

Principes des méthodes de séparation fondées sur les moments d'ordre supérieur

Principles of source separation methods based on higher order statistics



P. DUVAUT

Laboratoire de Traitement des Images et du Signal, ENSEA, Les chênes Pourpres, 95000 Cergy-Pontoise, France

Société Kurtosis Ingénierie, 5 avenue du Parc, Le Campus, 95032 Cergy-Pontoise, France

Ancien Elève de l'ENS de Cachan, agrégé de Sciences Physiques en 1981, Docteur en Sciences de l'Université d'Orsay en 1987, il dirige le Département Mathématiques et Traitement du Signal de l'Ecole Nationale Supérieure de l'Electronique et de ses Applications, ENSEA, à Cergy-Pontoise. Il poursuit des activités de recherche au sein de l'équipe de traitement des Images et du Signal, ETIS, de l'ENSEA. Ses principaux thèmes de recherche concernent le traitement des signaux Radar (en relation avec le Département DEM de la Thomson à Malakoff), l'utilisation des moments d'ordre supérieur en Détection et en Estimation, les relations entre non stationnaire, non linéaire et non gaussien. Animateur du Conseil Scientifique de la société Kurtosis Ingénierie, il co-dirige la collection Traitement du Signal des Editions Hermès.

RÉSUMÉ

Une présentation unifiée des méthodes de séparation de sources fondées sur les moments d'ordre supérieur est proposée. L'approche considérée part du recensement systématique des paramètres caractéristiques de toutes les méthodes, comme : les familles d'hypothèses statistiques concernant les données ; les catégories de modèles conjecturés (standards ou doublement orthogonaux) ; les critères de séparation (d'indépendance), qui conduisent à la restitution des sources, énoncés à partir de moments ou de cumulants ; les principes effectifs des méthodes de séparation (démarches directes utilisant une matrice de restitution des sources, démarches indirectes ou globales identifiant au préalable ou simultanément d'autres entités propres aux sources, amplitudes etc...). De manière générale, il est établi que l'ensemble de ces méthodes

conduisent à restituer, non pas un jeu de sources uniques, mais plutôt les éléments appartenant à l'intersection de deux classes d'équivalence. La première appelée classe du second ordre est associée à tous les vecteurs aléatoires de même matrice de covariance. La seconde qualifiée de classe d'indépendance, doit son existence à l'invariance de l'indépendance mutuelle de variables aléatoires dans toute opération de permutation-homothétie.

MOTS CLÉS

Séparation, indépendance, moments d'ordre supérieur, cumulants, analyse en composantes indépendantes, développement de Gram-Charlier, réseaux de neurones, rotations de Givens, décomposition de Cholesky.

SUMMARY

A unifying presentation of source separation methods based on higher order statistics is derived. The approach starts from the systematic scanning of the characteristic parameters of all the methods, like : the statistical hypotheses about the data ; the different kinds of models assumed, (standard or doubly orthogonal) ; the separation criteria, (independence), that lead to the sources restoration, written with moments or cumulants ; the effective principles of the separation, (direct trails based on a restoration matrix, indirect or global trails that estimate other characteristics of the sources, amplitudes etc..., before the true separation). From a general standpoint, it is shown that, these methods yield a restoration of

elements that belong simultaneously to two classes, more than a unique set of sources. The first class, named second order class is associated with all random vectors that have the same covariance matrix. The rationale of the second class, considered as a class of independence, stems from the invariance of mutual independence by any permutation operation.

KEY WORDS

Separation, independence, higher order statistics, cumulants, independant components analysis, Gram-Charlier expansion, neural nets, Givens rotation, Cholesky factorization.

1. Introduction

La séparation de sources indépendantes, fondée sur l'utilisation des moments d'ordre supérieur, MOS, [1], connaît depuis quelques années un essor tout particulier. Ce dernier se concrétise par l'apparition d'un nombre important de méthodes, [2], [3], [4], [5], [6], [7], [8], [9], qui, au

premier abord, semblent parfois très différentes, aussi bien au niveau des hypothèses de départ que des principes généraux sur lesquels reposent leurs méthodologies respectives. L'objectif principal de ce papier est de présenter, de façon unifiée et cohérente, à l'aide d'un formalisme unique, les fondements des principales méthodes connues et testées à ce jour. Le caractère novateur de ce travail est

donc, par essence même, le suivant. Il s'agit de préciser, en laissant subsister le moins d'ambiguïté possible, les *spécificités* de chaque approche : *modèle* des données, *hypothèses* sous jacentes, *intérêts* d'utiliser les moments d'ordre supérieur à deux, disponibles sous forme de *moments* traditionnels ou de *cumulants*, [1], *caractéristiques* de la séparation, etc... Cette étape en précède nécessairement une seconde, à plus long terme, qui fera l'objet d'une autre publication et qui vise à comparer les performances des algorithmes qui réalisent ces méthodes. Ces performances doivent être évaluées, d'ailleurs, à l'aide d'un ou plusieurs critères « ad hoc », suffisamment significatifs pour l'ensemble des approches. Le choix de tels critères demande encore, aujourd'hui, des approfondissements non négligeables. A cet égard, la référence [4] apporte une contribution en introduisant un *critère effectif de séparation* fondé sur une fonction dite de *contraste*.

L'organisation de l'article est la suivante. Le *paragraphe 2* est consacré à la présentation des différents *modèles* des données préconisés dans la plupart des méthodes recensées, à ce jour. L'utilisation de ces modèles est conditionnée, bien sûr, par différentes *hypothèses statistiques* variables suivant les références, [2], [3], [4], [5], [6], [7], [8], [9]. L'exposition de celles-ci prend place également dans le *paragraphe 2*. Les *objectifs* communs des différentes approches de la séparation de sources « bande étroite » y sont précisés aussi, sachant que le *paragraphe 4* les détaille davantage. On y trouve, pour finir, les différences avec les algorithmes traditionnels dits de localisation, [10], [11]. Le *paragraphe 3* exhibe les *ambiguïtés et difficultés* propres à la séparation qui interviennent au niveau du *modèle*, d'une part, et du *critère de séparation* effective des sources, d'autre part. Les limitations des modèles viennent du fait que ceux-ci ne représentent pas des signaux uniques mais plutôt *deux classes de signaux*. Une classe est liée à l'hypothèse d'indépendance statistique des sources, elle est appelée, en cela, *classe d'indépendance*. Sa définition repose sur le fait que toute opération de « permutation-homothétie » d'un certain nombre de variables aléatoires indépendantes conserve l'indépendance. Sous la seule hypothèse d'indépendance, on peut donc espérer, au mieux, atteindre la classe d'équivalence et non l'un de ses éléments. La seconde classe une apparaît lorsque l'on caractérise l'un des modèles uniquement par ses propriétés du second ordre qui se résument, ici, à une matrice de covariance. En effet, il existe une infinité de « factorisation en racine carrée » d'une matrice hermitienne définie positive, infinité engendrée par l'espace vectoriel des matrices unitaires de taille donnée, [12]. Si cette taille est égal à P , on déduit que « l'ordre de dégénérescence » de cette seconde classe est égal à $P(P-1)/2$, (les conditions d'orthonormalité génèrent $P(P+1)/2$ équations pour P^2 éléments). *Toutes les méthodes de séparation visent à lever totalement la dégénérescence de cette deuxième classe, en élaborant les $P(P-1)/2$ conditions supplémentaires manquantes.*

En dehors des ambiguïtés inhérentes aux modèles, quelques difficultés, non parmi les moindres, proviennent du *critère* testant la qualité de la séparation. Si, théoriquement, toutes les approches décrites dans [2], [3], [4], [5], [6], [7], [8], [9], tendent à restituer des sources statistique-

