Blind beamforming for non-Gaussian signals

J.F. Cardoso
A. Souloumiac

Indexing terms: Blind identification, Beamforming

Abstract: The paper considers an application of blind identification to beamforming. The key point is to use *estimates* of directional vectors rather than resort to their hypothesised value. By using estimates of the directional vectors obtained via blind identification, i.e. without knowing the array 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 modelling is necessary. Rather suprisingly, '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 on which blind identification relies is the statistical independence of the sources, which is exploited using fourth-order cumulants. A computationally efficient technique is presented for the blind estimation of directional vectors, based on joint diagonalisation of fourth-order cumulant matrices; its implementation is described, and its performance is investigated by numerical experiments.

1 Introduction

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

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

$$\boldsymbol{x}(t) = \boldsymbol{y}(t) + \boldsymbol{n}(t) \tag{2}$$

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

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

$$\boldsymbol{f}_p = R_x^{-1}\boldsymbol{a}_p \quad \text{with } R_x \overset{\text{def}}{=} E\{\boldsymbol{x}(t)\boldsymbol{x}(t)^*\} \tag{3}$$

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

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

where $^{\#}$ denotes the pseudo-inverse. Note that this filter can be computed without knowing the directional vectors except for the one corresponding to the signal of interest.

These two standard approaches, MVDR and LCMV, are based on knowledge of the directional vector associated with 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 reflections at the vicinity of the array (errors in modelling the propagation). Even if the array is perfectly known, pointing errors cause performance degradation. Finally, performance is limited by the use in eqn. 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 model-

Paper 9884F (E5), first received 11th March and in revised form 28th September 1993

The authors are with Télécom Paris, 46 rue Barrault, 75634 Paris, France

Comments and suggestions from 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.

ling of the propagation or array calibration. At first sight, such an approach may appear paradoxical as the array manifold is the link between the location of a source of interest and its associated directional vector. The relevance of blind identification to beamforming must therefore be discussed.

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 of eqns. 3 or 4 is straightforward. There is a clear benefit in processing without knowledge of the array manifold, as 'blind' beamforming is, in essence, insensitive to errors in the manifold model. In the presence of several coherent sources however, the blind approach cannot stand by itself as it typically yields the directional vectors associated with all sources. Unless all the received signals are of interest, without discrimination, some additional processing is then necessary to select among the estimated directional vectors those associated with the sources of interest. One approach is to form beams corresponding to all the detected signals and select the signals of interest using 'nonspatial' information (spectral content, modulation etc.). Another approach is to select among the blindly estimated directional vectors the closest to the directional vector predicted by physical modelling, 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 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 reflections (or multiple paths), however, cause a single emitter to contribute 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, whereas it severely affects parametric array processing. This is because two fully coherent sources (e.g. 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 output is $s_1(t)\boldsymbol{a}_1 + s_2(t)\boldsymbol{a}_2 = s_1(t)(\boldsymbol{a}_1 + \alpha\boldsymbol{a}_2)$, which is seen as a single source with a composite 'directional' vector $\boldsymbol{a}_1 + \alpha\boldsymbol{a}_2$. This would be a problem for any method assuming that each independent component is associated with some direction. The blind approach does not make any such assumption as 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, whereas 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 mentioning, in anticipation, that the columns of A are blindly estimated up to a scalar factor. This is irrevelant in many applications; in particular, such a factor in $\boldsymbol{a}_p$ does not change the SNR at the output of spatial filters such as eqns. 3 or 4.

2 Blind estimation of the array matrix

2.1 Blind identifiability

Before proceeding, it is important to specify the notion of blind identification. As each source contributes to the array output via the product $s_p(t)\boldsymbol{a}_p$, the process observed is unaffected by the exchange of a complex scalar factor between each source signal $s_p(t)$ and each vector $\boldsymbol{a}_p$. 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 Reference 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 $1 \leqslant p \leqslant n$, 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

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

This normalisation 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 there exists a matrix P such that $M = NP$ where P has exactly one nonzero 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 parameterised.

2.2 Notations and assumptions

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

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

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

H_0: the processes $\boldsymbol{n}(t), s_1(t), \ldots, s_n(t)$ are jointly stationary.

The *kurtosis* of the pth source is the real number

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

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

H_1: there is at most one nonkurtic source.

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

H_2: the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are linearly independent but otherwise arbitrary.

H_3: the variables $s_1(t), \ldots, s_n(t)$ are statistically independent for each t.

We will see that under H_{1-3}, 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

H_4: there exist consistent estimates for R_x and $\mathcal{Q}_x$.

