

L *a calibration d'antenne*

Identification aveugle aux ordres supérieurs

Blind Identification with higher order Statistics

par **J.-F. CARDOSO**

Télécom Paris, CNRS URA 820, GdR TdSI
46 rue Barrault, 75634 Paris Cedex 13

Résumé

Cette contribution propose une approche 'aveugle' au problème du traitement d'antennes mal calibrées en montrant qu'il est possible d'identifier les vecteurs directionnels même lorsque la variété d'antenne est complètement inconnue. Nous utilisons en contrepartie l'hypothèse d'indépendance statistique des émetteurs. Sous l'hypothèse de bande étroite, cette propriété est mise à profit en exploitant les cumulants d'ordre quatre des signaux reçus. Le mécanisme d'identification proposé ici repose sur un blanchiment spatial des observations suivi d'une 'diagonalisation conjointe' de matrices cumulantes. Ce dernier procédé permet de combiner l'efficacité numérique des méthodes algébriques et la stabilité des méthodes d'optimisation de contraste.

Mots clés : statistiques d'ordre supérieur, identification aveugle, séparation de sources.

Abstract

A 'blind approach' is proposed to address the issue of ill calibration. It is shown that, even when the array manifold is completely unknown, the directional vectors can be identified whenever the source signals can be assumed statistically independent and not normally distributed. The fourth-order cumulants of the received signals are exploited to this effect in a two-step procedure : the array output is first spatially whitened, some cumulant matrices of the resulting signals then are 'jointly diagonalized'. This procedure combines the numerical efficiency of algebraic identification methods with the stability of contrast optimization methods.

Key words : higher order statistics, blind identification, source separation.

1. Introduction

L'identification aveugle consiste, pour des observations obéissant au modèle linéaire $x = y + b = As + b$ du traitement d'antenne à bande étroite, que nous écrivons plus volontiers

$$x = y + b = \sum_{p=1,P} s_p a_p + b, \quad (1)$$

à estimer *sans la paramétrer* la matrice A ou encore ses colonnes a_p , *i.e.* les vecteurs directionnels. Puisque cette matrice représente la propagation entre les sources et les capteurs ainsi que l'instrumentation qui fournit finalement les signaux, on voit que l'absence de paramétrisation de cette matrice revient à abandonner toute velléité de modéliser les phénomènes physiques contribuant à la formation du signal. On ne conserve comme hypothèse de structure que le seul modèle linéaire (1) tandis qu'à l'information de

paramétrisation, on substitue celle portée par les statistiques d'ordre supérieur, exploitée sous l'hypothèse essentielle d'*indépendance statistique des signaux émis par les sources*.

Selon la nature des informations que l'on souhaite extraire des données reçues sur un réseau déformé, l'identification aveugle apporte des contributions différentes.

- Lorsque l'on s'intéresse, en télécommunications par exemple, aux signaux émis, l'identification de la déformation de l'antenne n'est pas une fin en soi. En effet, la connaissance de la matrice A suffit pour séparer les différents signaux reçus sur l'antenne. Ce type de problème aveugle est appelé *séparation de sources* dans la littérature récente, et l'identification aveugle en fournit une solution immédiate.
- Par contre, lorsque l'on s'intéresse à la localisation de sources, c'est-à-dire à l'estimation de paramètres physiques, il faut nécessairement recourir à une modélisation de la propagation. À

ce niveau, l'identification aveugle, lorsqu'elle est possible, peut constituer une première étape avantageuse puisqu'elle fournit individuellement les colonnes de la matrice A . Nul doute qu'il ne soit plus simple d'ajuster les paramètres physiques sur ces vecteurs plutôt que sur des données composites où l'ensemble de tous les paramètres interviennent.

Nous n'exposerons pas ici plus avant les liens entre identification aveugle et calibration. L'essentiel de cet article est consacré à la description d'une famille d'algorithmes utilisant conjointement des statistiques cumulantes d'ordre 2 et 4 pour l'identification aveugle.

Le choix des statistiques cumulantes d'ordre 2 et 4 se justifie de la façon suivante. Tout d'abord, les statistiques cumulantes d'ordre 2 – à savoir la covariance des observations – ne permettent pas à elles seules d'identifier en aveugle le modèle linéaire. Il faut donc nécessairement les compléter par un autre type d'information. Notre principe étant de renoncer à l'information paramétrique, on cherchera cette information dans les données elles-mêmes. Il est naturel de recourir aux statistiques cumulantes car leurs propriétés de linéarité et d'additivité correspondent bien au problème linéaire qui nous intéresse : la structure linéaire du modèle induit sur les cumulants une structure multilinéaire que nous mettrons à profit dans l'identification. Par ailleurs, des raisons de robustesse commandent de considérer des cumulants d'ordre aussi bas que possible. Or, en traitement d'antenne, les signaux étant souvent distribués circulairement¹, leurs statistiques d'ordre impair sont nulles. Ceci impose le choix des cumulants d'ordre quatre qui, par ailleurs, s'avèrent suffisants pour assurer l'identifiabilité en aveugle.

