Fourth-Order Cumulant-Based Blind Identification of Underdetermined Mixtures

Lieven De Lathauwer, *Senior Member, IEEE*, Joséphine Castaing, and Jean-François Cardoso

***Abstract*—In this paper we study two fourth-order cumulant-based techniques for the estimation of the mixing matrix in underdetermined independent component analysis. The first method is based on a simultaneous matrix diagonalization. The second is based on a simultaneous off-diagonalization. The number of sources that can be allowed is roughly quadratic in the number of observations. For both methods, explicit expressions for the maximum number of sources are given. Simulations illustrate the performance of the techniques.**

***Index Terms*—Cumulant, higher order statistics, higher order tensor, independent component analysis (ICA), parallel factor analysis, simultaneous diagonalization, underdetermined mixture.**

I. Introduction

CONSIDER the following basic linear mixture model:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}. \tag{1}$$

The stochastic vector $\mathbf{x} \in \mathbb{C}^J(\mathbb{R}^J)$ represents multichannel observations, the components of the stochastic vector $\mathbf{s} \in \mathbb{C}^R(\mathbb{R}^R)$ correspond to unobserved source signals, and $\mathbf{n} \in \mathbb{C}^J(\mathbb{R}^J)$ denotes additive noise. The *a priori* unknown mixing matrix $\mathbf{A} \in \mathbb{C}^{J\times R}(\mathbb{R}^{J\times R})$ characterizes the way the sources are combined in the observations. The goal of independent component analysis (ICA) [13], [31], or blind source separation (BSS), consists of the estimation of the source signals and/or the mixing matrix from observations of $\mathbf{x}$, assuming that the sources are statistically independent. The literature on ICA addresses for the most part the so-called *overdetermined* case, where $J \geqslant R$. Here, we consider the *underdetermined* or *overcomplete* case, where $J < R$.

A large class of algorithms for underdetermined ICA starts from the assumption that the sources are (quite) sparse [5], [26], [29], [33], [39]. In this case, the scatter plot typically shows high signal values in the directions of the mixing vectors. These extrema may be localized by maximization of some clustering measure [5], [26], [39]. Some of the clustering-based techniques perform an exhaustive search in the mixing vector space, and are therefore very expensive when there are more than two observation channels. In a preprocessing step a linear transform may be applied such that the new representation of the data is sparser (e.g., short-time Fourier transform in the case of audio signals) [5]. The method in [1] only requires that for each source one area in the time-frequency plane can be found where only that particular source is active; the signals may overlap anywhere else. In [24] the difference between long-time stationary sources and sources that are only short-time stationary is exploited to separate the latter.

There are two aspects to ICA: estimation of the mixing matrix and separation of the sources. In the overdetermined case, sources are usually separated by multiplying the observations with the pseudo-inverse of the mixing matrix estimate. This is no longer possible in the case of underdetermined mixtures: for each sample $\mathbf{x}_t$, the corresponding source sample $\mathbf{s}_t$ that satisfies $\mathbf{x}_t = \mathbf{A}\mathbf{s}_t$ is only known to belong to an affine variety of dimension $J - R$—hence the term "underdetermined." However, the mixing matrix and the source densities are still unique under mildly restrictive conditions [27]. Uniqueness of the source distributions allows for the joint estimation of sources and mixing matrix in a probabilistic framework [34]. However, even in the case of underdetermined mixtures, the estimation of the mixing matrix is an overdetermined problem (e.g., see Sections II and III), such that it makes sense to estimate the mixing matrix first, and then estimate the sources. The source values $\mathbf{s}_t$ may subsequently be estimated by maximizing the log posterior likelihood [34]. In the case of sparse sources, characterized by Laplacian densities, this can be formulated in terms of a linear programming problem [5], [10], [33]. If at most $J - 1$ sources can be active at the same time, then for each sample the active mixing vectors may be determined and the corresponding mixture inverted [29]. In the case of finite alphabet signals in telecommunication, one may perform an exhaustive search over all possible combinations. In this paper we focus on the estimation of the mixing matrix. The estimate of the mixing matrix may subsequently be used to separate the sources by means of the techniques mentioned earlier.