H_5: the additive noise is normally distributed and independent of the sources.

H_6: the additive noise is spatially white $R_n = \sigma I_m$, with unknown variance σ and $n < m$.

By H_5, an estimate of $\mathcal{Q}_x$ is also an estimate of $\mathcal{Q}_y$ as cumulants are additive for independent variables and as higher-order cumulants are zero for normally distributed variables. By H_6, an estimate of R_y can be classically constructed from the eigendecomposition of an estimate of R_x. We insist that assumptions H_{4-6} could be replaced by any other assumption set serving the same purpose: guaranteeing 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 $y(t)$ of the observation. This is done via a whitening matrix W, i.e. an $n \times m$ matrix such that $Wy(t)$ is spatially white. The whiteness condition is

$$I_n = WR_y W^H = WAA^H W^H \tag{8}$$

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

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

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 eqn. 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 $z(t) = Wx(t)$ still obeys a linear model:

$$z(t) \stackrel{\text{def}}{=} Wx(t) = W(As(t) + n(t)) = Us(t) + Wn(t) \tag{10}$$

The signal part of the whitened process is now 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 eqn. 10 to any other unitary matrix leaves unchanged the covariance of $z(t)$.

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 diagonaliser of an $n \times n$ 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 optimiser 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 optimisation problem. These approaches are reviewed in Sections 3.1 and 3.2; 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: with any $n \times n$ matrix M is associated a 'cumulant matrix' denoted by $Q_z(M)$, defined entrywise by

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

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 $M = \boldsymbol{b}_l \boldsymbol{b}_k^*$, where $\boldsymbol{b}_k$ denotes the $n \times 1$ vector with 1 in the kth 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' $\mathcal{M}^p$ as the set of all the parallel slices:

$$\mathcal{M}^p \stackrel{\text{def}}{=} \{Q_z(\boldsymbol{b}_l \boldsymbol{b}_k^*) \mid 1 \leqslant k, l \leqslant n\} \tag{12}$$

As $z(t)$ obeys the linear model eqn. 10, the cumulant matrices take a simple form. Using the cumulant properties (Gaussian rejection, additivity, multilinearity), it is straightforward to establish that

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

or equivalently

$$Q_z(M) = U \Lambda_M U^H \tag{14a}$$

$$\Lambda_M \stackrel{\text{def}}{=} \text{Diag}\,(k_1 \boldsymbol{u}_1^* M \boldsymbol{u}_1, \ldots, k_n \boldsymbol{u}_n^* M \boldsymbol{u}_n) \tag{14b}$$

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

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

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

Unitary diagonalisation of $Q_z(I_n)$ is not essentially determined, however, if some eigenvalues are identical. These eigenvalues, by eqn. 14, are $k_p \boldsymbol{u}_p^* I_n \boldsymbol{u}_p = k_p |\boldsymbol{u}_p|^2 = k_p$, so that the case of degeneracy is when some sources have identical kurtosis. It is suggested in Reference 5 that a linear combination of cumulant slices is diagonalised, that is $Q_z(M)$ is diagonalised for some matrix M. The pth eigenvalue of $Q_z(M)$ being $k_p \boldsymbol{u}_p^* M \boldsymbol{u}_p$, degeneracy is very unlikely. This approach suffers from two drawbacks. First, there is no guideline as to how to choose *a priori* the linear combination of cumulant slices, or equivalently, as to how to choose M before evaluating $Q_z(M)$. Secondly, 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 optimisation of cumulant criteria

Let V denote an $n \times n$ unitary matrix and further define $\boldsymbol{e}(t)$, as in Fig. 1,

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

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

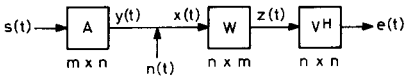

Fig. 1 *Inverting A by chaining a whitener and a unitary matrix*

It has then been proposed to determine U as the unitary minimiser of the sum of all the squared cross-cumulants in $\mathcal{Q}_e$. As 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 *maximising* under unitary constraint the criterion [6]

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

This criterion first appeared in Reference 7 where it was obtained via a fourth-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 Reference 8. Comon also describes an algorithm for maximising eqn. 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 maximiser of the criterion $c(V)$:

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

which is equivalent to minimising the sum of the squared cross-cumulants with distinct first and second indices. The main reason for considering the criterion of eqn. 18 is its links to underlying eigenstructures that allows for an efficient optimisation of it by means of 'joint diagonalisation'.