Plusieurs auteurs ont proposé des algorithmes d'identification aveugle fondés sur les cumulants empiriques d'ordres 2 et 4. L'approche de Lacoume et Gaeta[0] exploite un développement de Gram-Charlier de la vraisemblance des observations dont les premiers termes donnent une fonctionnelle des cumulants d'ordre 2 et 4. On peut aussi, comme le fait Comon[0], construire directement une fonctionnelle de contraste dépendant des cumulants d'ordre 2 et 4. Dans ces deux approches, comme dans celle discutée plus bas, on aboutit à un mécanisme analogue en deux étapes. Les cumulants d'ordre 2 (*i.e.* les éléments de la covariance) sont d'abord utilisés pour blanchir les observations; cette opération transforme corrélativement la matrice A en une matrice unitaire ou orthogonale. Cette matrice unitaire est ensuite identifiée comme celle qui maximise un certain critère dépendant de cumulants du quatrième ordre des observations blanchies. Un avantage essentiel de l'approche proposée ici réside dans l'implantation de la recherche d'une matrice unitaire. Celle-ci est obtenue dans une technique semblable à la technique de Jacobi pour la diagonalisation dont la complexité algorithmique est (essentiellement) inchangée lorsque l'on traite des données à valeurs complexes. Ce point est évidemment important pour le traitement d'antenne

1. C'est-à-dire que leur distribution de probabilité est invariante dans une rotation de phase.

à bande étroite. Signalons qu'il est possible [0] de n'utiliser que les cumulants d'ordre 4 pour l'identification, ce qui dispense de modéliser le bruit additif si celui-ci est gaussien. Signalons enfin que le problème 'dual' de séparation de sources est traité –mais dans le seul cas de données réelles– dans [0] et dans [0] comme un problème de filtrage adaptatif non linéaire, un peu à la façon de la déconvolution aveugle.

1.1. HYPOTHÈSES POUR L'IDENTIFICATION EN AVEUGLE

L'identification aveugle repose sur une hypothèse fondamentale : l'indépendance statistique des signaux sources, que nous supposons acquise dans toute la suite. Pour exploiter cette hypothèse au delà de l'ordre 2, il faut des signaux non gaussiens. Plus précisément, pour exploiter les cumulants d'ordre 4 comme nous le faisons dans la suite, nous supposons que chaque signal source présente un kurtosis non nul. Dans le cas complexe, qui nous intéresse ici, le kurtosis k_p de la p -ième source est défini par

$$k_p \stackrel{\text{def}}{=} \text{Cum}(s_p, s_p^*, s_p, s_p^*). \quad (2)$$

Rappelons que ce cumulants du quatrième ordre voit sa définition réduite à

$$\text{Cum}(s_p, s_p^*, s_p, s_p^*) = E\{|s_p|^4\} - 2E^2\{|s_p|^2\} \quad (3)$$

pour des signaux circulaires (en fait de circularité, il suffit ici que la distribution de s_p vérifie $E s_p = 0$ et $E s_p^2 = 0$).

Le bruit additif est supposé indépendant des signaux et normalement distribué. Sous cette dernière hypothèse, ses cumulants du quatrième ordre sont nuls de telle sorte qu'il n'est pas besoin d'en modéliser la structure spatiale. Par contre, le bruit contribue aux cumulants du second ordre qui ne sont donc exploitables que si la covariance de ce dernier peut être estimée d'une façon ou d'une autre (par exemple, en la supposant de la forme $\Gamma_b = \sigma^2 I$ et en estimant classiquement σ^2 à partir des petites valeurs propres de la covariance empirique).

Ces hypothèses statistiques sont évidemment destinées à placer l'identification aveugle dans une situation où sont disponibles des estimées consistantes des cumulants d'ordre 2 et 4 de Y , le signal 'utile'². Nous les complétons par une hypothèse géométrique faible : les colonnes de A sont linéairement indépendantes.

Avant de poursuivre, il importe de remarquer que, bien que l'objectif affiché soit d'identifier la matrice A , il est clair qu'une identification complète n'est pas possible en aveugle. En effet, le signal reçu est invariant dans l'échange d'un facteur complexe quelconque entre un vecteur a_p et le signal correspondant s_p puisque

2. Les hypothèses permettant l'estimation de la covariance du bruit peuvent être levées en n'exploitant que l'information du quatrième ordre qui garantit encore l'identifiabilité. Nous n'aborderons pas ici ces techniques d'identification au quatrième ordre seulement.

seul le produit $s_p a_p$ apparaît dans le signal observé, ce dernier constituant l'unique source d'information en identification aveugle. Ainsi, en l'absence de paramétrisation des vecteurs a_p , on ne peut faire mieux que d'identifier chacun de ces vecteurs à un facteur près. Nous profitons de cette indétermination fondamentale pour supposer, sans perte de généralité, que chaque source est de puissance unité :

$$\forall p = 1, \dots, P \quad \sigma_p^2 = 1. \quad (4)$$

Cette convention laisse indéterminée la phase de chaque vecteur a_p . Dans toute la suite, lorsque nous parlons de l'identification de A ou de ses colonnes, il est entendu qu'il s'agit d'une identification à une phase près sur chaque colonne a_p .

2. Identification par diagonalisation de matrices cumulantes

Nous définissons dans cette section la notion de 'matrice cumulante', qui sera l'outil de base pour représenter l'information du quatrième ordre. Dans le modèle de sources indépendantes, la structure très simple de ces matrices permet de ramener l'identification à un problème de diagonalisation.

