BLIND BEAMFORMING FOR NON GAUSSIAN SIGNALS.

Jean-François Cardoso, Antoine Souloumiac
Télécom Paris / CNRS URA 820 / GdR TdSI
46 rue Barrault, 75634 PARIS
email: cardoso@sig.enst.fr

Reprint from: 'IEE- Proceedings-F, vol. 140, no 6, pp. 362-370, dec. 1993.'

Note: This version (posted Jan. 97), corrects a typo in the published manuscript. In the description of step 1 and step 4 of the algorithm in section 4.1, one now correctly reads $\mathbf{h}_n$ rather that $\mathbf{h}_p$ as was first erroneously written.

Abstract

This paper considers an application of blind identification to beamforming. The key point is to use *estimates* of directional vectors rather than resorting to their hypothesized value. By using estimates of the directional vectors obtained via blind identification i.e. *without knowing the arrray manifold*, beamforming is made robust with respect to array deformations, distortion of the wave front, pointing errors, etc ... so that neither array calibration nor physical modeling are necessary. Rather surprisingly, 'blind beamformers' may outperform 'informed beamformers' in a plausible range of parameters, *even when the array is perfectly known to the informed beamformer.*

The key assumption blind identification relies on is the statistical independence of the sources, which we exploit using fourth-order cumulants. A computationally efficient technique is presented for the blind estimation of directional vectors, based on joint diagonalization of 4th-order cumulant matrices; its implementation is described and performance is investigated by numerical experiments.

1 Introduction

This paper is devoted to an application of blind identification to beamforming in the context of narrow-band array processing. Let us first recall the standard linear data model, where an array of m sensors receives waves emitted by n narrow-band sources. If $s_p(t)$ denotes the signal emitted by the p-th source, its contribution to the array output can be written $s_p(t)\mathbf{a}_p$ where $\mathbf{a}_p$ is a fixed m-vector acting, at the carrier frequency, as the spatial transfer function between the p-th emitting source and the array. Denoting $n(t)$ a possible additive noise, the array output $\mathbf{x}(t)$ is given by

$$\mathbf{y}(t) = \sum_{p=1,n} s_p(t)\mathbf{a}_p = A\mathbf{s}(t) \quad (1)$$

$$\mathbf{x}(t) = \mathbf{y}(t) + \mathbf{n}(t) \quad (2)$$

In (1), we also use the customary matrix-vector notation where the $m \times n$ *array matrix* A has vector $\mathbf{a}_p$ as its p-th column and where the $n \times 1$ vector $\mathbf{s}(t)$ has the signal $s_p(t)$ in p-th position. Each vector $\mathbf{a}_p$ is called the *directional vector* associated to the the p-th source since it depends on the direction (or on any relevant location parameters) of the source.

In the narrow-band context, the signal emitted by a spatially coherent source may be estimated by forming the inner product between the array output and a $m \times 1$ vector acting as a spatial filter. The review paper [1] is a good introduction to various strategies for designing spatial filters or 'beamformers'. Denote $\mathbf{f}_p$ the spatial filter designed to extract $s_p(t)$, the signal of interest. The simplest approach to coherently combine sensor outputs is to take $\mathbf{f}_p = \mathbf{a}_p$. However, beamforming may take into account the other signals and the noise in order to design optimal filters. For instance, if $s_p(t)$ is independent from the other contributions, the highest SNR at the filter output is obtained by forming a MVDR (minimum variance distortionless response) filter which is proportional to

$$\mathbf{f}_p = R_x^{-1}\mathbf{a}_p \quad \text{with} \quad R_x \stackrel{\text{def}}{=} E\{\mathbf{x}(t)\mathbf{x}(t)^*\}. \quad (3)$$

If the directional vectors associated to the other coherent sources are known, then one may also constrain the spatial filter to cancel these interfering signals, leading to a LCMV (linear constrained minimum variance) filter. In spatially white noise and for mutually independent source signals, the LCMV filter is easily found to be proportional to

$$\mathbf{f}_p = R_y^{\#}\mathbf{a}_p \quad \text{with} \quad R_y \stackrel{\text{def}}{=} E\{\mathbf{y}(t)\mathbf{y}(t)^*\} \quad (4)$$

where # denotes the pseudoinverse. Note that this filter can be computed without knowing the directional vectors but the one corresponding to the signal of interest.

These two standard approaches — MVDR and LCMV — are based on the knowledge of the directional vector associated to the desired signal and may be quite sensitive to errors in this vector. Such errors may be due to unknown deformation of the array, drift in the electronic hardware (calibration errors) or to multiple paths and/or wave reflexions in the vicinity of the array (errors in modeling the propagation). Even if the array is perfectly known, pointing errors cause performance degradation. Finally, performance is limited by the use in (3) or (4) of sample statistics in place of the true covariances R_x or R_y which cannot be perfectly estimated with finite sample size.

We propose to consider the use of *estimated* directional vectors in beamforming. It is the purpose of this contribution to describe a blind identification technique allowing the directional vectors to be estimated *without knowing the array manifold*, i.e. without physical modeling of the propagation or array calibration. At first sight, such an aproach may appear paradoxical since the array manifold is the link between the location of a source of interest and its associated directional vector. Relevance of the blind identification to beamforming must then be given some discussion.

If the field contains only one coherent source, whose associated directional vector can be reliably estimated in the blind fashion, then implementation of the spatial filters (3) or (4) is straightforward. There is a clear benefit in processing without the knowledge of the array manifold, since

'blind' beamforming is, by essence, insensitive to errors in the manifold model. In presence of *several* coherent sources though, the blind approach cannot stand by itself since it typically yields the directional vectors associated to *all* sources. Unless one is interested in all the received signals without discrimination, some additional processing is then necessary in order to select among the estimated directional vectors those associated to the sources of interest. One approach is to form beams corresponding to all the detected signals and select the signals of interest using 'non-spatial' information (spectral content, modulation, ...). Another approach is to select among the blindly estimated directional vectors the closest to the directional vector predicted by physical modeling which is a good guess if the array manifold is not too severely distorted and if the source position is known in advance.

Blind identification techniques rely on the assumption of mutual independence of the source signals received at a given time. The question of mutual independence deserves a specific discussion in the context of blind array processing. First note that the assumption of independence between sources is a statistically strong hypothesis but very plausible in practice for physically separated emitters. Wave reflexions (or multiple paths) though, cause a single emitter to contribute by several correlated wave fronts to the field impinging on the array. When these multiple paths correspond to similar propagation delays, these waves are fully coherent. We stress that this circumstance does *not* affect blind array processing, while it severely does so for parametric array processing. This is because two fully coherent sources (say sources 1 and 2) correspond to proportional signals: $s_1(t) = \alpha s_2(t)$ with α some complex number. The combination of these signals at the array ouput is $s_1(t)\mathbf{a}_1 + s_2(t)\mathbf{a}_2 = s_1(t)(\mathbf{a}_1 + \alpha\mathbf{a}_2)$ which is seen as a single source with a composite 'directional' vector $\mathbf{a}_1 + \alpha\mathbf{a}_2$. This would be a problem for any method assuming that each independent component is associated to some direction. The blind approach does not make any such assumption since it does not deal with directions of arrival. In other words, through multiple propagation paths, an 'informed array' sees several correlated sources, each with a directional vector corresponding to its location, while a 'blind array' sees only one source (in the full coherence case) with a unique composite 'directional' vector. Of course, the composite nature of this vector is irrelevant in the blind approach, where it makes no difference. Hence, the term 'directional vector' is misleading in the blind context and is not used in the following (except in the experimental section, where a 'fair' array is used).