3.3 Joint diagonalisation

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

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

where $|\text{diag}\,(\cdot)|$ 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 diagonalisation is equivalent to usual unitary diagonalisation. If the set $\mathcal{N}$ cannot be exactly jointly diagonalised (this is the case when *sample* cumulants are processed), the unitary maximisation of eqn. 19 defines a somewhat arbitrary but quite natural 'joint approximate diagonalisation'.

The link between optimisation-based and eigen-based blind-identification techniques is established by considering the joint diagonalisation 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 maximisation of $c(V)$ is equivalent to the joint diagonalisation of the parallel set. Blind identifiability via joint diagonalisation is guaranteed by the following proposition.

Proposition 2: Under H_{1-3}, a joint diagonaliser of $\mathcal{N}^p$ is essentially equal to U.

Proofs of these propositions are sketched in Appendix 8.2.

Joint diagonalisation corresponding to usual diagonalisation when only one hermitian matrix is involved, it is no surprise that the Jacobi technique can be extended to the joint diagonalisation of several matrices. This extension is described in Appendix 8.1 and offers a computation cost that is roughly s times the cost of diagonalising a single matrix. In addition, for the particular problem at hand, this cost can be further reduced by initialising the joint diagonaliser with the unitary matrix returned by the (ordinary) diagonalisation of a single cumulant matrix. A convenient choice is to diagonalise some $Q_z(M)$, where M is a random *hermitian* matrix, because then matrix $Q_z(M)$ also is hermitian.

3.4 Representation of fourth-order cumulants by eigenmatrices

The computational efficiency of joint diagonalisation 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 $\boldsymbol{v}$ with fourth-order cumulants, there exist d^2 real numbers $\lambda_1, \ldots, \lambda_{d^2}$ and d^2 matrices $M_1, \ldots, M_{d^2}$, called eigenmatrices, verifying

$$Q_v(M_r) = \lambda_r M_r \quad \text{Tr}\,(M_r M_s^H) = \delta(r, s)$$
$$1 \leqslant r, s \leqslant d^2 \tag{20}$$

The proof is straightforward by a classic 'stacking-unstacking' device: the relationship $N = Q_v(M)$ 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 $1 \leqslant a, b \leqslant d^2$, by

$$\tilde{N}_a = n_{ij} \quad \tilde{M}_a = m_{ij} \quad \tilde{Q}_{ab} = \text{Cum}\,(v_i, v_j^*, v_l, v_k^*)$$

with

$$a = i + (j-1)d \quad b = k + (l-1)d \tag{21}$$

Matrix $\tilde{Q}$ is easily checked to be hermitian. It then admits a set of d^2 real eigenvalues $\lambda_1, \ldots, \lambda_{d^2}$ and d^2 corresponding eigenvectors whose unstacking, as in eqn. 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 Reference 9.

The eigen-structure of $\mathcal{Q}_z$ derives from eqn. 13. It is readily checked that the set $\{\boldsymbol{u}_p \boldsymbol{u}_q^* | 1 \leqslant p, q \leqslant n\}$ verifies the properties of proposition 3. Orthonormality of the matrices in this set stems from U being unitary, and, by substitution into eqn. 13, it is found that $Q_z(\boldsymbol{u}_p \boldsymbol{u}_p^*) = k_p \boldsymbol{u}_p \boldsymbol{u}_p^*$, and $Q_z(\boldsymbol{u}_p \boldsymbol{u}_q^*) = 0$ 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 Reference 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

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

For our purpose, the eigen-set contains the relevant fourth-order information, as we have

Proposition 4: Under H_{0-3} 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 maximisation of $c(V)$ computationally attractive.

4 Blind-identification algorithms

4.1 The JADE algorithm

A blind-identification algorithm by joint approximate diagonalisation of eigen-matrices (JADE)* 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 fourth-order cumulants $\hat{\mathcal{Q}}_z$ of the whitened process $\hat{z}(t) = \hat{W}x(t)$; compute the n most significant eigenpairs $\{\hat{\lambda}_r, \hat{M}_r \mid 1 \leqslant r \leqslant n\}$.

Step 3: Jointly diagonalise the set $\hat{\mathcal{N}}^e = \{\hat{\lambda}_r \hat{M}_r \mid 1 \leqslant r \leqslant n\}$ 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 second-order statistics and is standard under $H_{5,6}$; 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 $\mu_1, \ldots, \mu_n$ the n largest eigenvalues and $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_n$ the corresponding eigenvectors of $\hat{R}_x$. A whitener is $\hat{W} = [(\mu_1 - \hat{\sigma})^{-1/2}\boldsymbol{h}_1, \ldots, (\mu_n - \hat{\sigma})^{-1/2}\boldsymbol{h}_n]^H$. We do not address the important issue of detecting the number n of sources.