2.1. MATRICES CUMULANTES

Bien que les mécanismes sous-jacents aux techniques d'identification proposées plus loin soient tensoriels et que la nature même des statistiques cumulantes d'ordre supérieur soit aussi tensorielle, il est possible ici (car nous ne considérons que le quatrième ordre) de présenter des algorithmes et des implantations qui ne fassent appel qu'à l'algèbre traditionnelle en 'matrice-vecteur'. Nous utiliserons les *matrices cumulantes*, dont la définition pour des signaux complexes circulaires est des plus élémentaires : à un vecteur aléatoire complexe circulaire x de covariance Γ_x et à toute matrice W de même taille que Γ_x , on associe la matrice cumulante $Q_x(W)$ définie par

$$Q_x(W) \stackrel{\text{def}}{=} E\{x^* W x x x^*\} - \Gamma_x W \Gamma_x - \Gamma_x \text{Trace}(W \Gamma_x). \quad (5)$$

Si W est choisie hermitienne, alors $Q_x(W)$ l'est aussi par construction.

On montre en appendice la relation étroite qui lie une matrice cumulante $Q_x(W)$ au tenseur des cumulants du quatrième ordre de x . Cette relation permet, si nécessaire, d'étendre au cas non circulaire la définition (5) en la modifiant de façon naturelle. Le rapport entre (5) et les cumulants explique la dénomination choisie et permet d'établir simplement la structure que prend, sous nos hypothèses, toute matrice cumulante d'un vecteur $x = As + b$:

$$Q_x(W) = \sum_{p=1, P} k_p a_p^* W a_p a_p a_p^*. \quad (6)$$

Cette structure est à comparer à celle de la covariance :

$$\Gamma_x = \sum_{p=1, P} \sigma_p^2 a_p a_p^* + \Gamma_b. \quad (7)$$

Une matrice cumulante possède donc une structure semblable à celle d'une covariance non bruitée, sans toutefois présenter la propriété d'être nécessairement non négative, puisque les kurtosis k_p et les termes quadratiques $a_p^* W a_p$ peuvent être de signes quelconques.

2.2. PRINCIPE DE L'IDENTIFICATION

Le principe d'identification aveugle que nous développons ici s'appuie sur l'information du second ordre (la covariance) pour orthogonaliser le signal utile y . En effet, toute matrice B de taille $P \times M$ vérifiant

$$B \Gamma_y B^* = I_P \quad (8)$$

permet de blanchir la variable x en ce sens que la variable z de dimension P , définie par

$$z \stackrel{\text{def}}{=} Bx = By + Bb \quad (9)$$

présente une partie signal By qui, par (8), est de covariance identité. Le blanchiment de x en z préserve toutes les propriétés 'linéaires' de x . En particulier, le bruit additif Bb reste gaussien, tandis que la partie signal By est toujours une somme de P composantes indépendantes :

$$z = \sum_{p=1, P} s_p c_p + Bb \quad \text{avec} \quad c_p \stackrel{\text{def}}{=} B a_p. \quad (10)$$

Par conséquent, la structure (6) est aussi vérifiée par toute matrice cumulante de z . On obtient donc

$$Q_z(W) = \sum_{p=1, P} k_p c_p^* W c_p c_p c_p^*. \quad (11)$$

Par ailleurs, le vecteur By (c'est-à-dire la partie signal de z) étant blanchi, sa covariance est l'identité :

$$\Gamma_{By} = I_P = \sum_{p=1, P} c_p c_p^* \quad (12)$$

où la dernière équation résulte de (4) et (8). Cette propriété (12) montre que les vecteurs blanchis c_p forment une base orthonormée de l'espace signal. Puisque les c_p sont orthonormés, la structure (11) peut encore être interprétée comme une décomposition propre où, pour toute matrice W , chaque c_p est un vecteur propre de $Q_z(W)$ associé à la valeur propre $k_p c_p^* W c_p$. A la base de nos algorithmes d'identification se trouve donc cette propriété que l'on peut encore formuler comme "toute matrice cumulante de z est diagonale dans la base des c_p ". Il est donc possible d'identifier le modèle linéaire via la diagonalisation d'une matrice cumulante. Ceci soulève cependant plusieurs questions dont la première est

celle de l'unicité de la diagonalisation : la base des c_p est elle la seule base de diagonalisation d'une matrice cumulée ?

Rappelons tout d'abord que dans le cas complexe, les vecteurs propres ne sont jamais déterminés qu'à une phase près. Cette indétermination correspond à celle de la phase des colonnes de A , déjà mentionnée et qui ne porte pas ici à conséquence. On sait par contre qu'une vraie indétermination des vecteurs propres apparaît lorsque les valeurs propres sont dégénérées. Les valeurs propres $Q_z(W)$ étant égales aux $k_p c_p^* W c_p$, il n'est pas possible, sans connaître les kurtosis k_p ni les vecteurs c_p , de leur assigner des valeurs toujours distinctes par un choix *a priori* de la matrice W .

Pour parer à une telle éventualité, nous proposons d'estimer les c_p par une *diagonalisation conjointe* d'un ensemble de matrices cumulantes. Nous faisons alors d'une pierre, deux coups. D'une part, la diagonalisation de plusieurs matrices cumulantes peut assurer l'identifiabilité (voir la section 2.4.). D'autre part, cette approche a pour effet d'augmenter la quantité totale d'information du quatrième ordre traitée et par conséquent d'augmenter l'efficacité statistique de l'estimation. Il nous faut toutefois définir la notion de diagonalisation conjointe. En effet, les erreurs d'estimation ne permettent pas en pratique d'obtenir des statistiques exactes : on ne peut donc espérer blanchir exactement la partie signal du processus, ou encore trouver une base où un ensemble de matrices cumulantes empiriques seraient exactement simultanément diagonales. Il faut alors considérer une diagonalisation conjointe *approchée*.