We close these remarks by mentionning, anticipating a bit, that the columns of A are blindly estimated up to a scalar factor. This is irrelevant in many applications; in particular such a factor in $\mathbf{a}_p$ does not change the SNR at the output of spatial filters like (3) or (4).

This paper is organized as follows. In section 2, the problem of blind estimation of the array matrix A is stated together with the relevant statistical hypothesis. It is recalled that the array output covariance determines A up to a unitary factor whose identification using 4th-order cumulants is discussed in section 3. To this purpose, the notion of 'joint diagonalization' is introduced. With this device, an efficient technique for blind identification is described in section 4, where related approaches are also reviewed (see also subsections 3.1 and 3.2). The last section 5 investigates performance in beamforming applications via numerical experiments.

2 Blind estimation of the array matrix

2.1 Blind identifiability

Before proceeding, it is important to specify the notion of blind identification. Since each source contributes to the array output via the product $s_p(t)\mathbf{a}_p$, the observed process is unaffected by the exchange of a complex scalar factor between each source signal $s_p(t)$ and each vector $\mathbf{a}_p$. Also note that the numbering of the sources is a pure notational convention but is otherwise immaterial. These simple remarks show that, without additional *a priori* information, matrix A can be at best identified up to permutation and scaling of its columns. More general considerations on blind identifiability and indetermination can be found in [2].

Advantage can be taken of this indetermination to assume, *without any loss of generality*, that the source signals have unit variance: $E\{|s_p(t)|^2\} = 1$ for $1 \leq p \leq n$ so that the dynamic range of the sources is accounted for by the magnitude of the corresponding column of A. For independent sources, we then have

$$R_s \stackrel{\text{def}}{=} E\{\mathbf{s}(t)\mathbf{s}(t)^*\} = I_n \quad \text{so that} \quad R_y = AA^H. \tag{5}$$

This normalization still leaves undetermined the ordering and the phases of the columns of A. The following definition is then in order: *Two matrices M and N are said to be* essentially equal *if it exists a matrix P such that $M = NP$ where P has exactly one non-zero entry in each row and column, these entries having unit modulus.* In this paper, blind identification of A is understood as the determination of a matrix essentially equal to A, without A being parameterized.

2.2 Notations and assumptions

Our approach to blind identification exploits the fourth-order cumulants of the array output. For $\mathbf{v}$ a complex d-dimensional random vector with coordinates $v_1, \cdots, v_d$ and finite 4th-order cumulants, we define a cumulant set denoted $\mathcal{Q}_v$ as:

$$\mathcal{Q}_v \stackrel{\text{def}}{=} \{\ Cum(v_i, v_j^*, v_k, v_l^*) \mid 1 \leq i, j, k, l \leq d\}. \tag{6}$$

For a complex stationary process $\mathbf{v}(t)$, we also denote $\mathcal{Q}_v$ rather than $\mathcal{Q}_{\mathbf{v}(t)}$ since the latter does not depend on t. We assume

H0 : *The processes $\mathbf{n}(t)$, $s_1(t), \cdots, s_n(t)$ are jointly stationary.*

The *kurtosis* of the p-th source is the real number

$$k_p \stackrel{\text{def}}{=} Cum(s_p(t), s_p^*(t), s_p(t), s_p^*(t)). \tag{7}$$

A source is said to be *kurtic* if it has a non zero kurtosis. We restrict ourselves to the case where:

H1 : *There is at most one non kurtic source.*

The crucial assumptions blind identification relies on are related to independence, exploited in this paper by assuming non Gaussian signals. More specifically, we assume:

H2 : *The vectors $\mathbf{a}_1, \cdots, \mathbf{a}_n$ are linearly independent but otherwise arbitrary.*

H3 : *The variables $s_1(t), \cdots, s_n(t)$ are statistically independent for each t.*

We will see that under H1-3, the array matrix A is essentially determined from R_y and $\mathcal{Q}_y$. For these quantities to be consistently estimated, it is further assumed that

H4 : *There exist consistent estimates for R_x and $\mathcal{Q}_x$.*

H5 : *The additive noise is normally distributed and independent from the sources.*

H6 : *The additive noise is spatially white $R_n = \sigma I_m$ with unknown variance σ and $n < m$.*

By H5, an estimate of $\mathcal{Q}_x$ also is an estimate of $\mathcal{Q}_y$ since cumulants are additive for independent variables and since higher-order cumulants are zero for normally distributed variables. By H6, an estimate of R_y can be classically constructed from the eigendecomposition of an estimate of R_x. We insist that assumptions H4-6 could be replaced by any other assumption set serving the same purpose: the existence of consistent estimates for R_y and $\mathcal{Q}_y$.

2.3 Using second-order information

We consider exploiting second order information by whitening the signal part $\mathbf{y}(t)$ of the observation. This is done via a *whitening matrix* W, i.e. a $n \times m$ matrix such that $W\mathbf{y}(t)$ is spatially white. The whiteness condition is

$$I_n = WR_yW^H = WAA^HW^H \tag{8}$$

where the last equality stems from (5) and I_n denotes the $n \times n$ identity matrix. Equation (8) implies that WA is a unitary matrix: for any whitening matrix W, it then exists a unitary matrix U such that $WA = U$. As a consequence, matrix A can be factored as

$$A = W^{\#}U = W^{\#}[\mathbf{u}_1, \cdots, \mathbf{u}_n] \tag{9}$$

where U is unitary. The use of second-order information – in the form of an estimate of R_y which is used to solve for W in (8) – reduces the determination of the $m \times n$ mixing matrix A to the determination of a unitary $n \times n$ matrix U. The whitened process $\mathbf{z}(t) = W\mathbf{x}(t)$ still obeys a linear model:

$$\mathbf{z}(t) \stackrel{\text{def}}{=} W\mathbf{x}(t) = W(A\mathbf{s}(t) + \mathbf{n}(t)) = U\mathbf{s}(t) + W\mathbf{n}(t) \tag{10}$$

The signal part of the whitened process now is a unitary mixture of the source signals. Note that all the information contained in the covariance is 'exhausted' after the whitening, in the sense that changing U in (10) to any other unitary matrix leaves unchanged the covariance of $\mathbf{z}(t)$.

3 Determining the unitary factor.

Two approaches for the determination of the unitary factor U in $A = W^{\#}U$ have been reported. In the first approach, U is computed as the diagonalizer of a $n \times n$ cumulant matrix. These 'eigen-based' techniques are computationally simple but, being based only on n^2 cumulant statistics, they may show poor statistical performance. Another approach obtains an estimate of U as the optimizer of some identification criterion which is a function of the *whole* cumulant set $\mathcal{Q}_z$: better performance is expected at the expense of solving an optimization problem. These approaches are reviewed in the next two subsections; we then describe our technique which combines advantages of both the eigen-based and the criterion-based approaches.