In step 2, computation of the eigenmatrices amounts to diagonalising an $n^2 \times n^2$ matrix made from the elements of $\hat{\mathcal{Q}}_z$. A standard algorithm for eigendecomposition of hermitian matrices will do perfectly, 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 diagonalisation is performed on the parallel set $\mathcal{N}^p$. An even simpler implementation is to form a set $\mathcal{N} = \{Q_z(C_r) \mid 1 \leqslant r \leqslant s\}$ (possibly using the sample counterpart of eqn. 15), where the C_rs 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 $\mathcal{N}^p$ is used.

Step 3 is implemented by extending the single-matrix Jacobi technique to several matrices, as described in Appendix 8.1. Note that when $n = 2$, the Jacobi technique is not iterative: a unique Givens rotation achieves (joint) diagonalisation. Also recall that joint diagonalisation may be initialised with the (usual) diagonaliser of a single cumulant matrix.

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

4.2 Related approaches

Besides References 3–8 already mentioned in Section 3, other contributions are related to blind identification of the model (eqns. 1 and 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 modelling could be applied, provided they are extended to the complex case. See for instance the cumulant-based approach in References 12–14 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 technique in Reference 15 closely parallels the unitary maximisation of eqn. 17 or the 'reverse criterion' of Reference 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 modelling the spatial structure of the noise as long as it is independent and normally distributed. References 10, 13 and 18–20 specifically consider the spatial problem.

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

Finally, simple solutions can also be implemented if the model (eqns. 1 and 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 diagonalising a correlation matrix $E\{z(t+\tau)z(t)^*\}$ rather than a cumulant matrix. This was independently proposed in References 24 and 27. As with cumulant matrices, indetermination problems may occur, and several correlation matrices (i.e. for various τ) may be jointly diagonalised for the sake of robustness, as shown in Reference 28. A necessary identifiability condition is that the source signals have different spectra. A safe approach may consist of the joint diagonalisation 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. Let $\hat{\boldsymbol{f}}_p$ denote an estimate of a spatial filter computed from T data samples when p is the signal of

* A Matlab implementation of JADE is available upon request from cardoso@sig.enst.fr.

interest. The estimated signal is

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

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

$$ISR_{pq} \overset{\text{def}}{=} E\{|\hat{f}_p^* a_q|^2\}$$

and

$$INSR_p \overset{\text{def}}{=} E\left\{\frac{\sigma|\hat{f}_p|^2 + \sum_{q \neq p} |\hat{f}_p^* a_q|^2}{|\hat{f}_p^* a_p|^2}\right\} \tag{24}$$

where the expectation is taken over realisations of T samples.

The first index is a pairwise interference-to-signal ratio (ISR) measuring the rejection of the qth source in the estimate of the pth signal. It is actually a ratio as it is implicitly normalised by the convention of eqn. 5, which implies that $ISR_{pq} = \delta(p, q)$ if $\hat{f}_p = f_p = R_y^{\#} a_p$. This index is used here to characterise the performance of LCMV filters as these are supposed to reject perfectly 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 $\hat{f}_p$ computed according to eqn. 3 or 4 using the true value of the directional vector a_p. We call 'blind beamformers' the same filters computed using the blind estimate of 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 $\hat{R}_x = \sum_{t=1,T} x(t)x(t)^*$. To estimate $R_y^{\#}$, the sample covariance is eigendecomposed into $\hat{R}_x = \sum_{r=1,m} \mu_r h_r h_r^*$. An estimate $\hat{\sigma}$ of the noise variance is the average of the $m - n$ smallest eigenvalues. After ordering the eigenpairs, we form $\hat{R}_y^{\#} = \sum_{r=1,n} (\mu_r - \hat{\sigma})^{-1} h_r h_r^*$.

5.2 Numerical experiments

A first series of experiments is intended to compare blind with informed beamforming, to determine to what extent our cumulant-based approach can accommodate significant noise levels and to obtain some indications as to 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 a_p takes the form $a_p = \sigma_p^{1/2} a(\phi_p)$, where σ_p is referred to as the power of the pth source; $\phi_p \in [-\frac{1}{2}, \frac{1}{2}]$ is the 'electric angle', and it depends on the physical location of the pth source; finally, the lth co-ordinate of vector $a(\phi)$ 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.