This paper presents new contributions to the class of algebraic algorithms for underdetermined ICA. In [15], [18], and [19] algorithms are derived for the specific case of two mixtures and three sources. An arbitrary number of mixing vectors can be estimated from two observation channels by sampling derivatives of sufficiently high order of the second characteristic function [38]. A more stable version of [38] is presented

Manuscript received March 9, 2006; revised September 22, 2006. This work was supported in part by the Research Council K.U. Leuven under Grant GOA-AMBioRICS, CoE EF/05/006 Optimization in Engineering, in part by the Flemish Government under F.W.O. Project G.0321.06, Tournesol 2005—Project T20013, and F.W.O. research communities ICCoS, ANMMM, and in part by the Belgian Federal Science Policy Office under IUAP P5/22. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tulay Adali.

L. De Lathauwer is with the Research Group ETIS, UMR 8051, F 95014 Cergy-Pontoise Cedex, France and also with the Research Group ESAT-SCD, Katholieke Universiteit Leuven, Leuven, Belgium (e-mail: delathau@ensea.fr).

J. Castaing is with the Research Group ETIS, UMR 8051, F 95014 Cergy-Pontoise Cedex, France (e-mail: castaing@ensea.fr).

J.-F. Cardoso is with the TSI Department, École Nationale Supérieure des Télécommunications, 75634 Paris Cedex 13, France (e-mail: cardoso@tsi.enst.fr).

Digital Object Identifier 10.1109/TSP.2007.893943

1053-587X/$25.00 © 2007 IEEE

in [17]. Algebraic underdetermined ICA is based on the decomposition of a higher order tensor in a sum of rank-1 terms. Some links with the literature on homogeneous polynomials are discussed in [14]. A simultaneous matrix diagonalization technique that may still be used when the number of sources exceeds the number of sensors is presented in [42]. In [3] the algebraic structure of the sixth-order cumulant tensor is partially exploited. A similar idea can be applied to a set of fourth-order cumulant tensors, corresponding to different time lags, when the individual source signals are dependent over some time interval [28]. In this paper we merely assume that the sources have nonzero kurtosis. For convenience, we also assume that the noise is Gaussian. Non-Gaussian noise leads to a perturbation of the equations. This is admissible as long as the perturbation stays relatively small, i.e., the signal-to-noise ratio (SNR) has to be sufficiently high.

The paper is organized as follows. A first fourth-order cumulant-based approach is discussed in Section II. The resulting algorithm is based on a simultaneous matrix diagonalization. A second approach, leading to a simultaneous off-diagonalization, is discussed in Section III. Simulation results are presented in Section IV. Section V is the conclusion. The proofs of the theorems are given in the Appendix. The presentation is in terms of complex signals. Whenever the results cannot be directly applied to real data, this is explicitly mentioned.

The foundations of Section II were laid in [6]. Some mathematical aspects are developed in more detail in [21]. In [22], [23] a variant of the technique is presented that generalizes simultaneous matrix diagonalization-based methods (involving a set of correlation matrices, for instance) to the underdetermined case.

Notation: Scalars are denoted by lowercase italic letters $(a, b, \ldots)$, vectors by lowercase boldface letters $(\mathbf{a}, \mathbf{b}, \ldots)$, matrices by boldface capitals $(\mathbf{A}, \mathbf{B}, \ldots)$, and tensors by calligraphic letters $(\mathcal{A}, \mathcal{B}, \ldots)$. Italic capitals are used to denote index upper bounds $(i = 1, 2, \ldots, I)$. The entry with row index i and column index j in a matrix $\mathbf{A}$, i.e., $(\mathbf{A})_{ij}$, is symbolized by a_{ij}. Likewise, we have $(\mathbf{A})_{i_1 i_2 \ldots i_N} = a_{i_1 i_2 \ldots i_N}$. The columns of $\mathbf{A}$ are denoted by $\mathbf{a}_1, \mathbf{a}_2, \ldots$ We will frequently use $(IJ \times KL)$ matrix representations of $(I \times J \times K \times L)$ tensors. To this end, we define