ment indépendantes dans leur ensemble, [13], *pratiquement*, néanmoins, elles mettent en œuvre des algorithmes dont le test d'arrêt revient à vérifier la réalisation de *conditions plus faibles*. La justification de ces palliatifs est contenue dans les idées suivantes. L'indépendance mutuelle d'un certain nombre de variables aléatoires, VA, se traduit par le *caractère additif* de la *seconde fonction caractéristique* du vecteur qui les regroupe, [13]. Cela signifie que tous les termes croisés du développement en série multidimensionnel de cette fonction sont nuls. Dans la mesure où il n'est pas envisageable de tester la nullité d'une infinité de termes, quelques méthodes, [4], [5], [8], [9], se limitent à la notion d'*indépendance jusqu'à l'ordre 4*, (annulation des termes croisés du développement en série jusqu'à l'ordre 4). Plus encore, elles se contentent, en dehors de celles contenues dans [4] et [8], de réaliser des *conditions nécessaires* d'indépendance d'ordre 4 et non suffisantes, en annulant, que quelques termes croisés d'ordre 4. La référence [4] montre comment, du fait de la *redondance inhérente* aux termes croisés d'ordre 4, le *choix pertinent* de termes à annuler conduit à une condition *suffisante* également, mais toujours limitée, bien sûr, à l'indépendance d'ordre 4. (Dans la suite on dénotera c_4 une telle condition d'indépendance d'ordre 4). Les références [2], [3], [6] et [7] introduisent elles aussi des conditions d'indépendances plus faibles, de définition beaucoup plus *vague*, (on ne connaît pas cette fois l'ordre d'indépendance atteint par la méthode), que les c_4 , dans la mesure où elles portent sur des *transformées non linéaires* des données à traiter, (on les symbolise par c_{NL} , dans la suite). Une contrainte très précise limite toutefois le nombre de c_4 et de c_{NL} : les conditions candidates doivent *lever la dégénérescence* d'ordre $P(P-1)/2$ vue plus haut. Comme il a déjà été indiqué brièvement, le *paragraphe 4*, livre les *principes généraux* des méthodes de séparation, contenues dans [2], [3], [4], [5], [6], [7], [8], [9], en distinguant deux familles, les approches *directes et indirectes*. Le cinquième *paragraphe*, quant à lui, présente sous forme de deux planches récapitulatives et comparatives l'ensemble des caractéristiques des méthodes étudiées dans l'article. Il est à noter également, qu'en dehors de la synthèse qu'il représente, ce travail contient, dans le *paragraphe 4*, une *justification nouvelle*, fondée sur le *produit de Kronecker*, PK, [14], de certains résultats de [8]. La référence [15] contient aussi une approche *non tensorielle*, [18], des concepts de [8]. Dans l'article [16], Friedlander et Porat introduisent une méthode de localisation spatiale utilisant les MOS, présentée, de même, à l'aide du PK. Pour finir, Comon et Cardoso, proposent dans [17], une synthèse des algorithmes de diagonalisation d'un tenseur d'ordre 4, et établissent comment les méthodes matricielles de [6] se déduisent par contraction des méthodes tensorielles de [8].

2. Modèles et hypothèses

2.1. LE MODÈLE STANDARD

On s'intéresse à P sources « bande étroite » noyées dans un bruit additif. P capteurs collectent l'information résultante, (d'emblée, déjà au niveau de la structure de

réception, une différence fondamentale avec les méthodes dites de *localisation* apparaît, telles que Music, [10], ou la norme minimale, [11], qui nécessitent au moins 1 capteur de plus que le nombre de sources à localiser). Sous forme vectorielle, on dispose, à l'instant t , de l'information

$$(2.1) \quad \mathbf{r}(t) = \mathbf{X}\mathbf{s}(t) + \mathbf{b}(t).$$

La matrice \mathbf{X} , *déterministe*, illustre les interactions entre les P sources et les P capteurs : nature de la propagation, forme des fronts d'onde, géométrie de l'antenne de réception, etc... Elle est appelée, en cela, matrice *spatiale*. Écrite comme concaténation de P vecteurs colonnes, \mathbf{x}_j , $1 \leq j \leq P$, elle fait apparaître ce qu'il convient d'appeler, les vecteurs *signatures* des sources.

Des hypothèses particulières relatives à la propagation, etc..., débouchent sur une *paramétrisation* de ces signatures, appelée, *paramétrisation spatiale*, dont la plus connue émane des méthodes de *localisation*, voir [10] et [11]. Pour des fronts d'onde plans et une antenne linéaire à capteurs équirépartis, chaque vecteur \mathbf{x}_j , appelé alors *vecteur de direction* (*steering vector*), est formé de nombres complexes de module unité, en progression géométrique, la raison dépendant directement de la position relative de la source et de l'antenne. Cette paramétrisation spatiale particulière est d'ailleurs à l'origine des algorithmes de localisation. Pour l'instant, notre propos ne repose sur aucune paramétrisation spatiale. Il fait néanmoins l'objet d'une première hypothèse :

HS les vecteurs \mathbf{x}_j sont algébriquement indépendants.

Cela signifie seulement que le modèle spatial de l'interaction sources-capteurs n'est pas *dégénéré*. Le vecteur $\mathbf{s}(t)$, *aléatoire*, admet comme composantes, les P messages, $s_j(t)$, contenant l'information propagée par toutes les sources. Le *découplage* des contributions *spatiales* et *informationnelles* apparent dans la relation (2.1) est dû au caractère « bande étroite » des sources : le terme modulant varie nettement moins vite, à l'échelle de l'antenne, que la porteuse, [10], [11]. Si tel n'est pas le cas, et moyennant une paramétrisation spatiale, on peut retrouver un découplage après transformée de Fourier, TF, des données, [10]. Le vecteur $\mathbf{b}(t)$, lui aussi aléatoire contient la contribution du bruit. L'introduction de considérations informationnelles induit des hypothèses supplémentaires susceptibles de caractériser les propriétés statistiques des sources ou du bruit et variables suivant les références, voir [2], [3], [4], [5], [6], [7], [8], [9], etc... :

HI à tout instant, t , les amplitudes des sources, $s_j(t)$, sont des variables aléatoires, VA, indépendantes, centrées, de variance unité

HD les VA $s_j(t)$ sont centrées, décorréliées de variance unité

HB0 le bruit n'est pas pris en considération, ($\mathbf{b}(t) = \mathbf{0}_P$)

HBB le bruit additif est centré, blanc, au sens fort, indépendant des sources, de puissance σ^2 , ceci, à tout instant, t

HBBG le bruit a toutes les qualités de HBB, et, de plus, est gaussien.

D'une manière générale, on remarque, bien sûr, que HI entraîne HD, la réciproque n'étant vraie que dans le cas gaussien, [13]. Les hypothèses HS, HI, et HB0 sont rencontrées dans les références, [2], [3], [5], [6], [7] et [8]. Dans [9], seules les hypothèses relatives au bruit sont modifiées : HB0 est remplacée par HBBG.

Remarque R1. Dans la mesure où la seconde fonction caractéristique d'un vecteur gaussien, [13], s'arrête à l'ordre deux, (elle a donc des termes d'ordre 4 a fortiori nuls), les hypothèses HB0 et HBBG sont *indissociables* sur les termes d'ordre 4 du développement en série de la seconde fonction caractéristique du vecteur observation. Leur distinction n'est donc pas fondamentale pour la plupart des méthodes, sauf dans la référence [9] qui nécessite complètement HBBG, et non seulement HB0, du fait de l'*argument maximum de vraisemblance généralisé* qu'elle développe.

Remarque R2. L'hypothèse HI est à nuancer dans la référence [4], dans la mesure où les variables aléatoires $s_j(t)$ tout en restant indépendantes n'y sont pas de variance unité. Plus générale, cette nouvelle hypothèse confère, à l'instant t , une matrice de covariance diagonale à éléments strictement positifs, notée \mathbf{A}^2 , au vecteur $\mathbf{s}(t)$. On se ramène néanmoins, à un terme multiplicatif près, à la situation décrite par (2.1) en définissant le vecteur $\mathbf{s}_n(t)$, déduit de $\mathbf{s}(t)$, par normalisation, $\mathbf{s}(t) = \mathbf{A}\mathbf{s}_n(t)$, et vérifiant lui, par voie de conséquence, l'hypothèse HI. En résumé, dans les références, [2], [3], [5], [6], [7], [8], [9], $\mathbf{s}_n(t)$ coïncide avec $\mathbf{s}(t)$ et dans [4] $\mathbf{s}(t) = \mathbf{A}\mathbf{s}_n(t)$.

Si l'on désire faire des comparaisons supplémentaires avec es méthodes de *localisation*, on note, dans ces dernières qu'en dehors de HS qui est conservée, HI est remplacée, en général, voir [10] et [11], par une hypothèse plus faible, la décorrélation des sources, soit HD, ici, et HB0 par HBB. Parmi les références traitant de séparation, on distingue celles qui considère les données *réelles*, [2], [3], [5], [9], et celles, [4], [6], [7], [8], qui tiennent compte de la *complexification* du signal reçu dû au caractère *analytique*, [19], du codage. Il semble d'ailleurs que l'utilisation de statistiques d'ordre supérieurs dans la séparation de sources « complexes » pose des problèmes de fond d'une nature plus fondamentale qu'un simple jeu d'écriture, comme la notion de *circularité d'ordre supérieur*. Ce concept, qui fait l'objet de quelques études à l'heure actuelle, vise à étendre la notion de « circularité », clairement définie dans le cas gaussien, [20], portant, donc, sur les moments d'ordre deux, à des processus non gaussiens, dont les propriétés statistiques dépendent, alors, des moments de rang plus élevés. Les contenus de [21] et [22] vont, à ce titre, dans cette direction. Dans ces conditions, sous les hypothèses conjointes HIBB la matrice de covariance de $\mathbf{r}(t)$ s'écrit

$$(2.2) \quad \mathbf{R} = \mathbf{X}\mathbf{X}^\dagger + \sigma^2 \mathbf{I}.$$

Remarque R3. Eu égard au contenu de la remarque R2 il faut remplacer le produit $\mathbf{X}\mathbf{X}^\dagger$, par $\mathbf{X}\mathbf{A}^2\mathbf{X}^\dagger$, dans la référence [4].