Figs. 2–4 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 three different values. Thus, each panel shows three solid lines corresponding to a blind beamformer and three broken lines corresponding to an informed beamformer. Dotted lines have also been added for ease of reference, as follows: the upper panel displays empirical values of ISR_{12} for BLCMV (solid) and ILCMV (broken), and the dotted

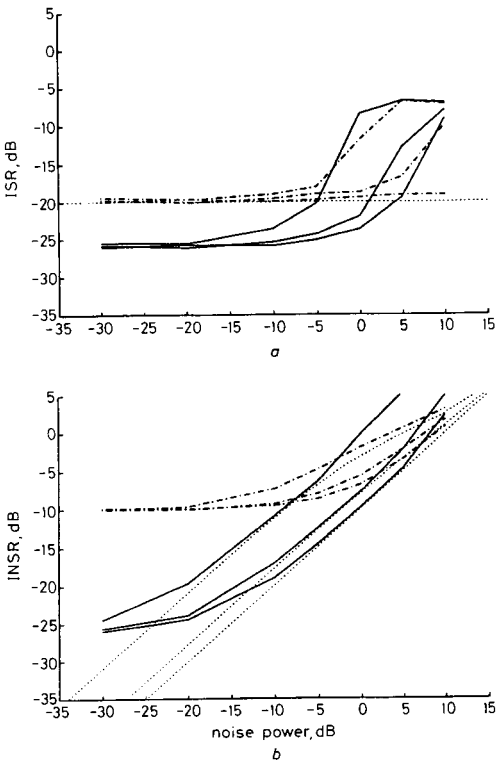

Fig. 2 *ISR_{12} and $INSR_1$ versus noise level σ*
$\phi_1 = 0$; $\phi_2 = 0.02, 0.05, 0.1$
a ISR_{12}
——— BLCMV
–·–·– ILCMV
····· reference level $1/T$
b $INSR_1$
——— BMVDR
–·–·– IMVDR
····· best theoretical value

line is the reference level $1/T$; the lower panel displays empirical values of $INSR_1$ for the BMVDR (solid) and for the IMVDR (broken), and the dotted lines give the best theoretical value, i.e. the value for the filter (eqn. 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, and decreasing source separation or source powers.

Fig. 2 shows the influence of noise level σ in decibels for three different source configurations: $\phi_1 = 0$ and $\phi_2 = 0.02$, 0.05, 0.1. Source levels are $\sigma_1 = \sigma_2 = 0$ dB. Fig. 3 shows, for three different noise levels $\sigma = -15$, -5, 5 dB, the influence of source separation: the first source is kept at $\phi_1 = 0$, and 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. Fig. 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 triplet of curves are obtained by letting $\sigma_1 = -10$, 0, 10 dB.

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 resist significant noise levels (regardless of the noise distribution) better than techniques involving higher-order statistics. In the case of interest, the benefit of consistent blind estimation is traded for potential higher

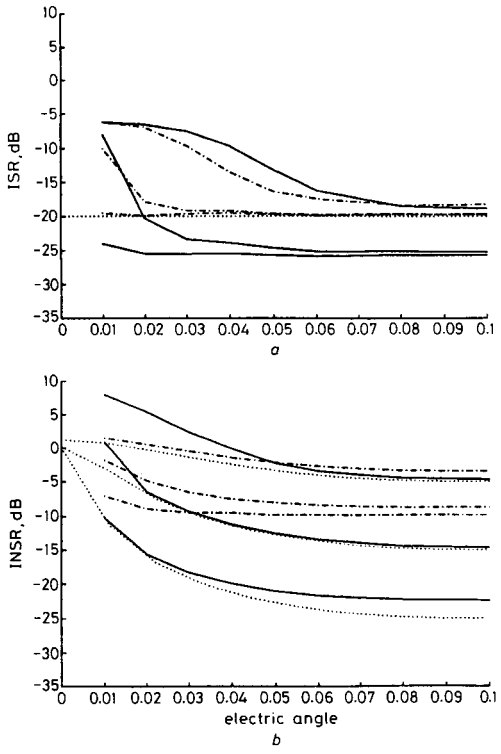

Fig. 3 *ISR_{12} and $INSR_1$ versus source separation $\phi_2 - \phi_1$*
$\sigma = -15, -5, 5$ dB
a ISR_{12} — BLCMV; -·-·- ILCMV; ····· reference level $1/T$
b $INSR_1$ — BMVDR; -·-·- IMVDR; ····· best theoretical value

variance at poor SNRs. However, it is a striking feature shown by these plots that, not only does the blind beamformer 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 Fig. 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 Fig. 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 as 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 $\hat{a}_p$ that appear to be 'nicely' correlated and partially cancel when the BLCMV filter $\hat{f}_p = \hat{R}_y^{\#} \hat{a}_p$ is computed. No such thing happens when the true, fixed directional vector a_p is used together with an estimate of $R_y^{\#}$ to form the ILCMV filter $\hat{f}_p = \hat{R}_y^{\#} a_p$.