3.1 Approaches based on eigendecomposition

We consider *cumulant matrices* defined as follows. To any $n \times n$ matrix M, is associated a 'cumulant matrix' denoted $Q_z(M)$, defined entrywise by

$$N = Q_z(M) \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad n_{ij} = \sum_{k,l=1,n} Cum(z_i, z_j^*, z_k, z_l^*) m_{lk} \quad 1 \leq i, j \leq n. \tag{11}$$

The (k,l)-th parallel cumulant slice is defined as the matrix whose (i,j)-th entry is $Cum(z_i, z_j^*, z_k, z_l^*)$. It is seen to be equal to $Q_z(M)$ by taking $M = \mathbf{b}_l \mathbf{b}_k^*$ where $\mathbf{b}_k$ denotes the $n \times 1$ vector with 1 in k-th position and 0 elsewhere. Note that a cumulant matrix $Q_z(M)$ may be seen as a linear combination of 'parallel cumulant slices' with the entries of M as coefficients. For later use, we define the 'parallel set' $\mathcal{N}^p$ as the set of all the parallel slices:

$$\mathcal{N}^p \stackrel{\text{def}}{=} \{Q_z(\mathbf{b}_l \mathbf{b}_k^*) | 1 \leq k, l \leq n\} \tag{12}$$

Since $\mathbf{z}(t)$ obeys the linear model (10), the cumulant matrices take a simple form. Using the cumulant properties — Gaussian rejection, additivity, multilinearity — it is straightforward to establish that

$$Q_z(M) = \sum_{p=1,n} k_p \ \mathbf{u}_p^* M \mathbf{u}_p \ \mathbf{u}_p \mathbf{u}_p^* \quad \forall M. \tag{13}$$

or equivalently

$$Q_z(M) = U \Lambda_M U^H \qquad \Lambda_M \stackrel{\text{def}}{=} Diag(k_1 \ \mathbf{u}_1^* M \mathbf{u}_1, \cdots, k_n \ \mathbf{u}_n^* M \mathbf{u}_n). \tag{14}$$

From equation (14) stems the basic idea for eigen-based blind identification : any cumulant matrix is diagonalized by U. Hence, the eigenvectors of a cumulant matrix, left multiplied by $W^{\#}$ as

in (9) give the columns of A. It is worth noticing that the fundamental indetermination of blind identification precisely corresponds to the indetermination of the eigendecomposition (provided the spectrum is not degenerate).

The simplest implementation of this idea is for circularly distributed signals where

$$Q_z(M) = E\{\mathbf{z}^* M \mathbf{z}\ \mathbf{z}\mathbf{z}^*\} - R_z M R_z - R_z Trace(M R_z) \quad \forall M \tag{15}$$

as can be seen by inserting in (11) the expression of cumulants in term of moments. Then, in the noiseless case $R_z = U R_s U^H = U U^H = I_n$ so that (15) gives $Q_z(I_n) = E\{|\mathbf{z}|^2 \mathbf{z}\mathbf{z}^*\} - (n+1) I_n$. It follows that $Q_z(I_n)$ and the 'weighted covariance' $E\{|\mathbf{z}|^2 \mathbf{z}\mathbf{z}^*\}$ have the same eigenvectors. Hence U may be identified as the unitary diagonalizer of the latter, i.e. without even computing the full cumulant matrix $Q_z(I_n)$ [3]. If some noise is present though, expression (15) must be evaluated, i.e. the corrective term $R_z^2 + R_z Tr(R_z)$ must be subtracted from the weighted covariance as shown in [4] in the real case.

Unitary diagonalization of $Q_z(I_n)$ is not essentially determined though, if some eigenvalues are identical. These eigenvalues, by (14), are $k_p \mathbf{u}_p^* I_n \mathbf{u}_p = k_p |\mathbf{u}_p|^2 = k_p$ so that the case of degeneracy is when some sources have identical kurtosis. It is suggested in [5] to diagonalize a linear combination of cumulant slices that is to diagonalize $Q_z(M)$ for some matrix M. The p-th eigenvalue of $Q_z(M)$ being $k_p \mathbf{u}_p^* M \mathbf{u}_p$, degeneracy is very unlikely. This approach suffers two drawbacks. First, there is no guideline as how to choose *a priori* the linear combination of cumulant slices or equivalently as how to choose M before evaluating $Q_z(M)$. Second, such a technique uses only a fraction of the fourth-order information: if one computes several randomly chosen cumulant matrices and retains the one with the largest spectrum spread, the information contained in the other cumulant matrices is lost.

3.2 Approaches based on optimization of cumulant criteria

Let V denote a $n \times n$ unitary matrix and further define $\mathbf{e}(t)$, as in figure 1,

$$\mathbf{e}(t) \stackrel{\text{def}}{=} V^H \mathbf{z}(t) = V^H U \mathbf{s}(t) + V^H W \mathbf{n}(t). \tag{16}$$

If $V = U$ then $V^H U = I_n$ and the coordinates of $\mathbf{e}(t)$ are the (noise corrupted) source signals.

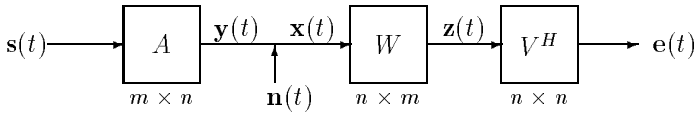

Figure 1: Inverting A by chaining a whitener and a unitary matrix.

More generally, if V is essentially equal to U, the coordinates of $\mathbf{e}(t)$ are the phase-shifted source signals, possibly permuted and corrupted by additive Gaussian noise, so that their higher-order *cross*-cumulants are zero.

It has then been proposed to determine U as the unitary minimize of the sum of all the squared cross-cumulants in $\mathcal{Q}_e$. Since the sum of the squared cross-cumulants plus the sum of the squared auto-cumulants of $\mathcal{Q}_e$ does not depend on V as long as V is kept unitary, this is equivalent to *maximizing* under unitary constraint the criterion [6]:

$$c'(V) \stackrel{\text{def}}{=} \sum_{i=1,n} |Cum(e_i, e_i^*, e_i, e_i^*)|^2. \tag{17}$$

This criterion first appeared in [7] where it is obtained via a 4th-order Gram-Charlier expansion of the likelihood function. Very interestingly, Comon [6] arrives at the same criterion by a different

approach based on contrast functions, which is reminiscent of [8]. Comon also describes an algorithm for maximizing (17) via products of Givens rotations. Unfortunately, the Givens angles at each step cannot be obtained in closed form in the complex case.

We propose to determine U as the unitary maximizer of the criterion $c(V)$:

$$c(V) \stackrel{\text{def}}{=} \sum_{i,k,l=1,n} |Cum(e_i, e_i^*, e_k, e_l^*)|^2 \tag{18}$$

which is equivalent to minimizing the sum of the squared cross-cumulants with distinct first and second indices. The main reason for considering criterion (18) is its link to underlying eigenstructures which allows for an efficient optimization of it by the mean of 'joint diagonalization'.

3.3 Joint diagonalization

Let $\mathcal{N} = \{N_r | 1 \leq r \leq s\}$ be a set of s matrices with common size $n \times n$. A *joint diagonalizer* of the set $\mathcal{N}$ is defined as a unitary maximizer of the criterion

$$C(V, \mathcal{N}) \stackrel{\text{def}}{=} \sum_{r=1,s} |diag(V^H N_r V)|^2 \tag{19}$$

where $|diag(.)|$ is the norm of the vector build from the diagonal of the matrix argument. When the set $\mathcal{N}$ contains only one hermitian matrix, joint diagonalization is equivalent to usual unitary diagonalization. If the set $\mathcal{N}$ cannot be exactly jointly diagonalized (this is the case when *sample* cumulants are processed), the unitary maximization of (19) defines a somewhat arbitrary but quite natural 'joint approximate diagonalization'.

The link between optimization-based and eigen-based blind identification techniques is established by considering the joint diagonalization of several cumulant matrices. In particular:

Proposition 1 *For any unitary matrix V, $c(V) = C(V, \mathcal{N}^p)$,*

which means that the unitary maximization of $c(V)$ is equivalent to the joint diagonalization of the parallel set. Blind identifiability via joint diagonalization is guaranteed by

Proposition 2 *Under H1-3, a joint diagonalizer of $\mathcal{N}^p$ is essentially equal to U.*

Proofs of these propositions are skecthed in appendix B.