2.3. DIAGONALISATION CONJOINTE APPROCHÉE

La diagonalisation d'une matrice consiste à trouver une base orthonormée où les éléments hors diagonale sont nuls. De façon équivalente, une telle base maximise la somme des modules carrés des éléments diagonaux. Le i -ème élément diagonal d'une matrice M de taille $P \times P$ dans une base orthonormée $D = \{e_i | i = 1, P\}$ étant donné par $e_i^* M e_i$, une base de diagonalisation pour la matrice M maximise donc le critère

$$C(M, D) \stackrel{\text{def}}{=} \sum_{i=1, P} |e_i^* M e_i|^2 \quad (13)$$

sur l'ensemble des bases orthonormées. Nous définissons alors une base de diagonalisation conjointe approchée d'un ensemble $\{M_k | k = 1, K\}$ de K matrices comme toute base orthonormée maximisant le critère

$$C(D) \stackrel{\text{def}}{=} \sum_{k=1, K} w_k C(M_k, D) = \sum_{k=1, K} w_k \sum_{i=1, P} |e_i^* M_k e_i|^2 \quad (14)$$

où l'on se réserve la possibilité de pondérer ce critère composite par une famille de poids positifs : $\{w_k | w_k > 0, k = 1, K\}$. La question de l'unicité d'une telle base est maintenant discutée dans le cas particulier qui nous intéresse.

2.4. IDENTIFIABILITÉ

Nous donnons une condition suffisante d'identifiabilité qui est ici équivalente à l'unicité de la diagonalisation conjointe. Nous définissons le 'FOSS' (Fourth Order Signal Subspace) comme l'ensemble de toutes les matrices cumulantes :

$$\text{FOSS} \stackrel{\text{def}}{=} \{M | M = Q_z(W)\}. \quad (15)$$

Le FOSS n'est pas autre chose que l'espace image de Q_z vu comme un opérateur sur les matrices. Puisqu'il est toujours possible de choisir une matrice W telle que les P nombres $k_p c_p^* W c_p$ prennent des valeurs arbitraires, le FOSS est aussi, compte tenu de (11), l'ensemble des combinaisons linéaires des $c_p c_p^*$:

$$\text{FOSS} = \{M | M = \sum_{p=1, P} \alpha_p c_p c_p^*, \alpha_p \in \mathcal{C}\}. \quad (16)$$

On peut encore dire que les matrices $c_p c_p^*$ forment une base (d'ailleurs orthonormée) du FOSS. Il est donc clair que le FOSS est un espace linéaire de matrices et que sa dimension est égale au nombre P des sources.

Soit alors une base orthonormée $\{d_p | p = 1, P\}$ qui diagonalise tous les membres d'une famille $\{M_k | k = 1, K\}$ de K matrices cumulantes. Chaque matrice M_k admet les d_p comme vecteurs propres et sa décomposition propre permet de l'exprimer comme une combinaison linéaire des $d_p d_p^*$. Supposons (c'est la condition cherchée) que la famille des matrices M_k engendre linéairement le FOSS. Alors toute matrice du FOSS – en particulier chaque $c_p c_p^*$ – est combinaison linéaire des M_k , qui sont elles-mêmes combinaisons linéaires des $d_p d_p^*$. On peut donc trouver, pour tout q , des coefficients λ_p tels que

$$c_q c_q^* = \sum_{p=1, P} \lambda_p d_p d_p^*. \quad (17)$$

Le membre de gauche est de rang un, tandis que celui de droite est de rang égal au nombre de coefficients non nuls. Ce nombre est donc égal à 1, ce qui montre que chaque $c_q c_q^*$ est nécessairement égal à un $d_p d_p^*$. On obtient donc l'unicité recherchée de la diagonalisation conjointe des M_k sous la condition suffisante que les matrices cumulantes M_k engendrent le FOSS. Si l'on se contente de travailler avec un nombre de matrices cumulantes égal au nombre de sources (qui est aussi la dimension du FOSS), cette condition est encore équivalente à celle de l'indépendance linéaire des matrices cumulantes puisque, par définition, les matrices cumulantes sont dans le FOSS. Pour une condition plus faible, voir [0].

3. Implantation

Le schéma général de l'algorithme, dont chaque étape est décrite plus loin, est le suivant :

1. Estimation du nombre de sources et du blanchisseur \hat{B} .
2. Blanchiment des données $\hat{z}(t) = \hat{B}x(t)$, $t = 1, \dots, T$.
3. Estimation de K matrices cumulantes \hat{M}_k , $k = 1, \dots, K$ de \hat{z} .
4. Diagonalisation conjointe approchée des matrices cumulantes dans une base $\{\hat{c}_p | p = 1, P\}$.
5. Les estimées \hat{a}_p des colonnes de A sont données par $\hat{a}_p = \hat{B}^\# \hat{c}_p$.

Quelques remarques :

- L'étape de blanchiment est indépendante du reste de l'algorithme. Elle dépend d'ailleurs des hypothèses sur la structure spatiale du bruit.
- Rappelons que l'étape de blanchiment ramène la dimension de l'observation à la valeur P du nombre de sources. Par conséquent, l'estimation des matrices cumulantes se fait dans un espace de dimension réduite à P .
- Les matrices cumulantes peuvent aussi être obtenues comme 'matrices propres' du tenseur des cumulants empiriques (voir la section 3.2.).