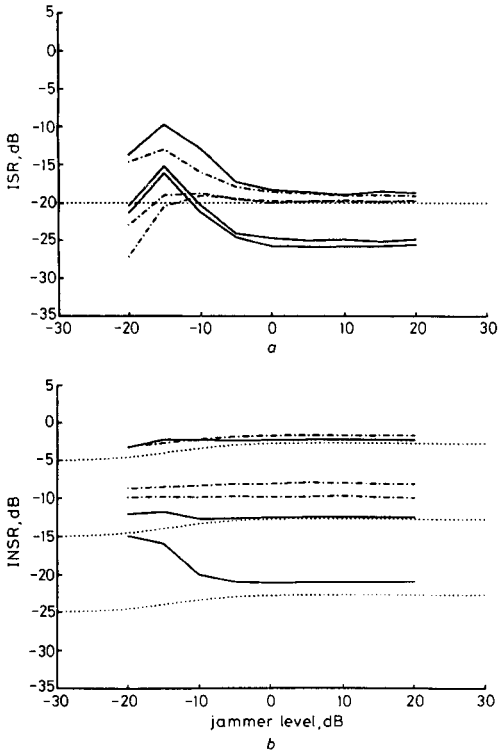

Fig. 4 *ISR_{12} and $INSR_1$ versus jammer level*
$\sigma = -5$ dB; $\phi_1 = -10, 0, 10$ dB; $\phi_2 = 0.05$
a ISR_{12} — BLCMV; -·-·- ILCMV; ····· reference level $1/T$
b $INSR_1$ — BMVDR; -·-·- IMVDR; ····· best theoretical value

Next, we illustrate the benefit of jointly diagonalising several cumulant matrices by comparing the performance of the JADE technique with the simulation results published in Reference 5, where the mixing matrix is estimated by diagonalising an unspecified linear combination of cumulant slices. In this example, matrix A is 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 realisations of $T = 7000$ samples. This Table can be compared with Table 1 of Reference 5. The extra bottom line is the columnwise RMS error (square root of the sum of the squared entries of the given column); it shows an even distribution of errors through

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		
0.0149	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 2: Entrywise and columnwise RMS error in $\hat{A}$ in a 4 × 3 example. JADE algorithm with $T = 700$ samples and binary 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

all columns. For ease of reference, we have computed the corresponding line from Reference 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 columnwise RMS errors) is computed to be 0.0654 through joint diagonalisation and 0.1527 through ordinary diagonalisation of Reference 5. Hence, in this example, joint diagonalisation achieves a better overall performance and it does so with a ten times smaller sample size. Of course, this is only indicative, as a better choice of the single cumulant matrix to be diagonalised may improve the performance reported in Reference 5. It is worth reporting that an exponential distribution does not seem to be very favourable 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.

6 Conclusions

Joint diagonalisation of cumulant matrices allows the whole fourth-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 narrowband beamforming. In this application, directional vectors are blindly estimated rather than modelled 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 beamformers may outperform informed beamformers (whose performance is limited by finite sample size), even when the latter use the true directional vector. This rather surprising fact should theoretically be supported by asymptotic performance analysis of the JADE estimator.