Joint diagonalization corresponding to usual diagonalization when only one hermitian matrix is involved, it is no surprise that the Jacobi technique can be extended to the joint diagonalization of several matrices. This extension is described in appendix A and offers a computation cost which is roughly s times the cost of diagonalizing a single matrix. In addition, for the particular problem at hand, this cost can be further reduced by *initializing* the joint diagonalizer with the unitary matrix returned by the (ordinary) diagonalization of a single cumulant matrix. A convenient choice is to diagonalize some $Q_z(M)$ where M is a random hermitian matrix, because then matrix $Q_z(M)$ also is hermitian.

3.4 Representation of the 4th-order cumulants by eigenmatrices

The computational efficiency of joint diagonalization can be further increased by downsizing $\mathcal{N}^p$ to a smaller set made of the significant 'eigenmatrices' of Q_z.

Proposition 3 *For any d-dimensional complex random vector* $\mathbf{v}$ *with 4th-order cumulants, there exist d^2 real numbers $\lambda_1, \cdots, \lambda_{d^2}$ and d^2 matrices $M_1, \cdots, M_{d^2}$, called* eigenmatrices *verifying*

$$Q_v(M_r) = \lambda_r M_r \qquad Tr(M_r M_s^H) = \delta(r, s) \qquad 1 \leq r, s \leq d^2. \tag{20}$$

The proof is straightforward by a classic 'stacking-unstacking' device : the relation $N = Q_v(M)$ is put in vector-matrix form $\tilde{N} = \tilde{Q}\tilde{M}$ by mapping the $d \times d$ matrices N and M into $d^2 \times 1$ vectors $\tilde{N}$ and $\tilde{M}$ and by mapping the set $\mathcal{Q}_v$ into a $d^2 \times d^2$ matrix. The simplest mapping is defined entrywise, for $1 \leq a, b \leq d^2$, by

$$\tilde{N}_a = n_{ij}, \quad \tilde{M}_a = m_{ij}, \quad \tilde{Q}_{ab} = Cum(v_i, v_j^*, v_l, v_k^*), \quad \text{with} \quad a = i+(j-1)d, \quad b = k+(l-1)d \tag{21}$$

Matrix $\tilde{Q}$ is easily checked to be hermitian. It then admits a set of d^2 real eigenvalues $\lambda_1, \cdots, \lambda_{d^2}$ and d^2 corresponding eigenvectors whose unstacking as in (21) yields the eigenmatrices. Eigenmatrices inherit the orthonormality property from the eigenvectors. The same results can be arrived at using a Kroneker product formulation as in [9].

The eigen-structure of $\mathcal{Q}_z$ derives from (13). It is readily checked that the set $\{\mathbf{u}_p\mathbf{u}_q^* | 1 \leq p, q \leq n\}$ verifies the properties of proposition 3. Orthonormality of the matrices in this set stems from U being unitary and by substitution into (13), one finds that $Q_z(\mathbf{u}_p\mathbf{u}_p^*) = k_p\mathbf{u}_p\mathbf{u}_p^*$ while $Q_z(\mathbf{u}_p\mathbf{u}_q^*) = 0$ for $p \neq q$. Hence the spectrum of Q_z is made of $n(n-1)$ zero eigenvalues and n eigenvalues equal to the kurtosis of the sources, (a similar device has been proposed in [10] for detecting the number of kurtic sources). With the notations of proposition 3 and after ordering the eigenvalues by decreasing order of magnitude, we define the *eigen-set* of $\mathcal{Q}_z$ as the matrix set :

$$\mathcal{N}^e \stackrel{\text{def}}{=} \{\lambda_r M_r | 1 \leq r \leq n\} \tag{22}$$

For our purpose, the eigen-set contains the relevant 4th-order information, since we have:

Proposition 4 *Under H0-3, for any unitary matrix V, $c(V) = C(V, \mathcal{N}^e)$.*

This reduced set of n matrices (rather than n^2 in $\mathcal{N}^p$) together with the extended Jacobi technique makes the maximization of $c(V)$ computationally attractive.

4 Blind identification algorithms

4.1 The JADE algorithm

A blind identification algorithm by Joint Approximate Diagonalization of Eigen-matrices (JADE)[1] can now be described by the following steps.

Step 1. Form the sample covariance $\hat{R}_x$ and compute a whitening matrix $\hat{W}$.

Step 2. Form the sample 4th-order cumulants $\hat{\mathcal{Q}}_z$ of the whitened process $\hat{\mathbf{z}}(t) = \hat{W}\mathbf{x}(t)$; compute the n most significant eigenpairs $\{\hat{\lambda}_r, \hat{M}_r | 1 \leq r \leq n\}$.

Step 3. Jointly diagonalize the set $\hat{\mathcal{N}}^e = \{\hat{\lambda}_r\hat{M}_r | 1 \leq r \leq n\}$ by a unitary matrix $\hat{U}$.

Step 4. An estimate of A is $\hat{A} = \hat{W}^{\#}\hat{U}$.

Some comments are in order about these successive steps.

Step 1 is concerned with 2nd-order statistics and is standard under H5-6 ; it is implemented via eigendecomposition of $\hat{R}_x$. Thanks to the white noise assumption, an estimate $\hat{\sigma}$ of the noise variance is the average of the $m-n$ smallest eigenvalues of $\hat{R}_x$. Denote $\mu_1, \cdots, \mu_n$ the n largest eigenvalues and $\mathbf{h}_1, \cdots, \mathbf{h}_n$ the corresponding eigenvectors of $\hat{R}_x$. A whitener is $\hat{W} = [(\mu_1 - \hat{\sigma})^{-1/2}\mathbf{h}_1, \cdots, (\mu_n - \hat{\sigma})^{-1/2}\mathbf{h}_n]^H$. We do not address the important issue of detecting the number n of sources.

In step 2, computation of the eigenmatrices amounts to diagonalizing a $n^2 \times n^2$ matrix made from the elements of $\hat{\mathcal{Q}}_z$. A standard algorithm for eigendecomposition of hermitian matrices will perfectly do, but more efficient implementations can also be devised, by taking into account additional cumulant symmetries or the fact that only the n most significant eigenpairs are needed [11]. Recall that computation of the eigenmatrices may be bypassed if, for simplicity, joint diagonalization is performed on the parallel set $\mathcal{N}^p$. An even simpler implementation is to form a set

[1] A Matlab implementation of JADE is available upon request or by anonymous FTP at `sig.enst.fr`.

$\mathcal{N} = \{Q_z(C_r) | 1 \leq r \leq s\}$ (possibly using the sample counterpart of (15)), where the C_r's are s arbitrary matrices in arbitrary number. Of course, identifiability cannot be guaranteed *a priori* and performance may be significantly lower than when $\mathcal{N}^e$ or $\mathcal{N}^p$ are used.

Step 3 is implemented by extending the single-matrix Jacobi technique to several matrices as described in appendix A. Note that when $n = 2$, the Jacobi technique is not iterative: a unique Givens rotation achieves (joint) diagonalization. Also recall that joint diagonalization may be initialized with the (usual) diagonalizer of a single cumulant matrix.

In step 4, the pseudo-inverse of $\hat{W}$ needs not be explicitly computed: the eigendecomposition of $\hat{R}_x$ may be recycled by $\hat{W}^{\#} = [(\mu_1 - \hat{\sigma})^{1/2}\mathbf{h}_1, \cdots, (\mu_n - \hat{\sigma})^{1/2}\mathbf{h}_n]$.

4.2 Related approaches

Besides the papers [3, 4, 5, 6, 7, 8] already mentioned in section 3, other contributions are related to blind identification of the model (1,2).