Sous HIB0, il suffit de prendre σ nul. Une fois introduits le modèle de travail ainsi que les différentes hypothèses possibles qui viennent conditionner son utilisation, les définitions de questions de *localisation* d'une part et de *séparation* d'autre part sont opportunes.

Localisation : Étant donnée une paramétrisation spatiale $\mathbf{X}(\mathbf{d})$ qui fait apparaître P directions relatives sources-capteurs, regroupées dans le vecteur \mathbf{d} , il s'agit d'estimer au mieux le vecteur des directions \mathbf{d} , en utilisant les moments d'ordre 2, sous les hypothèses HDBB, lorsque le nombre de capteurs dépasse de 1 le nombre de sources, voir [10], [11], ou, beaucoup plus récemment, en utilisant les moments d'ordre 2 et 4, sous les hypothèses HIBB, où cette fois, la contrainte liée au nombre de capteurs n'est pas nécessaire, voir [22]. À l'ordre 2, ces techniques, désormais, parfaitement connues, [10], [11], exploitent la décomposition spectrale, [12], de la matrice \mathbf{R} . À l'ordre 4, une approche semblable, fondée sur un formalisme tensoriel qui rejoint celui de [21], est développée dans [22]. Les concepts nouveaux qui y figurent sont d'ailleurs aussi dignes d'intérêt en séparation, comme il sera vu ultérieurement.

Séparation : Sans paramétrisation spatiale intrinsèque, c'est-à-dire, pouvant être reliée à un paramètre physique quelconque comme la géométrie des fronts d'onde, de l'antenne, des angles, etc..., de la matrice \mathbf{X} , il s'agit cette fois, dans la plupart des cas, [2], [3], [4], [5], [6], [7], [8], [9], sous les hypothèses HIB0, et plus rarement, sous les hypothèses HIBBG, [9], (voir remarque R1), de restituer, *le plus fidèlement possible*, le vecteur $\mathbf{s}(t)$.

Commentaires.

C1. Il convient de préciser davantage le sens de *paramétrisation intrinsèque*, puisque, tout compte fait, cette qualité est déterminante, dans la différence entre localisation et séparation. Pour ce faire, on oppose à la *paramétrisation intrinsèque* cristallisant des propriétés physiques très précises de l'environnement expérimental, la paramétrisation *algorithmique* de \mathbf{X} , comme, par exemple, sa factorisation sous forme du produit d'une matrice hermitienne et d'une matrice unitaire, [12], telle que celle préconisée dans [9]. Il est clair, au vu de sa nature, déconnectée de toute considération sur la « géométrie » du dispositif que l'exploitation d'une paramétrisation algorithmique ne peut déboucher sur la valeur d'un angle, la forme d'un front d'onde etc... Elle s'intègre donc, éventuellement, comme une étape de la séparation. Ceci est le cas dans [9].

C2. Les définitions précédentes s'entendent, bien sûr, au sens large : il est possible de faire de la localisation en estimant, au préalable, le vecteur \mathbf{s} , c'est d'ailleurs l'étape préliminaire des méthodes globales du maximum de vraisemblance, [23], [24] ; comme il est loisible de conduire une séparation après avoir estimé la matrice \mathbf{X} , [6], [7], [8], [9]. La différence notable à retenir vient de l'exploitation ou non d'une paramétrisation spatiale intrinsèque, qui évidemment supprime des degrés de liberté et lève, par conséquent, des ambiguïtés.

C3. L'expression, *le plus fidèlement possible*, est volontairement ambiguë car elle illustre bien l'imprécision émanant de la formulation du problème de séparation lorsque la seule contrainte des solutions est leur indépendance, par paires ou globales, voir la discussion de [4]. En effet, si l'on change, par exemple, l'ordre des sources l'indépendance subsiste. Il convient, dans ces conditions, de reformuler la séparation plus comme la restitution d'une classe de signaux statistiquement indépendants, qu'un signal unique. Ceci fait l'objet du paragraphe 3.

Avant de lever en parties ces ambiguïtés, on finit ce deuxième paragraphe par l'introduction d'une version spécifique du modèle défini par (2.1), exploitant l'idée de décomposition *doublement orthogonale*, [25], version appelée *modèle canonique*, [6], [7], [8].

2.2. LE MODÈLE CANONIQUE

On s'intéresse au passage de l'écriture,

$$(2.3) \quad \sum_{j=1, P} \mathbf{x}_j s_j(t)$$

préconisée dans ce qui précède et qui est simplement orthogonale,

$$(2.4) \quad E[s_j(t) s_k(t)] = \delta_{jk}$$

(les vecteurs \mathbf{x}_j sont seulement linéairement indépendants, voir HS), à l'écriture

$$(2.5) \quad \sum_{j=1, P} \mathbf{y}_j s_j(t)$$

doublement orthogonale, où, en plus de (2.4), on a

$$(2.6) \quad \mathbf{y}_j^\dagger \mathbf{y}_k = \delta_{jk}.$$

Cette propriété est accessible par changement de base de chaque réalisation du vecteur $\mathbf{r}(t)$, sous HIB0. En effet, l'exploitation des *informations contenues dans le second ordre*, de $\mathbf{r}(t)$, autrement dit de la matrice \mathbf{R} permet une telle orthonormalisation. En considérant la factorisation, [12], de la matrice hermitienne définie positive,

$$(2.7) \quad \mathbf{R} = \mathbf{C}^{1/2} [\mathbf{C}^{1/2}]^\dagger$$

les vecteurs

$$(2.8) \quad \mathbf{y}_j = [\mathbf{C}^{1/2}]^{-1} \mathbf{x}_j$$

associés à l'observation *anonique*,

$$(2.9) \quad \mathbf{c}(t) = [\mathbf{C}^{1/2}]^{-1} \mathbf{r}(t)$$

vérifient la propriété (2.6). On doit, alors, considérer un nouveau jeu de signatures spatiales, les \mathbf{y}_j , des sources. La représentation canonique, comme sa définition le suggère, apporte un intérêt aux seuls méthodes de séparation qui estiment d'abord les signatures spatiales des sources comme dans [6], [7] et [8].

3. Ambiguïtés et difficultés de la séparation

3.1. AMBIGUÏTÉS DU MODÈLE DE SOURCES INDÉPENDANTES, (CANONIQUE OU NON), CARACTÉRISÉ PAR UN SEUL SECOND ORDRE

On s'intéresse, ici, à quelques propriétés de la partie utile, $Xs(t)$, (mise éventuellement sous forme canonique $Ys(t)$, voir § 2.2.), sous les hypothèses HSI, voir § 2., de l'observation donnée par (2.1), ceci quelles que soient les hypothèses concernant le bruit, lorsque l'on dispose de l'information du second ordre, XX^\dagger .

Remarque R4. D'après les remarques R2 et R3 toute l'étude qui suit s'adapte au contenu de la référence [4] en remplaçant X par $X' = XA$.

On réunit ces qualificatifs sous forme d'un triplet, E , *modèle, hypothèses, connaissance*,

$$(3.1) \quad EI2 = \{Xs(t) \text{ resp. } (Ys(t)); \text{ HSI}; XX^\dagger \text{ resp. } (I)\}$$

plus concis à manipuler, (I fait référence à indépendance et 2 au second ordre). La donnée de E ne coïncide pas avec celle d'un vecteur $s(t)$ unique, mais plutôt celle de deux classes de signaux, bâties à partir de $s(t)$. La première classe émane de l'hypothèse d'indépendance statistique, HI. La seconde du fait que les connaissances disponibles sont limitées au second ordre.

3.1.1. Classe d'indépendance

On définit l'ensemble

$$(3.2) \quad CI(s) = \{v \in C^P \text{ tels que } \exists D \in \Delta \text{ et } P \in \Pi \text{ avec } v = DPss\}$$

où Δ est l'ensemble des matrices diagonales (P, P) dont les termes non nuls sont des nombres complexes de module unité et Π l'ensemble des matrices de permutation (P, P). Il est clair, d'après les propriétés élémentaires des matrices D et P , que tous les éléments v de $CI(s)$ sont des vecteurs aléatoires de composantes indépendantes de variance unité, (de matrice variance A^2 dans le cas de [4], voir R2). Ces deux dernières qualités ne définissent donc pas uniquement s , mais plutôt $CI(s)$.

3.1.2. Classe du second ordre

On considère l'ensemble

$$(3.3) \quad C2(X) = \{u \in C^P \text{ tels que } \exists Q \in U \text{ avec } u = Qs\}$$

où U est l'ensemble des matrices unitaires complexes (P, P). Tous les éléments u , de $C2(X)$ génèrent des observations utiles, Xu , dont la matrice de covariance est égale à XX^\dagger . En conséquence la donnée de XX^\dagger renvoie à l'ensemble $C2(X)$ et non à s seulement.