7 References

1 VAN VEEN, B.D., and BUCKLEY, K.M.: 'Beamforming: A versatile approach to spatial filtering', *IEEE ASSP Mag.*, Apr. 1988, pp. 4–24.
2 TONG, L., LIU, R., SOON, V., and HUANG, Y.: 'Indeterminacy and identifiability of blind identification', *IEEE Trans.*, 1991, **CS-38**, pp. 499–509
3 CARDOSO, J.-F.: 'Source separation using higher order moments'. Proc. ICASSP, 1989, pp. 2109–2112
4 SOON, V.C., TONG, L., HUANG, Y.F., and LIU, R.: 'An extended fourth order blind identification algorithm in spatially correlated noise'. Proc. ICASSP, 1990, pp. 1365–1368
5 TONG, L., INOUYE, Y., and LIU, R.: 'Waveform preserving blind estimation of multiple independent sources', *IEEE Trans.*, 1993, **SP-41**, pp. 2461–2470
6 COMON, P.: 'Independent component analysis'. Proc. Int. Workshop on Higher-order stat., Chamrousse, France, 1991, pp. 111–120
7 GAETA, M., and LACOUME, J.-L.: 'Source separation without a priori knowledge: the maximum likelihood solution'. Proc. EUSIPCO, 1990, pp. 621–624
8 DONOHO, D.: 'On minimum entropy deconvolution', in 'Applied time-series analysis II' (Academic Press, 1981), pp. 565–609
9 PORAT, B., and FRIEDLANDER, B.: 'Direction finding algorithms based on high-order statistics', *IEEE Trans.*, 1991, **SP-39**, pp. 2016–2025
10 GIANNAKIS, G., and SHAMSUNDER, S.: 'Modelling of non-Gaussian array data using cumulants: DOA estimation of more sources with less sensors', *Signal Processing*, 1993, **30**, pp. 279–297
11 CARDOSO, J.-F., and COMON, P.: 'Tensor based independent component analysis'. Proc. EUSIPCO, 1990
12 GIANNAKIS, G., INOUYE, Y., and MENDEL, J.: 'Cumulant-based identification of multichannel moving average models', *IEEE Trans.*, 1989, **AC-34**, pp. 783–787
13 TONG, L., and LIU, R.: 'A new eigenstructure based parameter estimation of multichannel moving average processes'. Proc. ICASSP, 1992, **5**, p. 249–252
14 LOUBATON, P., and REGALIA, P.A.: 'Blind deconvolution of multivariate signals: a deflation approach'. Proc. ICC, 1993, **3**, pp. 1160–1165
15 SHALVI, O., and WEINSTEIN, E.: 'New criteria for blind deconvolution of nonminimum phase systems (channels)', *IEEE Trans.*, 1990, **IT-36**, pp. 312–321
16 GODARD, D.: 'Self recovering equalization and carrier tracking in two-dimensional data communication systems', *IEEE Trans.*, 1980, **COMM-28**, pp. 1867–1875
17 MAYRARGUE, S.: 'Spatial equalization of a radio-mobile channel without beamforming using the constant modulus algorithm (CMA)'. Proc. ICASSP, 1993
18 CARDOSO, J.-F.: 'Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors'. Proc. ICASSP, 1991, pp. 3109–3112
19 CARDOSO, J.-F.: 'Iterative techniques for blind source separation using only fourth order cumulants'. Proc. EUSIPCO, 1992, pp. 739–742
20 CARDOSO, J.-F.: 'Fourth-order cumulant structure forcing. Application to blind array processing'. Proc. 6th SSAP Workshop on Statistical signal & array processing, Oct. 1992, pp. 136–139
21 CARDOSO, J.-F., and LAHELD, B.: 'Adaptive blind source separation for channel spatial equalization'. Proc. COST 229 Workshop on Adaptive signal processing, 1992, pp. 19–26
22 MOREAU, E., and MACCHI, O.: 'New self-adaptive algorithms for source separation based on contrast functions'. Proc. IEEE SP Workshop on Higher-order stat., Lake Tahoe, USA, 1993, pp. 215–219
23 JUTTEN, C., and HÉRAULT, J.: 'Une solution neuromimétique au problème de séparation de sources', *Traitement du Signal*, 1988, **5**, pp. 389–403
24 FÉTY, L., and VAN UFFELEN, J.P.: 'New methods for signal separation'. Proc. IEE 4th Int. Conf. on HF radio systems and techniques, London, Apr. 1988, pp. 226–230
25 DINÇ, A., and BAR-NESS, Y.: 'Bootstrap: A fast blind adaptive signal separator'. Proc. ICASSP, 1992, **2**, pp. 325–328
26 PHAM, D.T., GARRAT, P., and JUTTEN, C.: 'Separation of a mixture of independent sources through a maximum likelihood approach'. Proc. EUSIPCO, 1992, pp. 771–774
27 TONG, L., and LIU, R.: 'Blind estimation of correlated source signals'. Proc. Asilomar Conf., Nov. 1990
28 BELOUCHRANI, A., ABED MERAIM, K., CARDOSO, J.-F., and MOULINES, E.: 'Second-order blind separation of correlated sources'. Proc. Int. Conf. on DSP, Cyprus, 1993, pp. 346–351
29 GOLUB, G.H., and LOAN, C.F.V.: 'Matrix computations' (The Johns Hopkins University Press, 1989)