First note that this 'instantaneous spatial mixture' may be seen as a special case of more general spatio-temporal mixtures; in particular, blind identification techniques designed in the framework of multichannel ARMA modeling could be applied, provided they are extended to the complex case. See for instance the cumulant-based approach in [12, 13] and [14] for an adaptive approach. At the other extreme, stand purely temporal mixtures and the blind deconvolution problem, showing a structure similar to the purely spatial problem. For instance, the blind deconvolution techniques in [15] closely parallels the unitary maximization of (17) or the 'reverse criterion' of [10]. Similarly, the CMA algorithm [16] may be implemented in a spatial version [17].

Blind identification may be based on higher-order cumulants only (hence without second-order prewhitening), with the benefit that consistent estimation is possible without modeling the spatial structure of the noise as long as it is independent and normally distributed. The references [10, 13, 18, 19, 20] specifically considers the spatial problem.

Blind identification of model (1,2) is closely related to the 'source separation' problem since the latter consists in finding a 'separating matrix' B such that the coordinates of $B\mathbf{x}(t)$ are the source signals (up to the usual indeterminations), possibly corrupted by noise. Adaptive solutions may be based on cumulant criteria as in [21, 22, 14]. More generally, statistical independence at the output of a separating matrix (in the noiseless case) may be exploited by adapting B using non-linear functions of its output. A seminal paper is [23], which deals with real signals; see also [24, 25]. For i.i.d. source signals with known, differentiable probability densities, the maximum likelihood approach of [26] provides asymptotically optimal estimates in the noiseless case.

Finally, simple solutions can also be implemented if the model (1,2) holds with *temporally* correlated source signals, in which case non normality of sources is no longer necessary. The approach of section 3.1 may be followed, by diagonalizing a correlation matrix $E\{\mathbf{z}(t+\tau)\mathbf{z}(t)^*\}$ rather than a cumulant matrix. This was independently proposed in [24] and in [27]. As with cumulant matrices, indetermination problems may occur and several correlation matrices (i.e. for various τ) may be jointly diagonalized for the sake of robustness as shown in [28]. A necessary identifiability condition is that the source signals have different spectra. A safe approach may consist in the joint diagonalization of a set made of cumulant matrices *and* of correlation matrices.

5 Application to beamforming

5.1 Performance index

Applicability of these results to beamforming is now investigated. Denote $\hat{\mathbf{f}}_p$ an estimate of a spatial filter computed from T data samples when p is the signal of interest. The estimated signal is

$$\hat{s}_p(t) \stackrel{\text{def}}{=} \hat{\mathbf{f}}_p^* \mathbf{x}(t) = \sum_{q=1,n} s_q(t)\ \hat{\mathbf{f}}_p^* \mathbf{a}_q + \hat{\mathbf{f}}_p^* \mathbf{n}(t) \tag{23}$$

which contains the q-th signal with power $|\hat{\mathbf{f}}_p^* \mathbf{a}_q|^2$ and the noise with power $\sigma|\hat{\mathbf{f}}_p|^2$. We consider for any p and q the performance indices

$$\mathrm{ISR}_{pq} \stackrel{\mathrm{def}}{=} E\left\{|\hat{\mathbf{f}}_p^* \mathbf{a}_q|^2\right\} \quad \text{and} \quad \mathrm{INSR}_p \stackrel{\mathrm{def}}{=} E\left\{\frac{\sigma|\hat{\mathbf{f}}_p|^2 + \sum_{q \neq p} |\hat{\mathbf{f}}_p^* \mathbf{a}_q|^2}{|\hat{\mathbf{f}}_p^* \mathbf{a}_p|^2}\right\} \tag{24}$$

where the expectation is taken over realizations of T samples.

The first index is a pairwise Interference to Signal Ratio (ISR) measuring the rejection of the q-th source into the estimate of the p-th signal. It is actually a ratio since it is implicitely normalized by the convention (5) which implies that $\mathrm{ISR}_{pq} = \delta(p,q)$ if $\hat{\mathbf{f}}_p = \mathbf{f}_p = R_y^{\#} \mathbf{a}_p$. This index is used here to characterize the performance of LCMV filters since these are supposed to perfectly reject all the coherent jammers (i.e. the signals $s_q(t)$ for $q \neq p$). The second index is the natural measure of performance for the MVDR beamformer.

We call 'informed beamformers' the filters $\hat{\mathbf{f}}_p$ computed according to (3) or (4) using the *true* value of the directional vector $\mathbf{a}_p$. We call 'blind beamformers' the same filters computed using the *blind estimate* of $\mathbf{a}_p$ given by JADE. We refer to these filters by the obvious acronyms IMVDR, ILCMV, BMVDR and BLCMV.

In both cases, the sample statistics $\hat{R}_x$ or $\hat{R}_y$ are used. To be specific, the sample covariance is estimated by $\hat{R}_x = \sum_{t=1,T} \mathbf{x}(t)\mathbf{x}(t)^*$. To estimate $R_y^{\#}$, the sample covariance is eigendecomposed into $\hat{R}_x = \sum_{r=1,m} \mu_r \mathbf{h}_r \mathbf{h}_r^*$. An estimate $\hat{\sigma}$ of the noise variance is the average of the $m-n$ smallest eigenvalues. Ofter ordering of the eigenpairs, we form $\hat{R}_y^{\#} = \sum_{r=1,n} (\mu_r - \hat{\sigma})^{-1} \mathbf{h}_r \mathbf{h}_r^*$.

5.2 Numerical experiments

A first series of experiments is intended to compare blind versus informed beamformimg, to determine to which extent our cumulant based approach can accommodate significant noise levels and to get some indications as how this depends on the relevant parameters. We consider a linear $\lambda/2$ equispaced array of m unit-gain omnidirectional sensors and assume plane wave fronts. For convenience, we maintain the convention that the actual amplitude of each source is included in the corresponding directional vector. Thus, vector $\mathbf{a}_p$ takes the form $\mathbf{a}_p = \sigma_p^{1/2} \mathbf{a}(\phi_p)$ where σ_p is referred to as the power of the p-th source; $\phi_p \in [-\frac{1}{2}, \frac{1}{2}]$ is the 'electric angle', it depends on the physical location of the p-th source; finally, the l-th coordinate of vector $\mathbf{a}(\phi)$ is $exp(2j\pi l\phi)$. The experiments are conducted with temporally white signals, with $s_p(t)$ uniformly distributed on the unit circle for all p and t; the sample size is $T = 100$ for an array of $m = 10$ sensors.

Figures 2 to 4 have been obtained by averaging over $N = 500$ Monte Carlo runs. They obey an identical format: each figure shows blind and informed performance variation with respect to two parameters; the horizontal axis corresponds to the variation of one parameter; the other parameter takes 3 different values. Thus, each panel shows 3 solid lines corresponding to a blind beamformer and 3 dashed lines corresponding to a informed beamformer. Dotted lines have also been added for ease of reference, as follows: the left panel displays empirical values of ISR_{12} for BLCMV (solid) and ILCMV (dashed), and the dotted line is the reference level $1/T$; the right panel displays empirical values of INSR_1 for the BMVDR (solid) and for the IMVDR (dashed) and the dotted lines give the best theoretical value, i.e. the value for the filter (3) with perfect knowledge of the directional vector *and* of the covariance R_x. The curves need not be labelled: as expected performance decreases with increasing noise, decreasing source separation or source powers. Figure 2 shows the influence of noise level σ in dB for 3 different source configurations: $\phi_1 = 0$ and $\phi_2 = 0.02,\ 0.05,\ 0.1$. Source levels are $\sigma_1 = \sigma_2 = 0$ dB. Figure 3 shows, for three different noise levels $\sigma = -15,\ -5,\ 5$ dB, the influence of source separation: the first source is kept at $\phi_1 = 0$ while the second source is moved in the main lobe of the first one by varying ϕ_2. Source levels are $\sigma_1 = \sigma_2 = 0$ dB. Figure 4 shows the influence of the level σ_2 of the second source, considered as a 'jammer'. The noise level is kept at $\sigma = -5$ dB and the source locations are $\phi_1 = 0$ and $\phi_2 = 0.05$. The triples of curves are obtained by letting $\sigma_1 = -10,\ 0,\ 10$ dB.