3.1.3. Introduction de degrés de liberté dans le modèle

On déduit des considérations de 3.1.1. et 3.1.2. que E représente, en fait, un vecteur aléatoire, w , appartenant à l'intersection de $CI(s)$ et $C2(X)$, notée $CI2(s, X)$. En

conséquence, il convient de faire apparaître explicitement cette indétermination en réécrivant E sous la forme

$$(3.4) \quad E(Q; DP) = \{XQDPs(t) \text{ resp. } (YQDPs(t)); \text{ HSI}; XX^\dagger \text{ resp. } (I)\}.$$

En dehors des matrices D et P dont l'identification exacte, n'a pas fait encore, à notre connaissance, l'objet de travaux, E dépend de $P(P-1)/2$ degrés de libertés : la matrice unitaire Q à P^2 éléments, vérifie P équations de normalisation et $P(P-1)/2$ relations d'orthogonalité, il subsiste donc une « dégénérescence » d'ordre $P^2 - P - [P(P-1)/2]$.

3.1.4. Reformulation du problème de séparation

Compte tenu des discussions des § 3.11, 3.12 et 3.13 nous sommes en mesure maintenant de reformuler, de manière plus précise le problème de séparation :

Étant donné

$$(3.5) \quad E(Q; DP) = \{XQv \text{ resp. } (Ysv); \text{ HSI}; XX^\dagger \text{ resp. } (I)\}$$

il s'agit de restituer l'un quelconque, v , des éléments de la classe $CI(s)$, (D et P restant indéterminés), et de lever simultanément la dégénérescence d'ordre $P(P-1)/2$ de $C2(X)$ en estimant la matrice unitaire Q . (Remplacer X par X' dans (3.5) pour s'adapter aux contingences de [4], voir R4).

3.2. DIFFICULTÉS LIÉES AU CRITÈRE DE SÉPARATION

3.2.1. Position du problème

Les deux qualités que l'on va exiger d'un vecteur source restitué, noté a , sont les deux propriétés caractéristiques de la classe $CI(s)$, c'est-à-dire d'être formé de composantes indépendantes de variance unité, (de matrice variance A^2 dans [4]).

La dernière qualité s'obtient, pratiquement, sans difficulté, à l'aide d'une contrainte géométrique, telles que celles que l'on rencontre dans les domaines liés au filtrage d'antenne, [26], ou encore à l'analyse spectrale haute résolution, [27]. La première, par contre, appelle, comme il a déjà été précisé dans l'introduction, des conditions très fastidieuses, voire pratiquement irréalisables, du fait du manque d'information. En effet, parmi les conditions suffisantes d'indépendance, figurent la factorisation d'une densité de probabilité conjointe, celle d'une espérance mathématique, etc..., voir [13], qui ne sont pas d'une utilisation très commode.

3.2.2. Une condition suffisante d'indépendance fondée sur les cumulants

Plus récemment, avec l'introduction fréquente du Polyspectre, [1], et celle étroitement liée de « cumulants », [28], une condition suffisante, plus commode à vérifier, mais restant, malgré tout laborieuse, est apparue. Celle-ci, nécessite, pour être présentée simplement de définir les

cumulants. Le lecteur désirant des approfondissements pourra consulter la référence [1]. Soit une suite de P variables aléatoires b_k , $1 \leq k \leq P$, réelles, de première et seconde fonctions caractéristiques, [13], respectivement notées $\Phi_{\mathbf{b}}(\mathbf{z})$ et

$$(3.6) \quad \Psi_{\mathbf{b}}(\mathbf{z}) = \text{Log} [\Phi_{\mathbf{b}}(\mathbf{z})].$$

Les cumulants successifs de \mathbf{b} apparaissent dans les termes du développement en série de Mac-Laurin de $\Psi(\mathbf{z})$. On définit, en effet,

$$(3.7) \quad \text{CUM} [b_1^{\alpha_1}, \dots, b_p^{\alpha_p}] = (-j)^{\alpha_1 + \dots + \alpha_p} \frac{\partial^{\alpha_1}}{\partial z_1} \dots \frac{\partial^{\alpha_p}}{\partial z_p} \Psi(\mathbf{z})|_{\mathbf{z}=0}$$

Au regard de la définition (3.7) on déduit qu'à l'instar des moments, les cumulants sont des fonctions *multilinéaires et symétriques*, [1], [13]. Plus intéressante pour les préoccupations du paragraphe, sont les propriétés des *cumulants de variables indépendantes*. On déduit, en effet, du caractère additif

$$(3.8) \quad \Psi_{\mathbf{b}_1 \mathbf{b}_2}(\mathbf{z}) = \Psi_{\mathbf{b}_1}(\mathbf{z}_1) + \Psi_{\mathbf{b}_2}(\mathbf{z}_2)$$

lorsque \mathbf{b}_1 et \mathbf{b}_2 sont des vecteurs aléatoires, chacun de taille Q pour fixer les idées, à composantes indépendantes et où \mathbf{z} de taille $2Q$ s'écrit

$$(3.9) \quad \mathbf{z} = [\mathbf{z}_1^T \ \mathbf{z}_2^T]^T,$$

les deux propriétés :

PI1.

$$(3.10) \quad \text{CUM} [\mathbf{b}_1, \mathbf{b}_2] = 0$$

PI2.

$$(3.11) \quad \text{CUM} [\mathbf{b}_1 + \mathbf{b}_2] = \text{CUM} [\mathbf{b}_1] + \text{CUM} [\mathbf{b}_2].$$

Remarque R5. Par rapport à la relation de définition, (3.7), la notation utilisée dans (3.10) et (3.11) est légèrement différente, plus compacte, elle sous entend, bien sûr le cumulants relatif aux composantes du ou des vecteurs arguments.

Pour finir à propos des cumulants, on note que ceux d'ordre strictement supérieur à 2 de variables gaussiennes sont nuls, car la seconde fonction caractéristique est alors limitée à des termes quadratiques en \mathbf{z} , voir [2]. On précise, en fonction des moments, à toutes fins utiles, les cumulants jusqu'à l'ordre 4, de variables réelles qui s'obtiennent en partant de la définition

$$(3.12) \quad \text{CUM} [b_k] = E[x_k]$$

$$(3.13) \quad \text{CUM} [b_k b_m] = \text{COV} [b_k b_m]$$

$$(3.14) \quad \text{CUM} [b_k b_m b_r b_s] = E[b_k b_m b_r b_s] - E[b_k b_m] E[b_r b_s] - E[b_k b_r] E[b_m b_s] - E[b_k b_s] E[b_m b_r].$$

$E[\]$ désigne l'espérance mathématique et $\text{COV} [\]$ la covariance. On remarque alors que pour des variables aléatoires indépendantes, on a les deux conditions suivantes,

$$(3.15) \quad \text{CUM} (b_k b_m) = \text{VAR} (b_k) \delta(k, m)$$

$$(3.16) \quad \text{CUM} [b_k b_m b_r b_s] = \text{CUM} [b_k^4] \delta(k, m, r, s).$$

La quantité $\delta(k, m, r, s)$ étendant le symbole de Kronecker est toujours nulle sauf lorsque $k = m = r = s$, (elle vaut alors 1). En cohérence avec le discours de [28], on dira que (3.15) est la condition nécessaire et suffisante de blancheur ou d'indépendance d'ordre 2, cependant que (3.16) est la CNS de blancheur ou d'indépendance d'ordre 4, désignée, en accord avec l'introduction, par le sigle cI4.

Une cI4 complète revient, au vu de la propriété PI1, à annuler tous les cumulants croisés d'ordre 4, du vecteur des sources restituées, \mathbf{a} . En dehors de [4] et [8], les algorithmes fondés sur des cI4 testent l'annulation de seulement $P(P-1)/2$ cumulants croisés. Ceux déduits de cNL, voir l'introduction, annulent également, selon une procédure détaillée plus loin, $P(P-1)/2$ quantités. Dans les deux cas, cI4 ou bien cNL, cela coïncide très exactement avec le nombre de conditions susceptibles de supprimer les degrés de libertés inhérents à la matrice unitaire \mathbf{Q} qui caractérise la classe $\text{C2}(\mathbf{X})$, voir § 3.12.

Remarque R6. Il est établi dans [4] comment, du fait d'une certaine *redondance*, l'annulation de quelques cumulants croisés d'ordre 4 suffit à *diagonaliser totalement le tenseur cumulants* de même ordre. La référence [8] le diagonalise directement sans exploiter cette redondance. L'algorithme de Comon [4], nécessite donc *moins de calculs et de stockage en mémoire*, mais c'est évidemment l'effet de la compensation, est moins robuste que la diagonalisation de Cardoso [8], qui, à notre connaissance, est la seule méthode, du fait même de cette *robustesse*, capable d'identifier un nombre de sources supérieur au nombre de capteurs disponibles.