3.1. IMPLANTATION DU BLANCHIMENT

Nous donnons ici une implantation classique du blanchiment dans le cas où la covariance du signal est obtenue par soustraction à la covariance empirique d'une estimée de la covariance du bruit, supposée de la forme $\sigma^2 I$. On forme donc d'abord, sur la base de T échantillons du signal reçu, une estimée $\hat{\Gamma}_x$ de la covariance empirique

$$\hat{\Gamma}_x = \frac{1}{T} \sum_{t=1, T} x(t)x^*(t) \quad (18)$$

que l'on diagonalise en

$$\hat{\Gamma}_x = \sum_{m=1, M} \lambda_m v_m v_m^* \quad (19)$$

La puissance du bruit peut classiquement être estimée par

$$\hat{\sigma}^2 = \frac{1}{M - P} \sum_{m > P} \lambda_m \quad (20)$$

si les éléments propres sont indexés en ordre décroissant. Blanchiment et restriction à l'espace signal sont alors obtenus en projetant x sur les vecteurs

$$b_p \stackrel{\text{def}}{=} (\lambda_p - \hat{\sigma}^2)^{-1/2} v_p \quad \forall p = 1, P \quad (21)$$

c'est-à-dire que la p -ième composante z_p de z est calculée par $z_p = b_p^* z$. Insistons sur le fait que les cumulants s'estiment dans l'espace de z et donc pour un vecteur de dimension P seulement. La matrice \hat{B} de blanchiment peut se construire explicitement en prenant b_p pour p -ième colonne de \hat{B}^* .

3.2. ESTIMATION DES MATRICES CUMULANTES

Nous proposons deux approches pour obtenir un jeu de matrices cumulantes. La première s'appuie naturellement sur la définition (5) et suppose un choix *a priori* de K matrices hermitiennes de pondération $\{W_k | k = 1, K\}$. L'estimateur naturel des K matrices cumulantes correspondantes est alors

$$\hat{M}_k = \frac{1}{T} \sum_{t=1, T} \hat{z}^*(t) W_k \hat{z}(t) \hat{z}(t) \hat{z}^*(t) - \hat{\Gamma}_z W_k \hat{\Gamma}_z - \hat{\Gamma}_z \text{Trace}(W_k \hat{\Gamma}_z) \quad (22)$$

avec

$$\hat{\Gamma}_k = \frac{1}{T} \sum_{t=1, T} \hat{z}(t) \hat{z}^*(t) \quad (23)$$

Cette première approche est d'implantation simple et peu coûteuse : le coût est proportionnel à K fois celui de l'estimation d'une covariance ordinaire (car pour chaque échantillon, le produit $\hat{z}(t) \hat{z}^*(t)$ n'est calculé qu'une fois pour toutes les matrices cumulantes. La trace du produit de ce terme avec chaque W_k donne le terme quadratique. Si W_k est une matrice creuse, ce dernier coût est négligeable). De plus, le choix du nombre K de matrices cumulantes détermine le volume de statistiques à estimer, ce qui permet un compromis entre efficacité statistique et coûts d'estimation/calcul.

Toutefois, nous n'avons pas trouvé de principe permettant de choisir les matrices de pondération W_k et leur nombre de façon à assurer à coup sûr l'identifiabilité, à moins de choisir un ensemble de P^2 matrices de pondération W_k linéairement indépendantes. Dans la pratique, on peut choisir par exemple

- $K = 1$ et $W_1 = I$: on obtient l'algorithme FOBI [0] dans sa forme étendue [0], d'une très grande simplicité puisqu'il se ramène à la diagonalisation de la seule matrice $E(|z|^2 z z^*) - \Gamma_z^2 - \Gamma_z \text{Trace}(\Gamma_z)$. Typiquement, l'identification n'est pas possible pour des sources de kurtosis identiques.
- $K = P$ avec $W_k = \hat{c}_p \hat{c}_p^*$ si l'on possède déjà des estimées des vecteurs c_p . C'est une des façons d'intégrer de l'information *a priori*. Si ces estimées sont bonnes, on voit par (11) que chaque matrice cumulée est proche de $c_p c_p^*$.
- $K = P^2$ et les matrices W_k sont creuses, le choix le plus simple étant que chaque m_k ne possède qu'un élément non nul pris égal à 1. Les matrices M_k forment alors une base orthonormée de l'espace des matrices et, conformément au (2.4.), l'identifiabilité est garantie dans ce cas.

- Traitement des $K = P$ matrices 'propres' du tenseur cumulant (cf *infra* et [0]).

Il faut souligner qu'une possible indétermination avec $K < P^2$ reste de probabilité faible. En effet, une seule matrice cumulante, construite sur une matrice de pondération W , possède des valeurs propres données par $k_p c_p^* W c_p$, qui sont donc 'presque sûrement' distinctes si W est choisie 'au hasard', ceci assurant alors l'unicité de la décomposition propre et donc l'identifiabilité.