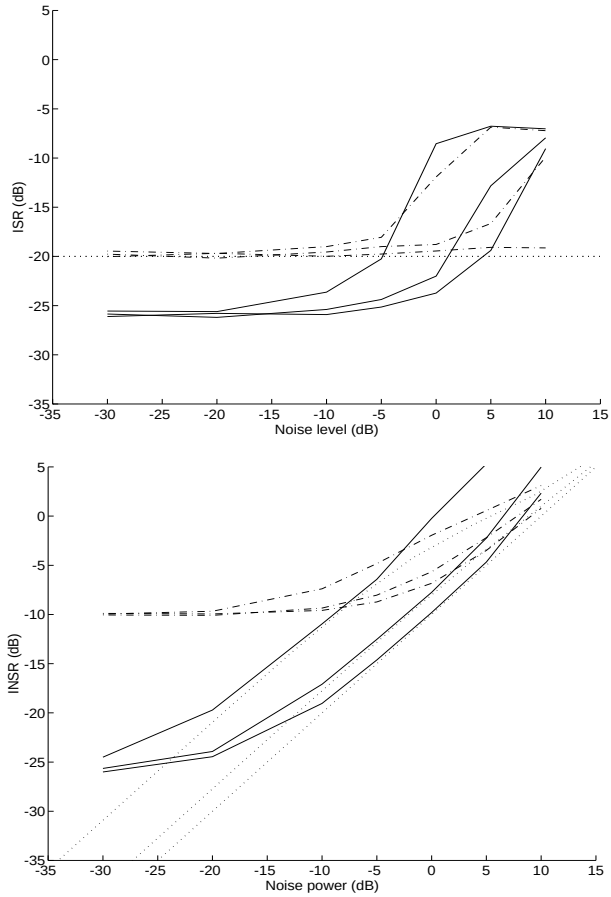

Figure 2: ISR_{12} (left) and INSR_1 (right) versus noise level σ.

The main conclusion to be drawn from these figures is that blind beamforming performs better than informed beamforming as long as the situation is 'not too hard' (poor conditioning of A). This is an instance of a more general statement that covariance based techniques better resist to significant noise levels (regardless of the noise distribution) than techniques involving higher order statistics. In the case of interest, the benefit of consistent blind estimation is traded for potential higher variance at poor SNRs. However it is a striking feature shown by these plots that, not only does the blind performer perform better than the informed one at high SNRs, but that it does so with a small $T = 100$ sample size even down to moderately low SNRs, depending on the parameters governing the 'hardness' of the task, such as source closeness. In fig. 2 for instance, blind advantage is maintained up to $\sigma > 5$ dB for $\phi_2 - \phi_1 = 0.1$ (this is the easiest case, where the directional vectors are orthogonal). Another striking feature, seen in figure 2 for instance, is that the ISR level tends to the same limit as σ goes to zero for various values of $\phi_2 - \phi_1$. This limit, as given by figure 2, is $1/T$ for the informed beamformer and is 6 dB lower for the blind beamformer. Performance of blind beamforming should also be studied as a function of the distribution of the source signals. In this respect, the 6 dB advantage observed with constant modulus sources is not expected to hold for other distributions (see below an illustration in the real case).

We have no complete explanation to why blind beamforming may perform better than informed beamforming. We cannot but acknowledge that errors in the estimate of R_y induce errors in $\hat{\mathbf{a}}_p$

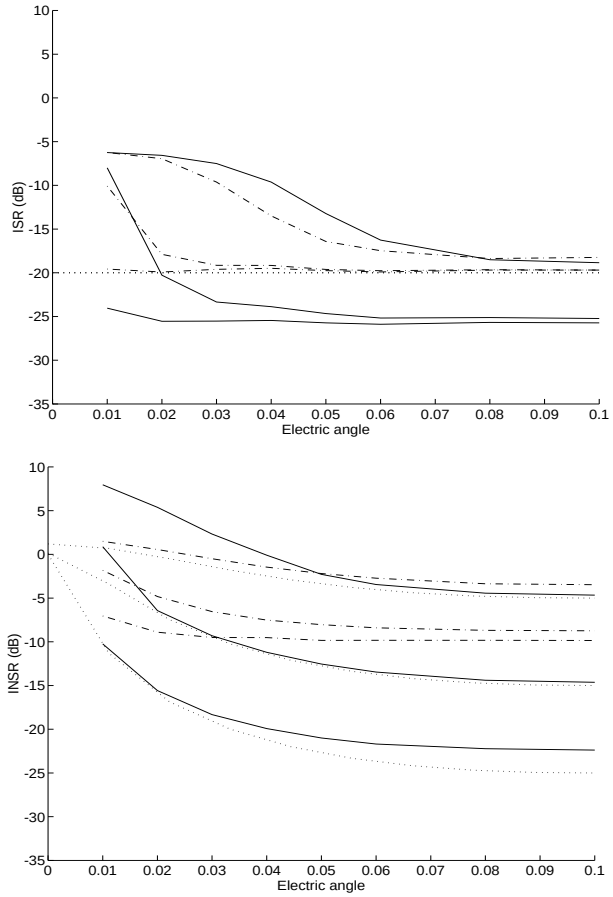

Figure 3: ISR_{12} (left) and INSR_1 (right) versus source separation $\phi_2 - \phi_1$.

which appear to be 'nicely' correlated and partially cancel when the BLCMV filter $\hat{\mathbf{f}}_p = \hat{R}_y^{\#}\hat{\mathbf{a}}_p$ is computed. No such thing happens when the true, fixed directional vector $\mathbf{a}_p$ is used together with an estimate of $R_y^{\#}$ to form the ILCMV filter $\hat{\mathbf{f}}_p = \hat{R}_y^{\#}\mathbf{a}_p$.

Next, we illustrate the benefit of jointly diagonalizing several cumulant matrices by comparing the performance of the JADE technique with the simulation results published in [5], where the mixing matrix is estimated by diagonalizing an unspecified linear combination of cumulant slices. In this example, matrix A is a real 4×3 matrix, the i.i.d. sources follow a one-sided exponential distribution and the noise level is $\sigma = 20$ dB. Table 1 shows the JADE RMS error for each entry of A, evaluated over 100 realizations of $T = 7000$ samples. This table can be compared with table 1 of [5]. The extra bottom line is the column-wise RMS error (square root of the sum of the squared entries of the given column); it shows an even distribution of the errors through all columns. For ease of reference, we have computed the corresponding line from [5]; it is [0.0213 0.0650 0.1365] which has a smaller first entry than in table 1 but larger next entries. The overall RMS (square root of the sum of the squared column-wise RMS errors) is computed to be 0.0654 through joint diagonalization and 0.1527 through ordinary diagonalization of [5]. Hence, in this example, joint diagonalization achieves a better overall performance and it does so with a ten times smaller sample size. Of course, this is only indicative since a better choice of the single cumulant matrix to be diagonalized may improve the performance reported in [5]. It is worth reporting that an