3.2.3. Conditions d'indépendance, retenues dans les approches de la séparation

Les conditions d'indépendance d'ordre 2, (c'est-à-dire portant sur les moments ou cumulants du second ordre), sont en fait, plus ou moins déjà contenues dans la donnée de \mathbf{E} , voir (3.5). En effet, l'écriture de la matrice de covariance de la partie utile du vecteur de données sous la forme $\mathbf{X}\mathbf{X}^T$, ($\mathbf{X}\mathbf{A}^2\mathbf{X}^T$ pour [4]), tient compte de la condition (3.15) avec, par hypothèse $\text{VAR} (b_k) = \text{VAR} (s_k) = 1$, ($\text{COV} (\mathbf{s}(t)) = \mathbf{A}^2$ pour [4]), voir (2.1). Les méthodes vont donc différer par

le choix de conditions d'indépendance devant lever la dégénérescence d'ordre $P(P-1)/2$ de la classe du second ordre, $\text{C2}(\mathbf{X})$, voir § 3.12.

Les cI4 portent sur les cumulants et les cNL sur les moments.

a. cI4 portant sur les cumulants, références [4], [5], [8], [9].

Ruiz et Lacoume, dans [5], chiffrent l'indépendance des

composantes du vecteur source restitué, \mathbf{a} , en imposant la condition

$$(3.17) \quad \text{CUM} [a_k a_k^* a_m^* a_m] = 0, \quad \text{pour } 1 \leq k < m \leq P$$

ce qui fournit bien les $P(P-1)/2$ relations manquantes. *Remarque R7.* La notion de cumulants s'étend simplement, jusqu'à l'ordre 4, à des variables aléatoires à valeurs complexes,

$$(3.18) \quad \text{CUM} [a_k a_r^* a_m^* a_s] = E[a_k a_r^* a_m^* a_s] - E[a_k a_r^*] E[a_m^* a_s] - E[a_s a_r^*] E[a_m^* a_k]$$

La relation (3.18) met en avant l'idée, selon laquelle, le cumulants d'ordre 4 mesure l'écart à la loi *gaussienne complexe circulaire* [20].

Dans [4], Comon établit qu'après une transformation linéaire régulière d'un vecteur de composantes mutuellement indépendantes, l'*indépendance par paire entraîne l'indépendance globale* et devient, par la même, une condition suffisante, a fortiori à l'ordre 4. La transformation unitaire \mathbf{Q} inconnue et destinée à blanchir à l'ordre 4 les sources restituées, peut être alors paramétrisée sous forme d'un produit de $P(P-1)/2$ rotations planes de Givens et d'un nombre complexe de module unité.

Chaque rotation élémentaire est déterminée de façon à générer une *paire* de variable aléatoire *indépendante à l'ordre 4*, c'est-à-dire telle que tous les cumulants croisés d'ordre 4 soient nuls,

$$(3.19) \quad \text{CUM} [|a_1|^2 a_1^* a_2] = 0$$

$$(3.20) \quad \text{CUM} [|a_2|^2 a_2^* a_1] = 0$$

$$(3.21) \quad \text{CUM} [|a_1|^2 a_2^2] = 0$$

Remarque R8. L'indépendance d'ordre 2 est atteinte par mise sous forme canonique, voir § 2.2.

Un seul *parcours* de $P(P-1)/2$ rotations planes, appelé en cohérence avec le vocabulaire de l'analyse numérique matricielle traditionnelle, [29], *balayage*, ou encore *salve*, (*sweep* en anglais), s'avère, bien sûr, *insuffisant* à l'*orthogonalisation totale* d'ordre 4. En effet, chaque nouvelle rotation plane, membre d'un balayage donné, perturbe la *géométrie d'ordre 4* établie par les rotations précédentes. Comon montre néanmoins que lorsque le nombre de balayage augmente, la somme des carrés de module des cumulants croisés standardisés, (cumulants de vecteurs canoniques au sens du § 2.2), de l'ensemble des sources restituées tend vers zéro. En d'autres termes, la procédure converge. Cardoso diagonalise complètement dans [8] le tenseur cumulants d'ordre 4, ce qui correspond aux liens révélés dans la remarque R6 avec l'algorithme de Comon. Pour finir, Gaeta et Lacoume, dans [9] spécifient la forme du développement de Gram-Charlier [30], de la première fonction caractéristique de \mathbf{a} à l'aide de (3.16).

Remarque R9. Le développement de Gram-Charlier de la première fonction caractéristique, $\Phi_{\mathbf{a}}(z)$, d'un vecteur aléatoire \mathbf{a} réel dont on connaît la matrice de covariance \mathbf{G} et les cumulants d'ordre 4 revient à approximer le second

ordre de $\Phi_{\mathbf{a}}(z)$ par une gaussienne et introduire un terme complémentaire qui tient compte de l'ordre 4. On a, dans ces conditions, voir [30] et [9],

$$(3.22) \quad \Phi_{\mathbf{a}}(z) = \exp * \left\{ - \sum_{k,m} g(k,m) a_k a_m \right\} [1 + (1/4!) \sum_{k,m,r,s} \text{CUM}[a_k, a_m, a_r, a_s] z_k z_m z_r z_s]$$

où $g(k,m)$ est le terme k,m de la matrice de covariance \mathbf{G} . On note, alors, en quoi (3.15) et (3.16) imposent une forme spécifique à (3.22).

b. Conditions Non Linéaires, cNL, références [2], [3], [6].

Jutten et Héroult figurent parmi les premiers, voir [3], à avoir abordé le problème de la séparation de sources, ceci dans le contexte du *calcul neuromimétique*. En supposant que les densités de probabilité des sources restituées a_k sont paires et en considérant des fonctions $f(\cdot)$ et $g(\cdot)$ impaires et fortement non linéaires, leur condition s'écrit

$$(3.23) \quad E[f(a_k)g(a_m)] = 0, \quad \text{pour } 1 \leq k < m \leq P$$

Pratiquement, leurs algorithmes ont été implémentés avec

$$(3.24) \quad f(\cdot) = (\cdot)^3$$

$$(3.25) \quad *g(\cdot) = \text{Arctg}(\cdot)$$

On peut justifier, en parties, le choix de (3.23) si l'on prend en considération les développements en série des fonctions $f(\cdot)$ et $g(\cdot)$ qui feront intervenir d'autant plus de termes avec un poids significatif, donc d'autant plus de moments croisés dans (3.23), qu'elles exhibent des non linéarités. Fety dans [2] fonde l'optimalité de sa méthode sur des arguments semblables, en l'occurrence les conditions

$$(3.26) \quad E[h(a_k)a_m] = 0, \quad \text{pour } 1 \leq k < m \leq P$$

où $h(\cdot)$ est choisie de façon à amplifier les écarts de puissance entre les sources. Plusieurs fonctions ont été étudiées à cet effet. La fonction « racine nième » semble apporter les meilleures performances. Cardoso, dans [6], conditionne la structure de la matrice de covariance des vecteurs signatures canoniques modifiés,

$$(3.27) \quad \tilde{\mathbf{y}}_k = \left[\sum_{k=1,P} |a_k|^2 \right]^{1/2} \mathbf{y}_k$$

où les \mathbf{y}_k sont les signatures canoniques, voir § 2.2, par les contraintes

$$(3.28) \quad E[a_i a_j^* |a_k|^2] = 0, \quad \text{si } i \neq j$$

$$(3.29) \quad E[a_i a_j^* |a_k|^2] = E[|a_k|^4], \quad \text{si } i = j = k$$

$$(3.30) \quad E[a_i a_j^* |a_k|^2] = 1, \quad \text{si } i = j \neq k$$

Là encore, lorsque i, j, k , varient, chacun entre 1 et P , l'ensemble des relations déduites de (3.28), (3.29) et

(3.30) permettent de déterminer, au préalable, (avant la séparation à proprement parler), la structure de la matrice de covariance de \tilde{y}_k .

4. Principes des méthodes de séparation

4.1. CONSIDÉRATIONS GÉNÉRALES

Rendu à ce stade, et d'après les contenus des paragraphes 3.14 et 3.23 on voit que la séparation de sources indépendantes se résume à la restitution d'un élément, noté \mathbf{a} , de la classe d'indépendance, CI(s), en laissant dans l'ombre une indétermination de phase et de permutation, voir § 3.11. Cette restitution doit, de plus, s'accompagner du lever de dégénérescence d'ordre $P(P-1)/2$ de la classe du second ordre, C2(X), voir § 3.12. Ceci se réalise, dans la pratique, en imposant des conditions d'indépendance, CI4 ou cNL, sur les composantes du vecteur restitué. En dehors des considérations précédemment énumérées telles que, le modèle, les hypothèses, la nature des données, (réelles ou complexes), le type de conditions testées, il reste encore un point important à développer qui différencie un peu plus, encore, les approches des références [2], [3] [4], [5], [6], [7], [8], [9]. Il s'agit de la façon dont on va obtenir \mathbf{a} . Les méthodes *indirectes* [6], [7], [8], et du maximum de vraisemblance approché, (*globales*), [8], traversent, au préalable, [6], [7], [8], ou simultanément, [8], (avant ou pendant la restitution de \mathbf{a}), une étape d'estimation de la matrice spatiale X. Les méthodes *directes* [2], [3], [4], [5], quant à elles, se contentent de séparer les sources. La distinction ébauchée ici n'a rien de systématique, (on peut imaginer estimer X, dans un deuxième temps, à l'aide d'une méthode directe), elle permet malgré tout de regrouper les algorithmes partageant quelques points communs afin d'unifier davantage les techniques auxquelles on a recours.

