp. 2158–2168, Aug. 1994.

[204] V. ZIVOJNOVIC, "Higher-order statistics and Huber's robustness", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp. 236–240.

[205] I. G. ZURBENKO, *The spectral analysis of time series*, North-Holland, 1985.

6.3 Annexes

Pour ne pas encombrer inutilement le document, ce sont essentiellement les articles de revue qui sont rassemblés dans cette annexe.

Sommaire