8 Appendix

8.1 Joint diagonalisation algorithm

The Jacobi technique [29] for diagonalising a unique hermitian matrix is extended for the joint approximate diagonalisation of a set $\mathcal{N} = \{N_r | 1 \leqslant r \leqslant s\}$ of arbitrary $n \times n$ matrices. It consists in minimising the diagonalisation criterion (eqn. 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 $V^H N_r V$, optimisation of eqn. 19 amounts to finding θ and ϕ such that $\sum_r |a'_r|^2 + |d'_r|^2$ is maximised. Note that $2(|a'_r|^2 + |d'_r|^2) = |a'_r - d'_r|^2 + |a'_r + d'_r|^2$ and that the trace $a'_r + d'_r$ is invariant in a unitary transformation; maximisation of the criterion of eqn. 19 is equivalent to maximisation of Q:

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

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 $r = 1, \ldots, s$. Then by defining the vectors

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

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

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

the s eqns. 28 may be written in the form $u = Gv$ where $G^T \overset{\text{def}}{=} [g_1, \ldots, g_s]$, so that Q also reads

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

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

8.2 Proofs

To prove propositions 1 and 4, we establish a more general lemma.

Lemma: For any set $\{B_r | 1 \leqslant r \leqslant n^2\}$ of orthonormal $n \times n$ matrices, the identity $c(V) = C(V, \{Q_z(B_r) | 1 \leqslant r \leqslant n^2\})$ holds for any unitary matrix $V = [v_1, \ldots, v_n]$.

Proof: By the following chain of identities:

$$C(V, \{Q_z(B_r) | 1 \leqslant r \leqslant n^2\}) \overset{1}{=} \sum_{i,r} |v_i^* Q_z(B_r) v_i|^2 \overset{2}{=} \sum_{i,r} |\text{Trace } (B_r Q_z(v_i v_i^*))|^2 \overset{3}{=} \sum_i \|Q_z(v_i v_i^*)\|^2_{\text{Fro}} \overset{4}{=} \sum_{i,k,l} |v_k^* Q_z(v_i v_i^*) v_l|^2 \overset{5}{=} \sum_{i,k,l} |\text{Cum } (e_k, e_l^*, e_i, e_i^*)|^2 \overset{6}{=} c(V)$$

Equality 1 is a rewriting of the joint diagonalisation criterion. Equality 2 is an instance of the identity $v^* Q_z(B) v = \text{Trace } (B Q_z(vv^*))$ resulting, for any matrix B and vector v, from the definition of eqn. 11. The matrix sets $\{B_r^H | 1 \leqslant r \leqslant n^2\}$ and $\{v_k v_l^* | 1 \leqslant k, l \leqslant n\}$ are two orthonormal bases for the space of $n \times n$ matrices; expressing the Frobenius norm of $Q_z(v_i v_i^*)$ onto each of these two sets yields equalities 3 and 4, respectively. Equality 5 comes from the multilinearity of the cumulants using $e_i = v_i^* z$. Finally, the last equality 6 uses the cumulant symmetries.

Proof of proposition 1: Using the lemma, as $\{(b_l b_k^*) | 1 \leqslant k, l \leqslant n\}$ 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 $\mathcal{N} = \{M_r | M_r = V \Lambda_r V^H, 1 \leqslant r \leqslant s\}$ where each Λ_r is diagonal and V is unitary, and if $\tilde{V}$ is a joint diagonaliser 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 diagonaliser of $\mathcal{N}^p$, each matrix $\tilde{V}^H Q_z(b_l b_k^*) \tilde{V}$ is diagonal. By linearity of Q_z, this is also true for any linear combination of the matrices $b_l b_k^*$, 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 is holds only $\tilde{U}$ is essentially equal to U.

Assume first that all the sources are kurtic and set $N = \sum_{p=1,n} p k_p^{-1} u_p u_p^*$. The eigenvalues of $Q_z(N)$, which are $1, \ldots, n$ according to eqn. 13, are distinct indeed, so that the unitary diagonaliser of $Q_z(N)$ is essentially unique and then essentially equal to U. Secondly, 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 except for one which is zero. These are distinct numbers, and the conclusion still holds.

Proof of proposition 3: In text.

Proof of proposition 4: Apply the lemma with the orthonormal basis $\{M_r | 1 \leqslant r \leqslant n^2\}$ made of the eigenmatrices of Q_z. These verify $Q_z(M_r) = \lambda_r M_r$ for $1 \leqslant r \leqslant n^2$, but, under H_{0-3}, there are at most n nonzero eigenvalues. Discarding $n(n-1)$ matrices, $\lambda_r M_r$ with $\lambda_r = 0$ does not affect the criterion of eqn. 19 and yields the eigen-set.