L'approche la plus sûre consiste à utiliser toute l'information d'ordre 4 ce qui garantit l'identifiabilité. Plutôt que d'estimer $K = P^2$ matrices cumulantes, on peut choisir de travailler directement avec les cumulants. Le rapport entre matrices cumulantes et cumulants au sens strict est donné en appendice. Nous montrons ici comment ce rapport mène à l'idée de matrice propre. Il suffit de remarquer que l'application qui à toute matrice W associe la matrice $Q(W)$ est une application linéaire sur l'espace des matrices. On montre aisément qu'elle est auto-adjointe. Elle possède donc des matrices propres associées à des valeurs propres réelles. Il est alors naturel de choisir comme statistiques cumulantes les matrices propres associées aux valeurs propres de plus grand module. Cette approche, qui implique d'estimer l'ensemble des cumulants de \hat{z} , puis de calculer une décomposition propre du tableau obtenu, est détaillée dans [0, 0]. Le coût d'estimation/calcul n'est pas nécessairement prohibitif puisque l'on travaille dans l'espace signal, c'est-à-dire avec une dimensionnalité égale au nombre de sources.

L'avantage procuré par les matrices propres tient à ce qu'il n'en existe que P qui soient associées à des valeurs propres significatives; ce sont ces matrices que l'on diagonalise. On résume en fait l'information du quatrième ordre dans un ensemble exhaustif de P matrices; cette exhaustivité ne peut être atteinte *a priori* dans l'approche par les matrices cumulantes qu'en choisissant un jeu complet de $K = P^2$ matrices cumulantes.

3.3. IMPLANTATION DE LA DIAGONALISATION CONJOINTE

Pour implanter un algorithme de diagonalisation conjointe approchée, nous proposons d'adapter la technique de Jacobi. Nous en rappelons d'abord brièvement le principe, mais ceci dans une formulation non traditionnelle afin de rendre évidente la généralisation à la diagonalisation de plusieurs matrices. Dans la technique de Jacobi, une transformation unitaire U telle que $U^H M U$ soit diagonale est obtenue par composition de rotations de Givens complexes. Chaque matrice de Givens travaille sur une paire ($i \neq j$) d'indices et son effet sur la somme des carrés des modules des termes diagonaux de M se réduit à modifier $|m_{ii}|^2 + |m_{jj}|^2$. La trace $m_{ii} + m_{jj}$ étant invariante dans une transformation unitaire, la maximisation de $|m_{ii}|^2 + |m_{jj}|^2$ est encore équivalente à celle de $|m_{ii} - m_{jj}|^2$. En choisissant pour les matrices de Givens complexes la paramétrisation suivante :

$$V = \begin{bmatrix} \cos\theta & -\sin\theta e^{-i\phi} \\ \sin\theta e^{i\phi} & \cos\theta \end{bmatrix}, \quad (24)$$

on observe que la quantité $m_{ii} - m_{jj}$ est transformée en $m'_{ii} - m'_{jj}$ selon

$$m'_{ii} - m'_{jj} = g^T u \quad (25)$$

$$u \stackrel{\text{def}}{=} [m_{ii} - m_{jj}, m_{ij} + m_{ji}, i(m_{ij} - m_{ji})] \quad (26)$$

$$g \stackrel{\text{def}}{=} [\cos 2\theta, \sin 2\theta \cos \phi, \sin 2\theta \sin \phi] \quad (27)$$

Si M est hermitienne, le vecteur u est à coefficients réels et les angles de θ et ϕ sont facilement choisis pour annuler $m'_{ii} - m'_{jj}$. L'algorithme de Jacobi dit 'cyclique' pour la diagonalisation d'une seule matrice hermitienne balaye alors en plusieurs fois l'ensemble des paires d'indices ordonnées, en mettant à jour la matrice unitaire U et la matrice à diagonaliser.

Dans le cas qui nous intéresse, la diagonalisation conjointe d'un ensemble quelconque de matrices M_k non nécessairement hermitiennes est définie au sens du critère (14). Nous procédons de la même façon en représentant une transformation unitaire par un produit de rotation de Givens. A chaque pas, et donc pour chaque paire d'indices (i, j), les angles de Givens θ et ϕ doivent alors maximiser

$$\sum_{k=1, K} w_k |g^T u_k|^2 = g^T \left(\sum_{k=1, K} w_k u_k u_k^* \right) g \quad (28)$$

où, pour chaque k , le vecteur u_k de dimension 3 est extrait de la matrice M_k comme dans (26). La somme qui apparaît ci-dessus entre parenthèses est une matrice 3×3 hermitienne. Sa partie réelle est donc anti-symétrique et disparaît dans la forme quadratique. Seule importe donc la matrice F symétrique réelle de taille 3×3 définie par

$$F = \text{Re} \left(\sum_{k=1, K} w_k u_k u_k^* \right) \quad (29)$$

correspondant à une forme quadratique $g^T F g$ qu'il faut maximiser en θ et ϕ . La solution est fort simple, puisque la paramétrisation particulière (27) du vecteur g est équivalente à $g^T g = 1$. Le problème de maximisation quadratique sous contrainte unitaire admet notoirement pour solution un vecteur propre de F associé à la plus grande valeur propre. La matrice F étant réelle de taille 3×3 , ce vecteur propre se calcule explicitement en fonction des éléments de F et les éléments de la rotation de Givens s'en déduisent alors algébriquement [0].

On voit donc que, comme dans la diagonalisation d'une seule matrice et ceci même dans le cas complexe, les paramètres des rotations de Givens sur chaque paire s'obtiennent par une formule algébrique. Une fois ces paramètres calculés, on procède à une remise à jour des matrices cumulantes. Cette étape est coûteuse mais seulement K fois plus que pour une seule matrice; ayant de plus lieu dans l'espace signal, le coût dépend du nombre de sources et non de celui des capteurs. Le coût total de la diagonalisation est ensuite proportionnel au nombre d'itérations de Givens. Ce nombre peut être éventuellement réduit par une

initialisation de la procédure. En effet, la diagonalisation d'une seule matrice cumulée fournit, sauf cas de dégénérescence, des estimées consistantes des c_p . Il est donc possible d'initialiser la diagonalisation conjointe avec la matrice de rotation obtenue par diagonalisation d'une seule des matrices cumulantes.