$$\mathbf{T} = \mathrm{mat}(\mathcal{T}) \Leftrightarrow (\mathbf{T})_{(i-1)J+j,(k-1)L+l} = (\mathcal{T})_{ijkl}.$$

Analogously, $(I \times J)$ matrices will often be stacked in IJ-dimensional vectors

$$\mathbf{m} = \mathrm{vec}(\mathbf{M}) \Leftrightarrow (\mathbf{m})_{(i-1)J+j} = (\mathbf{M})_{ij}.$$

The inverse of the latter operation is denoted by $\mathbf{M} = \mathrm{unvec}(\mathbf{m})$. Vectorization of an $(I \times J \times K \times L)$ tensor is done as follows:

$$\begin{aligned} \mathbf{m} &= \mathrm{vec}(\mathcal{M}) \\ &\Leftrightarrow (\mathbf{m})_{(i-1)JKL+(j-1)KL+(k-1)L+l} \\ &= (\mathcal{M})_{ijkl}. \end{aligned}$$

The symbol δ_{ij} stands for the Kronecker delta, i.e., $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. $\mathbb{H}^{J \times J}$ denotes the space of $(J \times J)$ Hermitean matrices. Finally, we recall the definition of the Kronecker product $\otimes$ and the Khatri–Rao product $\odot$ [35]

$$\mathbf{A} \otimes \mathbf{H} \stackrel{\mathrm{def}}{=} \begin{pmatrix} a_{11}\mathbf{H} & a_{12}\mathbf{H} & \cdots \\ a_{21}\mathbf{H} & a_{22}\mathbf{H} & \cdots \\ \vdots & \vdots & \end{pmatrix}$$

$$\mathbf{A} \odot \mathbf{H} \stackrel{\mathrm{def}}{=} (\mathbf{a}_1 \otimes \mathbf{h}_1 \quad \mathbf{a}_2 \otimes \mathbf{h}_2 \quad \cdots).$$

II. FOOBI Algorithm

Consider the quadricovariance $\mathcal{C}^{\mathbf{x}} = \mathrm{Cum}\{\mathbf{x}, \mathbf{x}^*, \mathbf{x}^*, \mathbf{x}\} \in \mathbb{C}^{J \times J \times J \times J}$. Due to the multilinearity property of cumulant tensors, we have

$$c^{\mathbf{x}}_{ijkl} = \sum_{r=1}^{R} \kappa_r a_{ir} a^*_{jr} a^*_{kr} a_{lr} \tag{2}$$

in which κ_r is the kurtosis of the rth source. This is a decomposition of a symmetric fourth-order tensor in a sum of symmetric rank-1 terms, cf. [9], [14], [16], [20], [21], [30], and the references therein. The minimal number of rank-1 tensors in which a higher order tensor can be decomposed is called its rank. In terms of $\mathbf{C}^{\mathbf{x}} = \mathrm{mat}(\mathcal{C}^{\mathbf{x}}) \in \mathbb{C}^{J^2 \times J^2}$ and $\tilde{\mathbf{C}}^{\mathbf{s}} = \mathrm{diag}(\kappa_1, \ldots, \kappa_R) \in \mathbb{C}^{R \times R}$, (2) can be written as

$$\mathbf{C}^{\mathbf{x}} = (\mathbf{A} \odot \mathbf{A}^*) \cdot \tilde{\mathbf{C}}^{\mathbf{s}} \cdot (\mathbf{A} \odot \mathbf{A}^*)^H. \tag{3}$$

Note that each term in (2) consists of the contribution of one distinct source to $\mathcal{C}^{\mathbf{x}}$. Hence, in terms of the quadricovariance, "mixture identification" amounts to the computation of decomposition (2)–(3). We will work via a second decomposition, which is introduced in the following theorem.