4.2. LES MÉTHODES DIRECTES [2], [3], [4], [5]

4.2.1. Matrice de séparation

Ces méthodes légitiment la recherche directe, sous les hypothèses HSI0, d'un vecteur \mathbf{a} de CI(s) sous la forme

$$(4.1) \quad \mathbf{a} = \mathbf{Hr}$$

par la *linéarité* de la relation entre \mathbf{r} et \mathbf{s} , voir (2.1). La matrice \mathbf{H} , inconnue, est appelée matrice de *restitution* ou de *séparation*. Le vecteur \mathbf{a} devant nécessairement appartenir à CI(s), on note, au vu du § 3.11, qu'il existe une matrice diagonale \mathbf{D} , (dont les termes non nuls valent ± 1), et une matrice de permutation \mathbf{P} , telles que

$$(4.2) \quad \mathbf{HX} = \mathbf{DP}$$

$$(4.22 \text{ bis}) \quad \mathbf{HXA} = \mathbf{DP}, \text{ pour [4],}$$

Une paramétrisation et une série de contraintes relatives à \mathbf{H} , différentes suivant les méthodes, permet de passer de P^2 inconnues à $P(P-1)/2$.

4.2.2. Paramétrisation et contraintes relatives à la matrice de séparation

a. Utilisation du calcul neuromimétique, [3].

Jutten et Héroult dans [3] fondent la paramétrisation de \mathbf{H} sur l'obtention de \mathbf{a} par un procédé récursif, appelé mémoire associative, très proche du fonctionnement de synapses [3], qui se traduit par l'équation

$$(4.3) \quad \mathbf{a}(t) = \mathbf{r}(t) - \mathbf{Na}(t)$$

$$(4.4) \quad \text{avec } \text{diag}(\mathbf{N}) = \mathbf{0}.$$

La relation (4.4) exprime le fait que tous les éléments diagonaux de \mathbf{N} , considérée comme une matrice de *rétroaction*, sont nuls. Les motivations « neuromimétiques » de (4.3) sont liées à la reconstruction possible de \mathbf{a} à partir de \mathbf{r} en se débarrassant de l'influence des autres sources par combinaison linéaire. Les relations (4.1), (4.3) et (4.4) se synthétisent sous forme du couple de formules, paramétrisation-contraintes,

$$(4.5) \quad \mathbf{H} = (\mathbf{I} + \mathbf{N})^{-1}$$

$$(4.6) \quad \text{diag}(\mathbf{H}^{-1}) = \mathbf{I}$$

(4.6) lève P degrés de liberté de \mathbf{H} . Les conditions nécessaires du second ordre,

$$(4.7) \quad E[a_k a_m] = \delta(k, m), \text{ pour } k \neq m$$

en suppriment, $P(P-1)/2$. Les $P(P-1)/2$ degrés résiduels sont levés, finalement, par les CNI, (3.23).

b. Décomposition de Cholesky de la matrice de covariance [4].

Dans [4], Comon, considère la structure

$$(4.8) \quad \mathbf{H} = \mathbf{AQ}^\dagger \mathbf{L}^{-1}$$

sous la contrainte

$$(4.9) \quad E[\mathbf{aa}^\dagger] = \mathbf{A}^2$$

\mathbf{Q} est une matrice de rotation et \mathbf{L} une matrice triangulaire inférieure. Ceci impose

$$(4.10) \quad \mathbf{R} = \mathbf{LL}^\dagger.$$

Par conséquent, l'utilisation des informations du second ordre, sous forme de la factorisation de Cholesky (4.10), de \mathbf{R} fixe les $P(P+1)/2$ termes de \mathbf{L} . Les $P(P-1)/2$ degrés de libertés restant, (\mathbf{Q}), sont levés en déterminant les $P(P-1)/2$ rotations de Givens qui paramétrisent \mathbf{Q} , voir § 3.23a.

c. Approche de Ruiz et Lacoume [5].

Aucune paramétrisation spécifique n'est préconisée dans [5]. La contrainte

$$(4.11) \quad \text{diag}(\mathbf{H}) = \mathbf{I}$$

fournit P relations. L'ordre 2 est utilisé comme dans le a., Jutten et Héroult) et les CNI d'ordre 4 sont les mêmes que celles du b., (Comon), en l'occurrence, (3.17). Une

planche récapitulative des caractéristiques des méthodes directes est fournie dans le paragraphe 5.

d. Approche de Fety, [2].

Fety impose aux sources restituées une variance unité. Ceci fige les $P(P+1)/2$ relations manquantes sous forme de la relation,

$$(4.12) \quad \mathbf{R}\mathbf{R}^T = \mathbf{I}.$$

4.3. MÉTHODES INDIRECTES [6], [7], [8] ET GLOBALES, [8]

4.3.1. Méthodes indirectes [6], [7], [8]

Ces méthodes visent à exploiter la décomposition spectrale [12], d'une matrice de moments d'ordre 4 [6] et [7], (dans la mesure où la différence, minime, entre [6] et [7] provient de la prise en compte supplémentaire d'un bruit additif dans [7], seul le contenu de [6] est détaillé), ou d'un tenseur cumulant de même ordre [8]. Le modèle canonique, voir § 2.2, est retenu.

a. Matrice de moments d'ordre 4, Cardoso M [6], [7].

On s'intéresse à [6] seulement. Sous les cNL (3.28, 29, 30) on constate sans peine que

$$(4.13) \quad \tilde{\mathbf{R}}_4 = \mathbf{E}[\tilde{\mathbf{Y}}\mathbf{a}\mathbf{a}^\dagger \tilde{\mathbf{Y}}^\dagger] = \sum_{k=1, P} (\mu_k + P - 1) \mathbf{y}_k \mathbf{y}_k^\dagger$$

où les \mathbf{y}_k sont les signatures spatiales canoniques orthonormées, voir § 2.2. La matrice $\tilde{\mathbf{Y}}$ réunit les vecteurs colonnes $\tilde{\mathbf{y}}_k$, voir (3.26), avec, de plus,

$$(4.14) \quad \mu_k = \mathbf{E}[|a_k|^4].$$

Par conséquent la décomposition spectrale de la matrice $\tilde{\mathbf{R}}_4$, fournit, non seulement les signatures canoniques, \mathbf{y}_k , (comme ils sont orthonormés, ce sont nécessairement les vecteurs propres de $\tilde{\mathbf{R}}_4$), mais encore les moments d'ordre 4 des sources. On passe aux signatures « standards », \mathbf{x}_k , voir § 2.1, en utilisant la relation

$$(4.15) \quad \mathbf{x}_k = \mathbf{C}^{1/2} \mathbf{y}_k$$

$\mathbf{C}^{1/2}$ étant défini par (2.7).

b. Tenseur cumulants d'ordre 4, Cardoso, C [8].

Des résultats semblables à ceux de [8] mais obtenus par une approche différente sont fournis ci-après. L'approche nouvelle proposée rejoint le formalisme de [21] et est fondée sur l'utilisation du produit de Kronecker [14], de 2 matrices. Étant données les matrices \mathbf{A} , de taille $[\mathbf{M}, \mathbf{N}]$, et \mathbf{B} de taille $[\mathbf{R}, \mathbf{Q}]$ on définit la matrice \mathbf{C} , de taille $[\mathbf{MR}, \mathbf{NQ}]$, produit de Kronecker de \mathbf{A} et \mathbf{B} , notée

$$(4.16) \quad \mathbf{C} = \mathbf{A} \otimes \mathbf{B}$$

par son bloc de rang i, j

$$(4.17) \quad [\mathbf{C}](i, j) = a_{ij} \mathbf{B}$$

Autrement dit, on multiplie tous les termes de la matrice \mathbf{B}

par la matrice \mathbf{A} . Une propriété, qui sera abondamment considérée dans la suite, s'écrit,

$$(4.18) \quad (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AD} \otimes \mathbf{BD}.$$

Pour finir, le produit de Kronecker n'est pas commutatif. On s'intéresse à la forme des cumulants d'ordre 4, collectés dans une matrice ou un tenseur [8], des composantes du vecteur observation canonique, voir § 2.2,

$$(4.19) \quad \mathbf{c} = \mathbf{Y}\mathbf{s}.$$

Pour ce faire, on introduit un vecteur observation d'ordre deux de taille P^2 ,

$$(4.20) \quad \mathbf{c}_2 = \mathbf{c} \otimes \mathbf{c}^*$$

qui d'après (4.18) et (4.19) se réécrit

$$(4.21) \quad \mathbf{c}_2 = (\mathbf{Y} \otimes \mathbf{Y}^*)(\mathbf{s} \otimes \mathbf{s}^*).$$