Dans la pratique, il n'est pas nécessaire de pousser la convergence de l'algorithme jusqu'à la précision du calculateur. En effet, l'analyse de performance de l'identification montre que la variance d'estimation sur les angles est au mieux de $1/4T$ si T échantillons indépendants sont disponibles pour estimer les statistiques. On peut donc arrêter les itérations de Givens dès que sur chaque paire d'indices l'angle ϕ est inférieur à une certaine fraction de $\sqrt{1/4T}$. Autrement dit, on dispose d'un critère d'arrêt de nature statistique, plus rapidement atteint (pour des tailles raisonnables de T et de mot-machine!) qu'un critère numérique.

3.4. ESTIMATION DES COLONNES DE A

La diagonalisation conjointe approchée étant obtenue dans une base orthonormée $\{\hat{c}_p | p = 1, P\}$, on peut choisir d'estimer les signaux s_p (séparation de sources) et/ou les colonnes de A (identification).

Lorsque le but du traitement est de séparer les signaux sources, une estimation bruitée est obtenue directement grâce à l'orthonormalité par

$$\hat{s}_p(t) = \hat{c}_p^* \hat{z}(t). \quad (30)$$

Pour obtenir les estimées des vecteurs directionnels a_p , il faut enfin inverser l'effet du blanchiment. La pseudo inverse $\hat{B}^\#$ du blanchisseur \hat{B} se construit directement comme la matrice dont la p -ième colonne est $(\lambda_p - \sigma^2)^{1/2} v_p$. Les colonnes de A s'en déduisent par $\hat{a}_p = \hat{B}^\# \hat{c}_p$.

4. Performances

Nous donnons ici quelques indications sur les performances que l'on peut attendre des techniques d'identification aveugle.

Tout d'abord, il est important de signaler l'identification aveugle utilisant les cumulants d'ordre 4 ne saurait résister à de forts niveaux de bruit. Précisons cette assertion. Les estimateurs proposés ici sont fortement consistants dès que l'on dispose d'estimées fortement consistantes des cumulants eux-mêmes. Pour des valeurs suffisamment grandes de T (qui désigne le nombre d'échantillons), la variance d'estimation décroît alors en $1/T$ quelque soit le niveau de bruit. La vraie question est de déterminer le coefficient de proportionnalité à $1/T$. Alors que les techniques du second ordre doivent, dans les mauvais niveaux de bruit, affronter une variance en σ^4/T , les techniques du quatrième ordre font face à un σ^8/T car c'est en σ^8 que varie, à bruit fort, la covariance des estimées des cumulants du quatrième ordre. Ainsi,

il faut un nombre d'échantillons croissant comme la quatrième puissance du niveau de bruit pour obtenir à *bruit fort* un niveau donné de performances. Cette contrainte peut facilement devenir rédhibitoire : il est plus raisonnable, comme nous le faisons désormais de considérer des configurations de RSB supérieur à 1.

Nous citons ici les résultats obtenus par A. Souloumiac dans sa thèse [0] consacrée à la séparation de sources. Le critère de qualité retenu pour évaluer les performances de l'identification aveugle est un taux de rejection : après identification aveugle par diagonalisation (pondérée à l'unité) des matrices propres des cumulants empiriques, on forme les estimées des signaux source comme dans (30) et l'on mesure pour chaque signal estimé la puissance relative des autres signaux source. Les simulations ont révélé un fait frappant : pour des signaux dont la modulation se fait à module constant, et dans les situations où les signaux dominent en puissance le bruit, les erreurs d'estimation des paramètres dépendent essentiellement des erreurs commises dans l'estimation de la covariance, bien plus que des erreurs commises dans l'estimation des cumulants d'ordre 4. Cette circonstance est heureuse pour la caractérisation de ces algorithmes puisqu'il suffira –dans le domaine 'bruit faible' qui nous intéresse– de calculer les performances asymptotiques sous les seules erreurs du second-ordre, *i.e.* sous les seules erreurs de blanchiment. On obtient alors un résultat remarquablement simple : après identification aveugle et filtrage inverse, le taux moyen de rejection de chaque source est $1/4T$ pour des signaux à module constant. Ce résultat est pleinement confirmé par les simulations numériques. Il est remarquable que le taux de rejection obtenu ne dépende asymptotiquement que du nombre d'échantillons disponibles (supposés indépendants). Les puissances des signaux et du bruit, les caractéristiques géométriques, les kurtosis des sources, influencent bien entendu les performances asymptotiques, mais de façon indirecte : ces dernières quantités déterminent le seuil au delà duquel le taux de rejection est simplement égal à $1/4T$. On peut par ailleurs montrer que cette valeur de $1/4T$ est une borne aux performances des algorithmes basés sur un préblanchiment des observations [0].