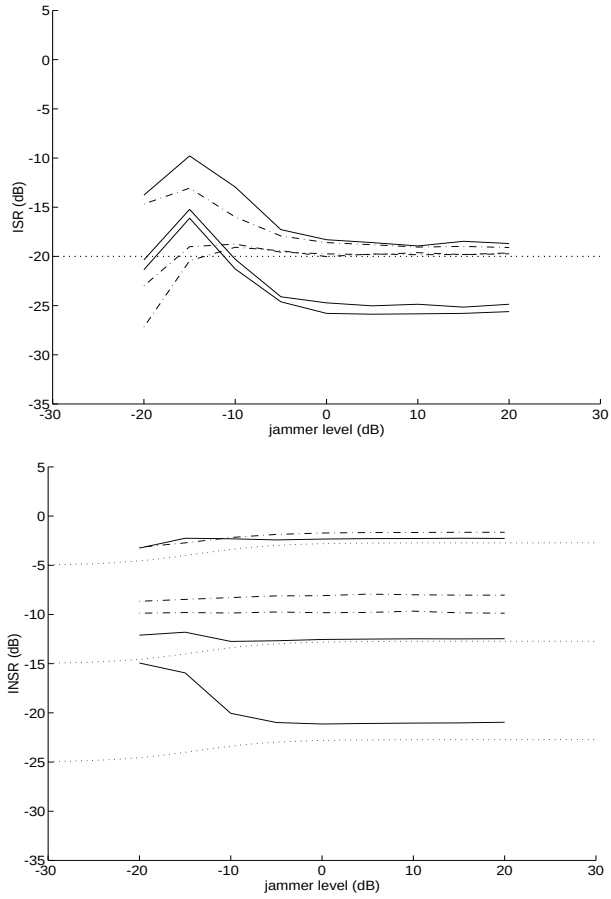

Figure 4: ISR_{12} (left) and INSR_1 (right) versus jammer level.

exponential distribution does not seem to be very favorable to blind identification: in the same experiment conducted with binary sources, the JADE technique achieves an overall RMS value of 0.051 with $T = 700$ samples; entrywise RMS appears in table 2.

Conclusion

Joint diagonalization of cumulant matrices allows the whole 4th-order cumulant set to be processed with a computational efficiency similar to eigen-based techniques. The resulting blind identification scheme (JADE) has been applied to narrow band beamforming. In this application, directional vectors are *blindly estimated* rather than modeled via a (possibly problematic) array manifold, making the blind technique insensitive to array mismatch and pointing errors. Numerical simulations show that, in a significant range of parameters, blind beamforming may outperform informed beamformers (whose performance is limited by finite sample size) even when the latters use the true directional vector. This rather surprizing fact should be theoretically supported by asymptotic performance analysis of the JADE estimator.

Acknowledgements

Comments and suggestions by the anonymous reviewers have greatly contributed to improving the first version of this paper. Part of this work has been supported by Thomson CSF/RGS/STS.

RMS errors		
0.0149	0.0153	0.0173
0.0162	0.0170	0.0179
0.0196	0.0229	0.0229
0.0194	0.0216	0.0194
0.0353	0.0389	0.0390

Table 1: Entrywise and columnwise RMS error in $\hat{A}$ in a 4×3 example. JADE algorithm with $T = 7000$ samples and one-sided exponentially distributed sources.

RMS errors		
0.0150	0.0083	0.0133
0.0168	0.0102	0.0122
0.0088	0.0176	0.0229
0.0194	0.0134	0.0109
0.0310	0.0258	0.0311

Table 2: Entrywise and columnwise RMS error in $\hat{A}$ in a 4×3 example. JADE algorithm with $T = 700$ samples and binary sources.

Appendix A: A joint diagonalization algorithm.

The Jacobi technique [29] for diagonalizing a unique hermitian matrix is extended for the joint approximate diagonalization of a set $\mathcal{N} = \{N_r | 1 \leq r \leq s\}$ of arbitrary $n \times n$ matrices. It consists in minimizing the diagonalization criterion (19) by successive Givens rotations. We start by describing the 2×2 case and we denote

$$N_r = \begin{bmatrix} a_r & b_r \\ c_r & d_r \end{bmatrix} \tag{25}$$

for $r = 1, \ldots, s$. A complex 2×2 Givens rotation is

$$V = \begin{bmatrix} \cos\theta & -e^{j\phi}\sin\theta \\ e^{-j\phi}\sin\theta & \cos\theta \end{bmatrix}, \tag{26}$$

Denoting a'_r, b'_r, c'_r and d'_r the coefficients of $V^H N_r V$, optimization of (19) amounts to finding θ and ϕ such that $\sum_r |a'_r|^2 + |d'_r|^2$ is maximized. Note that $2(|a'_r|^2 + |a'_r|^2) = |a'_r - d'_r|^2 + |a'_r + d'_r|^2$ and that the trace $a'_r + d'_r$ is invariant in a unitary transformation, maximization of criterion (19) is equivalent to maximization of Q:

$$Q \stackrel{\text{def}}{=} \sum_r |a'_r - d'_r|^2 \tag{27}$$

It is easily checked that

$$a'_r - d'_r = (a_r - d_r)\cos 2\theta + (b_r + c_r)\sin 2\theta \cos\phi + j(c_r - b_r)\sin 2\theta \sin\phi \tag{28}$$

for r=1, ..., s. Then by defining the vectors

$$u \stackrel{\text{def}}{=} [a'_1 - d'_1, \cdots, a'_s - d'_s]^T \tag{29}$$

$$v \stackrel{\text{def}}{=} [\cos 2\theta, \sin 2\theta \cos\phi, \sin 2\theta \sin\phi]^T \tag{30}$$

$$g_r \stackrel{\text{def}}{=} [a_r - d_r, b_r + c_r, j(c_r - b_r)]^T \tag{31}$$

the s equations (28) may be written in the form $u = Gv$ where $G^T \stackrel{\text{def}}{=} [g_1, \cdots, g_s]$ so that Q also reads

$$Q = u^H u = v^T G^H G v = v^T \, Real(G^H G) \, v \tag{32}$$

where we have used that, $G^H G$ being hermitian by construction, its imaginary part is anti-symmetric, hence contributing nothing to the above quadratic form. The last step is to recognize that the particular parameterization (30) of v is equivalent to the condition $v^T v = 1$. Thus the optimal v is the eigenvector of $Re(G^H G)$ associated to the largest eigenvalue, which is easily computed for a real 3×3 symmetric matrix. Further, the entries of the Givens rotation can computed from the coordinates of v without even using trigonometrics as in the standard Jacobi technique [29].

Appendix B : Proofs

In order to prove propositions 1 and 4, we establish a more general lemma.
Lemma. *For any set $\{B_r | 1 \leq r \leq n^2\}$ of orthonormal $n \times n$ matrices, the identity $c(V) = C(V, \{Q_z(B_r) | 1 \leq r \leq n^2\})$ holds for any unitary matrix $V = [\mathbf{v}_1, \cdots, \mathbf{v}_n]$.*
Proof: by the following chain of identities.

$$\begin{aligned} C(V, \{Q_z(B_r) | 1 \leq r \leq n^2\}) &\stackrel{1}{=} \sum_{i,r} |\mathbf{v}_i^* Q_z(B_r) \mathbf{v}_i|^2 \stackrel{2}{=} \sum_{i,r} |\text{Trace}(B_r Q_z(\mathbf{v}_i \mathbf{v}_i^*))|^2 \\ \stackrel{3}{=} \sum_i ||Q_z(\mathbf{v}_i \mathbf{v}_i^*)||^2_{\text{Fro}} &\stackrel{4}{=} \sum_{i,k,l} |\mathbf{v}_k^* Q_z(\mathbf{v}_i \mathbf{v}_i^*) \mathbf{v}_l|^2 \stackrel{5}{=} \sum_{i,k,l} |Cum(e_k, e_l^*, e_i, e_i^*)|^2 \stackrel{6}{=} c(V) \end{aligned}$$