On considère, ensuite la matrice $[\mathbf{P}^2, \mathbf{P}^2]$, appelée matrice ou tenseur des cumulants d'ordre 4, dont les coefficients sont les cumulants des coefficients de la matrice

$$(4.22) \quad \mathbf{c}_2 \mathbf{c}_2^\dagger$$

notée, symboliquement,

$$(4.23) \quad \text{CUM} [\mathbf{c}_2 \mathbf{c}_2^\dagger].$$

La multilinéarité des cumulants, voir [1], entraîne, en respectant les notations de (4.23)

$$(4.24) \quad \text{CUM} [\mathbf{c}_2 \mathbf{c}_2^\dagger] = (\mathbf{Y} \otimes \mathbf{Y}^*) \times \text{CUM} [(\mathbf{s} \otimes \mathbf{s})(\mathbf{s}^\dagger \otimes \mathbf{s}^T)](\mathbf{Y}^\dagger \otimes \mathbf{Y}^T).$$

Au regard de (4.18), il vient,

$$(4.25) \quad \text{CUM} [\mathbf{c}_2 \mathbf{c}_2^\dagger] = (\mathbf{Y} \otimes \mathbf{Y}^*) \times \text{CUM} [(\mathbf{s}\mathbf{s}^\dagger \otimes \mathbf{s}^* \mathbf{s}^T)](\mathbf{Y}^\dagger \otimes \mathbf{Y}^T).$$

Sous l'hypothèse d'indépendance, il reste une expression qui généralise ce que l'on a à l'ordre deux,

$$(4.26) \quad \text{CUM} [\mathbf{c}_2 \mathbf{c}_2^\dagger] = \sum_{k=1, P} \gamma_k (\mathbf{y}_k \otimes \mathbf{y}_k^*) (\mathbf{y}_k^\dagger \otimes \mathbf{y}_k^T)$$

où, d'après le caractère canonique des signatures \mathbf{y}_k , voir § 2.2, on a

$$(4.27) \quad (\mathbf{y}_k^\dagger \otimes \mathbf{y}_k^T)(\mathbf{y}_j \otimes \mathbf{y}_j^*) = d(k, j)$$

et

$$(4.28) \quad \gamma_k = \text{CUM} [|s_k|^4].$$

En d'autres termes, la décomposition (4.26) est la représentation spectrale de la matrice $\text{CUM} [\mathbf{c}_2 \mathbf{c}_2^\dagger]$. Les vecteurs propres associés aux valeurs propres non nulles s'identifient aux signatures $(\mathbf{y}_k \otimes \mathbf{y}_k^*)$. L'intérêt de la démonstration proposée ici par rapport à celle de [8] est liée au fait qu'elle n'utilise pas d'algèbre tensorielle, (elle s'étend d'ailleurs, sans difficulté, à tout ordre), on trouvera une approche semblable dans [15].

4.3.2. Méthodes globales fondées sur le maximum de vraisemblance [9]

La matrice \mathbf{XQ} , qui intervient dans (3.5), étant paramétrisée sous la forme

$$(4.29) \quad \mathbf{XQ} = \mathbf{U}\Sigma\mathbf{V}$$

(\mathbf{U} et \mathbf{V} sont deux matrices unitaires et Σ une matrice diagonale), le bruit additif étant supposé blanc, gaussien de variance σ^2 , indépendant du signal utile, il s'agit d'estimer, au sens du maximum de vraisemblance, le jeu de paramètres déterministes $\xi = \{\mathbf{U}, \Sigma, \mathbf{V}, \sigma^2, \text{CUM}\}$, (si CUM représentent les cumulants d'ordre 4 des sources) pour un nombre de sources fixé égal à P . La résolution d'un tel problème [31], exige que l'on connaisse la densité de probabilité de l'observation,

$$(4.30) \quad \mathbf{r} = \mathbf{U}\Sigma\mathbf{V}\mathbf{v} + \mathbf{b}$$

notée $p(\mathbf{r}; \xi)$. L'hypothèse d'indépendance entre les sources et le bruit, conjuguée à la relation de Parseval [32], permet d'écrire

$$(4.31) \quad p(\mathbf{r}; \xi) = \int \Phi_{\mathbf{v}}(\mathbf{u}) \beta^*(\mathbf{r}, \mathbf{u}) d\mathbf{u}$$

où $\Phi_{\mathbf{v}}(\mathbf{u})$ est la fonction caractéristique de \mathbf{v} et avec

$$(4.32) \quad \beta(\mathbf{r}, \mathbf{u}) = \text{TF}[p_b(\mathbf{r} - \mathbf{U}\Sigma\mathbf{V}\mathbf{v})].$$

$p_b(\cdot)$ désigne la densité de probabilité du bruit. Comme il a été supposé gaussien, le terme $\beta(\mathbf{r}, \mathbf{u})$ s'obtient sans difficulté. La fonction $\Phi_{\mathbf{v}}(\mathbf{u})$, quand à elle fait l'objet d'un développement de Gram-Charlier [30], selon (3.23). Après quoi l'écriture de la log vraisemblance, (on se contente, ici, vu le caractère très « technique » des développements, de commenter qualitativement les résultats obtenus, le lecteur intéressé consultera la référence [9]), fait apparaître deux termes. L'un est identique à celui que l'on obtiendrait dans le cas gaussien qui permet d'estimer σ^2 , \mathbf{U} et Σ . Le second contient des informations supplémentaires liées à l'ordre 4. Sa prise en compte aboutit à \mathbf{V} et CUM. En résumé, cette méthode confirme, selon la démarche du maximum de vraisemblance et sous l'hypothèse d'un bruit gaussien, le rôle assuré par les moments d'ordre 4 pour lever les $P(P-1)/2$ degrés de libertés contenus ici dans la matrice unitaire \mathbf{V} .

5. Classification des méthodes

On donne dans ce paragraphe, sous forme de deux tableaux récapitulatifs, une présentation synthétique et comparative des méthodes *directes* [2], [3], [4], [5], d'une part et *indirectes* [6], [8], [9], d'autre part. Y figurent l'ensemble des caractéristiques des approches qui ont été découvertes et justifiées tout au long de l'article.

Les méthodes synthétisées sur le tableau 1 visent donc à restituer des sources indépendantes par une *transformation linéaire* agissant sur les observations, *fonction* des P^2

éléments d'une matrice \mathbf{H} . Les lignes « Information d'ordre 2 » et « contraintes » montrent comment on obtient déjà une série de relations, (fondées exclusivement sur les moments d'ordre deux), destinées à lever partiellement les degrés de liberté induits. Ce nombre est complété par $P(P-1)/2$ cNL, (conditions non linéaires), ou c14, (conditions d'indépendance d'ordre 4). La somme des nombres apparaissant sur une même colonne est, dans ces conditions, égale à P^2 .

Les méthodes données sur le tableau 2, qualifiées dans l'article d'indirectes et globales, produisent, en plus de la séparation, des caractéristiques supplémentaires des signaux, (la matrice des signatures par exemple), par *diagonalisation d'une matrice de moments* d'ordre 4 [6], d'un *tenseur cumulant d'ordre 4* [8], et, pour finir, à l'aide d'une méthode dite *maximum de vraisemblance approché* [19].

6. Conclusion

Le contenu de cet article apporte une contribution à la présentation de la séparation de sources indépendantes comme la restitution d'un élément d'une classe de signaux indépendants de variance unité, en laissant dans l'ombre une indétermination de signe et de permutation. Cette restitution doit, de plus, s'accompagner du lever de dégénérescence d'ordre $P(P-1)/2$ émanant de toute factorisation en racine carrée d'une matrice de covariance. Ceci se réalise, dans la pratique, en imposant des conditions d'indépendance à l'ordre 4, c14, ou non linéaires, cNL, sur les composantes du vecteur restitué. Les c14 portent sur les cumulants et les cNL prennent effet sur les moments. En dehors de considérations telles que, le modèle, les hypothèses, la nature des données (réelles ou complexes), le type de conditions mises en œuvre, un point important différencie un peu plus, encore, les approches des méthodes de séparation recensées à ce jour. Il s'agit de la façon dont on sépare les sources, à proprement parler. Les méthodes *indirectes* et du maximum de vraisemblance approché (*globales*) traversent, au préalable ou simultanément (avant ou pendant la restitution des sources), une étape d'estimation de l'information spatiale contenue dans les données. Les méthodes *directes* quant à elles se contentent de séparer les sources. La distinction ébauchée ici n'a rien de systématique, elle permet malgré tout de regrouper les algorithmes partageant quelques points communs afin d'unifier davantage les techniques auxquelles on a recours. Les méthodes *indirectes* exploitent systématiquement la décomposition spectrale d'une matrice de moments d'ordre 4 ou d'un tenseur de cumulant de même ordre. On note, au passage, qu'une justification nouvelle de la forme spécifique de la représentation spectrale d'un tenseur cumulant d'ordre 4 de variables indépendantes a été présentée. Ses différences par rapport à celle de [8] proviennent du fait que l'algèbre tensorielle d'ordre 4 n'est pas utilisée. Les méthodes *globales* ou du *maximum de vraisemblance*, sous l'hypothèse d'un bruit blanc gaussien additif indépendant