Donnons quelques chiffres tirés de [0] pour fixer les idées. On considère une antenne linéaire de 5 capteurs espacés d'une demie longueur d'onde, recevant les signaux de 2 sources i.i.d. modulées en PSK, de puissances identiques et égales à σ_s^2 , situées aux angles θ_1 et θ_2 par rapport au plan normal à l'antenne. Dans une identification aveugle avec 100 échantillons, on obtient typiquement un taux de rejection en $1/4T$ voisin de 26 dB et donc une bonne séparation des sources. Les simulations numériques confirment cette valeur à quelques décibels près, mais elle n'est atteinte, lorsque $\theta_1 = 0^\circ$ et $\theta_2 = 8^\circ$, que si $\sigma_s^2 > 10\sigma^2$. Bien entendu, ce seuil lui-même croît lorsque les deux sources se rapprochent. On renvoie à [0] et [0] pour une étude plus complète.

5. Appendice

5.1. MATRICES CUMULANTES ET TENSEUR CUMULANT

On utilise ici la notation tensorielle indicée où un vecteur x est représenté par ses coordonnées x_i dans une base orthonormée quelconque. Les coordonnées de x^* , son transconjugué, sont notées x^i . La 'quadrivariance' Q_x est le tenseur des cumulants du quatrième ordre de x , c'est-à-dire le tenseur 2 fois covariant et 2 fois contravariant de coordonnées q_{ik}^{jl} définies par

$$q_{ik}^{jl} \stackrel{\text{def}}{=} \text{Cum}(x_i, x^j, x_k, x^l). \quad (31)$$

Cette définition fournit une extension évidente de celle utilisée au second ordre, puisqu'on peut voir la covariance Γ_x comme le tenseur de coordonnées γ_i^j définies par

$$\gamma_i^j = \text{Cum}(x_i, x^j). \quad (32)$$

Pour des variables aléatoires circulaires (en fait dès que $E x_i = 0$ et $E x_i x_j = 0$ pour tout i et j), la définition des cumulants d'ordre 4 se réduit à

$$\begin{aligned} \text{Cum}(x_i, x^j, x_k, x^l) &= E\{x_i x^j x_k x^l\} - E\{x_i x^j\} E\{x_k x^l\} \\ &\quad - E\{x_i x^l\} E\{x_k x^j\}. \end{aligned} \quad (33)$$

Le modèle (1) se réécrit en notations indicées comme

$$x_i = \sum_p s_p p_i + b_i \quad (34)$$

où nous prenons la liberté de noter p_i la i -ième coordonnée du vecteur a_p . Sous nos hypothèses, ce modèle confère à la quadrivariance la structure

$$q_{ik}^{jl} = \sum_p k_p p_i p^j p_k p^l \quad (35)$$

qui s'obtient en conséquence directe des trois propriétés fondamentales des cumulants, à savoir i) additivité dans l'addition de variables indépendantes ii) linéarité dans chaque argument iii) nullité pour des variables gaussiennes aux ordres supérieurs à deux.

Cette structure particulière donne directement la structure des matrices cumulantes, car $M = Q_x(W)$ n'exprime pas autre

chose, en termes tensoriels, que la contraction de W sur Q_x . Considérons en effet les coordonnées w_i^j et m_i^j de matrices reliées par les relations

$$m_i^j = \sum_{kl} q_{ik}^{jl} w_l^k. \quad (36)$$

En reportant cette expression dans (33), on retrouve précisément la définition (5) des matrices cumulantes. Sur l'autre main, en reportant (36) dans la structure (35) de la quadrivariance, on obtient la structure (6) des matrices cumulantes.

BIBLIOGRAPHIE

- [1] M. GAETA and J.-L. LACOUME, "Estimateurs du maximum de vraisemblance tendus la séparation de sources non gaussiennes," *Traitement du Signal*, vol. 7, no. 5, pp. 419-434, 1990.
- [2] P. COMON, "Independent component analysis," in *Proc. Int. Workshop on Higher-Order Stat., Chamrousse, France*, pp. 111-120, 1991.
- [3] J.-F. CARDOSO, "Iterative techniques for blind source separation using only fourth order cumulants," in *Proc. EUSIPCO*, pp. 739-742, 1992.
- [4] C. JUTTEN and J. HÉRAULT, "Blind separation of sources : an adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1-10, 1991.
- [5] D. T. PHAM, P. GARRAT, and C. JUTTEN, "Separation of a mixture of independent sources through a maximum likelihood approach," in *Proc. EUSIPCO*, pp. 771-774, 1992.
- [6] J.-F. CARDOSO and A. SOULOUMIAC, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, pp. 362-370, Dec. 1994.
- [7] J.-F. CARDOSO, "Source separation using higher order moments," in *Proc. ICASSP*, pp. 2109-2112, 1989.
- [8] V. C. SOON, L. TONG, Y. F. HUANG, and R. LIU, "An extended fourth order blind identification algorithm in spatially correlated noise," in *Proc. ICASSP*, pp. 1365-1368, 1990.
- [9] A. SOULOUMIAC, "Utilisation des statistiques d'ordre supérieur en traitement d'antenne," *Thèse de doctorat. Télécom Paris*, Février 1993.
- [10] A. SOULOUMIAC and J.-F. CARDOSO, "Givens angles for simultaneous diagonalization," soumis à *SIAM J. Mat. Anal. Appl.*.
- [11] A. SOULOUMIAC and J.-F. CARDOSO, "Performances en séparation de sources," in *Proc. GRETSI, Juan les Pins, France*, pp. 321-324, 1993.

Remerciements à Jean-François Bercher pour avoir suggéré plusieurs améliorations du manuscrit.