Equality 1 is a rewriting of the joint diagonalization criterion. Equality 2 is an instance of the identity $\mathbf{v}^* Q_z(B) \mathbf{v} = Trace(B Q_z(\mathbf{v}\mathbf{v}^*))$ resulting, for any matrix B and vector $\mathbf{v}$, from definition (11). The matrix sets $\{B_r^H | 1 \leq r \leq n^2\}$ and $\{\mathbf{v}_k \mathbf{v}_l^* | 1 \leq k, l \leq n\}$ are two orthonormal basis for the space of $n \times n$ matrices; expressing the Frobenius norm of $Q_z(\mathbf{v}_i \mathbf{v}_i^*)$ onto each of these two sets yields equalities 3 and 4 respectively. Equality 5 comes by the multilinearity of the cumulants using $e_i = \mathbf{v}_i^* \mathbf{z}$. Finally, 6 uses the cumulant symmetries.□
Proof of proposition 1. Using the lemma, since $\{(\mathbf{b}_l \mathbf{b}_k^*) | 1 \leq k, l \leq n\}$ is an orthonormal set.□
Proof of proposition 2. We state without proof the following simple property. *If $\mathcal{N}$ is a set of s matrices in the form $\mathcal{N} = \{M_r | M_r = V \Lambda_r V^H, 1 \leq r \leq s\}$ where each Λ_r is diagonal and V is unitary, and if $\tilde{V}$ is joint diagonalizer of $\mathcal{N}$, then matrix $\tilde{V}^H M_r \tilde{V}$ is diagonal for any M_r in $\mathcal{N}$.* Thus, if $\tilde{V}$ is a joint diagonalizer of $\mathcal{N}^p$, each matrix $\tilde{V}^H Q_z(\mathbf{b}_l \mathbf{b}_k^*) \tilde{V}$ is diagonal. By linearity of Q_z, this is also true for any linear combination of the matrices $\mathbf{b}_l \mathbf{b}_k^*$ and we conclude that $\tilde{V}^H Q_z(N) \tilde{V}$ is then diagonal for any matrix N. Of course this property holds for $\tilde{V} = U$; we have to establish that it holds only if $\tilde{U}$ is essentially equal to U.

Assume first that all the sources are kurtic and set $N = \sum_{p=1,n} p \, k_p^{-1} \mathbf{u}_p \mathbf{u}_p^*$. The eigenvalues of $Q_z(N)$, which are $1, \cdots, n$ according to (13), are distinct indeed so that the unitary diagonalizer of $Q_z(N)$ is essentially unique and then essentially equal to U. Second if one source has a zero kurtosis, we set its contribution in N to zero so that $Q_z(N)$ now has the first n integers as eigenvalues but for one which is zero. These are distinct numbers and the conclusion still holds.□
Proof of proposition 4. In text.
Proof of proposition 3. Apply the lemma with the orthonormal basis $\{M_r | 1 \leq r \leq n^2\}$ made of the eigenmatrices of Q_z. These verify $Q_z(M_r) = \lambda_r M_r$ for $1 \leq r \leq n^2$ but under H0-3, there are at most n non zero eigenvalues. Discarding $n(n-1)$ matrices $\lambda_r M_r$ with $\lambda_r = 0$ does not affect criterion (19) and yields the eigen-set.□

References

[1] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, pp. 4–24, Apr. 1988.

[2] L. Tong, R. Liu, V. Soon, and Y. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Tr. on CS*, vol. 38, pp. 499–509, May 1991.

[3] J.-F. Cardoso, "Source separation using higher order moments," in *Proc. ICASSP*, pp. 2109–2112, 1989.

[4] 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.

[5] L. Tong, Y. Inouye, and R. Liu, "Waveform preserving blind estimation of multiple independent sources," *IEEE Tr. on SP*, vol. 41, pp. 2461–2470, July 1993.

[6] P. Comon, "Independent component analysis," in *Proc. Int. Workshop on Higher-Order Stat., Chamrousse, France*, pp. 111–120, 1991.

[7] M. Gaeta and J.-L. Lacoume, "Source separation without a priori knowledge: the maximum likelihood solution," in *Proc. EUSIPCO*, pp. 621–624, 1990.

[8] D. Donoho, "On minimum entropy deconvolution," in *Applied time-series analysis II*, pp. 565–609, Academic Press, 1981.

[9] B. Porat and B. Friedlander, "Direction finding algorithms based on high-order statistics," *IEEE Tr. on SP*, vol. 39, pp. 2016–2025, Sept. 1991.

[10] G. Giannakis and S. Shamsunder, "Modelling of non Gaussian array data using cumulants: DOA estimation of more sources with less sensors," *Signal Processing*, vol. 30, pp. 279–297, July 1993.

[11] J.-F. Cardoso and P. Comon, "Tensor based independent component analysis," in *Proc. EUSIPCO*, 1990.

[12] G. Giannakis, Y. Inouye, and J. Mendel, "Cumulant-based identification of multichannel moving average models," *IEEE Tr. on AC*, vol. 34, pp. 783–787, 1989.

[13] L. Tong and R. Liu, "A new eigenstructure based parameter estimation of multichannel moving average processes," in *Proc. ICASSP*, vol. 5, pp. 249–252, 1992.

[14] P. Loubaton and P. A. Regalia, "Blind deconvolution of multivariate signals: a deflation approach," in *Proc. ICC*, vol. 3, pp. 1160–1165, 1993.

[15] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Tr. on IT*, vol. 36, no. 2, pp. 312–321, 1990.

[16] D. Godard, "Self recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Tr. on Comm.*, vol. 28, no. 11, pp. 1867–1875, 1980.

[17] S. Mayrargue, "Spatial equalization of a radio-mobile channel without beamforming using the constant modulus algorithm (CMA)," in *Proc. ICASSP*, 1993.

[18] J.-F. Cardoso, "Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors," in *Proc. ICASSP*, pp. 3109–3112, 1991.

[19] J.-F. Cardoso, "Iterative techniques for blind source separation using only fourth order cumulants," in *Proc. EUSIPCO*, pp. 739–742, 1992.

[20] J.-F. Cardoso, "Fourth-order cumulant structure forcing. Application to blind array processing," in *Proc. 6th SSAP workshop on statistical signal and array processing*, pp. 136–139, Oct. 1992.

[21] J.-F. Cardoso and B. Laheld, "Adaptive blind source separation for channel spatial equalization," in *Proc. of COST 229 workshop on adaptive signal processing*, pp. 19–26, 1992.

[22] E. Moreau and O. Macchi, "New self-adaptive algorithms for source separation based on contrast functions," in *Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA*, pp. 215–219, 1993.

[23] C. Jutten and J. Hérault, "Une solution neuromimétique au problème de séparation de sources," *Traitement du Signal*, vol. 5, no. 6, pp. 389–403, 1988.

[24] L. Féty and J. P. Van Uffelen, "New methods for signal separation," in *Proc. of 4th Int. Conf. on HF radio systems and techniques*, (London), pp. 226–230, IEE, Apr. 1988.

[25] A. Dinç and Y. Bar-Ness, "Bootstrap: A fast blind adaptive signal separator," in *Proc. ICASSP*, vol. 2, pp. 325–328, 1992.

[26] 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.

[27] L. Tong and R. Liu, "Blind estimation of correlated source signals," in *Proc. Asilomar conference*, Nov. 1990.

[28] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and Éric Moulines, "Second-order blind separation of correlated sources," in *Proc. Int. Conf. on Digital Sig. Proc.*, (Cyprus), pp. 346–351, 1993.

[29] G. H. Golub and C. F. V. Loan, *Matrix computations*. The Johns Hopkins University Press, 1989.