Tableau 1. Méthodes directes de séparation

Méthodes	[2] Féty	[3] Héroult	[4] Comon	[5] Lacoume
Mod. standard Mod. canonique	X_s 0	X_s 0	$X A s_n$ 0	X_s 0
HSIO	1	1	1	1
Don. réelles Don. complexes	1 0	1 0	0 1	0 1
Informations d'ordre 2	R	$E[a_k a_m] = \delta(k, m)$ $(k \neq m) P(P-1)/2$	$R = LL^T$ $P(P+1)/2$	$E[a_k a_m^*] = \delta(k, m)$ $(k \neq m) P(P-1)/2$
cNL $1 \leq k < m \leq P$ cI4	$E[h(a_k) a_m] = 0$ $P(P-1)/2$	$E[f(a_k) g(a_m)] = 0$ $P(P-1)/2$	$P(P-1)/2$ indépendance/paire	$P(P-1)/2$ $CUM[a_k ^2 a_m ^2] = 0$
Ap. directe restitution contraintes nombre	$HX = DP$ H $HRH^T = I$ $P(P+1)/2$	$HX = DP$ $(I+N)^{-1}$ $diag[H^{-1}] = I$ P	$HXA = DP$ $AQ^T L^{-1}$ $HRH^T = A^2$	$HX = DP$ H $diag[H] = I$ P

Tableau 2. Méthodes indirectes et globales

Méthodes	[6] Cardoso M	[8] Cardoso C	[9] Gaeta-Lacoume
Mod. standard Mod. canonique	Y_s	Y_s	$X_s + b$
Hsio Hsibbg	1 0	0 1	0 1
Don. réelles Don. complexes	0 1	0 1	1 0
Ordre 2	$XX^T = C^{1/2}(C^{1/2})^T$	$XX^T = C^{1/2}(C^{1/2})^T$	XX^T
cI4 moments Ordre 4 cI4 cumulants	$E[a_i a_j^* a_k ^2]$ (3.28, 29, 30)	$CUM[a_k a_m^* a_r a_s^*]$ (3.22)	$CUM[a_k a_m a_r a_s]$ (3.17)
Approche INDIRECTE Moments Approche INDIRECTE Cumulants	$E[a^T a Y a a^T Y^T]$ (4.13, 14)	$\sum (y_k \otimes y_k^*)(y_k^T \otimes y_k^T)$ (4.26)	
Approche GLOBALE Paramétrisation Algorithme			$XQ = USV$ (4.29) Gram-Charlier (3.22)

de données, développent la log vraisemblance des données en approximant la première fonction caractéristique d'une VA de covariance et de cumulants d'ordre 4 connus par un développement de Gram-Charlier. Deux termes apparaissent alors. Le premier qui contient les informations du second ordre coïncide avec l'écriture relative à l'hypothèse gaussienne. Le second, d'ordre 4 lève la dégénérescence évoquée plus haut. Les méthodes *indirectes* restituent les sources à partir des observations par une opération linéaire que matérialise une matrice de séparation. Une paramétri-

sation de celle-ci, assortie de quelques contraintes diminue le nombre d'inconnues à rechercher. On distingue les paramétrisations fondées sur le calcul neuromimétique et celle déduite d'une factorisation sous forme du produit d'une matrice de rotation et d'une matrice triangulaire. Toutes les méthodes présentées dans cet article, en dehors de celles relatées par les références [7] et [9], ne tiennent pas compte du bruit. Il semble néanmoins, pour le moment, que l'extension au modèle bruité n'aille pas sans poser de problèmes sérieux car la plupart des méthodes de

[3], [4], [5], [6] et [7] testées en présence de bruit n'ont pas encore atteint des performances vraiment fiables. Un effort tout particulier devrait donc, très prochainement, aller vers la prise en compte d'informations bruitées, même au quatrième ordre (considérer un bruit gaussien additif n'introduit aucune modification à l'ordre 4 dans la mesure où ses cumulants de rang quatre sont identiquement nuls).

BIBLIOGRAPHIE

- [1] D. R. BRILLINGER, « An introduction to polyspectra », *Ann. Math. Stat.*, Vol. 36, pp. 1351-1374, October 1965.
- [2] L. FETY, *Méthodes de Traitement d'antenne adaptées aux radio-communications*, thèse de doctorat, ENST, Paris, juin 1988.
- [3] C. JUTTEN et J. HÉRAULT, « Une solution neuromimétique au problème de séparation de sources », *TS*, Vol. 5, n° 6, pp. 389-403, 1988.
- [4] P. COMON, « Analyse en composantes Indépendantes et Identification aveugle », *Numéro spécial Non linéaire-Non gaussien*, *TS*, à paraître novembre 1990.
- [5] J. L. LACOUME et P. RUIZ, « Source identification, a solution based on the cumulants », *Proc. of the 4-th ASSP Workshop on spectral estimation and modeling*, pp. 199-203, Aug. 1988.
- [6] J. F. CARDOSO, « Sources separation using higher order moments », *Proc. of ICASSP, Glasgow, 1989*.
- [7] V. C. SOON and L. TONG, « An extended FOBI algorithm in spatially correlated noise », *Proc. of ICASSP*, pp. 1365-1368, Albuquerque, avril 1990.
- [8] J. F. CARDOSO, « Blind identification of Independent components with higher-order statistics », *Proc. of Workshop on higher order spectral analysis*, pp. 157-162, Vail, June 1989.
- [9] M. GAETA et J. L. LACOUME, « Estimateurs du maximum de vraisemblance étendus à la séparation de sources non-gaussiennes », *TS, Numéro spécial Non linéaire Non gaussien*, 1990.
- [10] R. O. SCHMIDT, *A signal subspace approach to multiple emitter location and spectral estimation*, Ph. D. dissertation, Stanford University, Cal.
- [11] R. KUMARESAN and D. TUFTS, « Estimating the angles of arrival of multiple plane waves », *IEEE trans. Aerospace and Electronic Systems*, Vol. AES-19, pp. 134-139.
- [12] M. MARCUS, basic theorems in Matrix theory, National Bureau Standards, *Applied Math. Ser.* 57, January 22, 1960.
- [13] P. BREMAUD, *Introduction aux probabilités*, Springer, New York, 1984.
- [14] A. GRAHAM, *Kronecker Products and Matrix Calculus Applications*, John Wiley and Sons, New York 1981.
- [15] J. F. CARDOSO, « Blind Independent Component analysis », *Proc. of ICASSP*, Albuquerque, avril 1990.
- [16] B. PORAT and B. FRIEDLANDER, « Direction finding Algorithms based on high-order statistics », *Proc. of ICASSP*, pp. 2675-2678, Albuquerque, avril 1990.
- [17] J. F. CARDOSO et P. COMON, *Tensor based Independent Component Analysis*, to be presented, EUSIPCO, Barcelona, Septembre 1990.
- [18] L. SHWARTZ, *Les Tenseurs*, Masson, Paris 1981.
- [19] B. PICINBONO, *Éléments de Théorie du Signal*, Dunod, Paris, 1981.
- [20] B. PICINBONO, *Traitement Statistique du Signal*, polycopié de cours ESE, 1989.
- [21] P. CHEVALIER, P. DUVAUT, B. PICINBONO, « Le filtrage de Volterra transverse en traitement du signal », *TS, Numéro spécial Non linéaire Non gaussien*, 1990.
- [22] J. F. CARDOSO, « Localisation par la quadricovariance », *TS, Numéro spécial Non linéaire Non gaussien*, 1990.
- [23] Y. BRESLER, « Exact maximum likelihood parameter estimation of superimposed exponential signals in noise », *IEEE trans.* October 1986, ASSP-34, pp. 1081-1089.
- [24] P. STOICA, « Novel Eigenanalysis method for direction estimation », *IEE Proc.*, Vol. 137, No. 1, Feb. 1990.
- [25] A. BLANC-LAPIERRE et B. PICINBONO, *Fonctions Aléatoires*, Masson, Paris, 1981.
- [26] O. L. FROST, « An algorithm for linearly constrained Adaptive Array processing », *Proc. of the IEEE*, 60, No 8, pp. 926-935, August 1972.
- [27] S. HAYKIN, *Adaptive Filter theory*, Prentice Hall, New Jersey, 1989.
- [28] P. BONDON et B. PICINBONO, « De la blancheur et de ses transformations », *TS, Numéro spécial Non linéaire Non gaussien*, 1990.
- [29] P. LASCAUX, *Analyse Numérique Matricielle Appliquée à l'art de l'Ingénieur*, Masson, Paris, 1986.
- [30] A. ANGOT, *Compléments de Mathématiques*, Masson, Paris, 1982.
- [31] H. L. VAN TREES, *Detection, Estimation and Modulation theory*, part 1, Wiley, New York, 1968.
- [32] A. RODDIER, *Distributions et Transformation de Fourier*, Mac-Graw-Hill, New York, 1985.

Manuscrit reçu le 20 mars 1990