Theorem 1: A tensor $\mathcal{T} \in \mathbb{C}^{J \times J \times J \times J}$, satisfying the symmetries $t_{klij} = t^*_{ijkl}$ and $t_{jilk} = t^*_{ijkl}$, can be eigendecomposed as

$$t_{ijkl} = \sum_{r}^{R} \lambda_r (\mathbf{E}_r)_{ij} (\mathbf{E}_r)^*_{kl} \tag{4}$$

in which the matrices $\mathbf{E}_r$ are Hermitean and mutually orthonormal w.r.t. the Euclidean inner product, and in which the values λ_r are real and nonzero. R is the rank of $\mathbf{T} = \mathrm{mat}(\mathcal{T}) \in \mathbb{C}^{J^2 \times J^2}$. Denote $\mathbf{e}_r = \mathrm{vec}(\mathbf{E}_r), r = 1, \ldots, R, \mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_R) \in \mathbb{C}^{J^2 \times R}$ and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_R) \in \mathbb{C}^{R \times R}$. Then (4) is equivalent to

$$\mathbf{T} = \mathbf{E} \cdot \mathbf{\Lambda} \cdot \mathbf{E}^H \tag{5}$$

in which $\mathbf{E}$ is columnwise orthonormal, with $e_{(i-1)J+j,r} = e^*_{(j-1)J+i,r}, i, j = 1, \ldots, J$, and in which the values λ_r are real and nonzero.

From Theorem 1 we have

$$(\mathcal{C}^{\mathbf{x}})_{ijkl} = \sum_{r}^{R} \lambda_r (\mathbf{E}_r)_{ij} (\mathbf{E}_r)^*_{kl} \tag{6}$$

$$\mathbf{C}^{\mathbf{x}} = \mathbf{E} \cdot \mathbf{\Lambda} \cdot \mathbf{E}^H. \tag{7}$$

The latter equation is essentially a matrix eigenvalue decomposition (EVD), which may easily be computed. The eigenvectors

have to be normalized in order to make the matrices $\mathbf{E}_r$ Hermitean; in the Appendix it is explained how this can be done. Note that, if $\mathbf{A} \odot \mathbf{A}^*$ is full column rank and if $R \leqslant J^2$ (complex case) or $R \leqslant J(J+1)/2$ (real case), the number of sources is given by the rank of $\mathbf{C}^{\mathbf{x}}$. (We notice that in array processing applications, properties of the array may cause $\mathbf{A} \odot \mathbf{A}^*$ to be column rank deficient—we refer to Remark 2.) We will now show how (2)–(3) and (6)–(7) are linked. Assume at this point that all sources are super-Gaussian, i.e., $\kappa_r > 0, r = 1, \ldots, R$. The more general situation will be addressed in Remark 1. From (3) it is clear that $\mathbf{C}^{\mathbf{x}}$ is positive (semi)definite. We have the following theorem.

Theorem 2: Let $\mathbf{C}^{\mathbf{x}}$ be positive (semi)definite and assume that it can be decomposed as in (3) and (7). Then we have

$$(\mathbf{A} \odot \mathbf{A}^*) \cdot (\tilde{\mathbf{C}}^{\mathbf{s}})^{\frac{1}{2}} = \mathbf{E} \cdot \mathbf{\Lambda}^{\frac{1}{2}} \cdot \mathbf{Q} \tag{8}$$

in which $\mathbf{Q}$ is real $(R \times R)$ orthogonal.

After the computation of $\mathbf{E}$ and $\mathbf{\Lambda}^{(1/2)}$ from (7), the next step is the computation of $\mathbf{Q}$. Equation (8) shows that multiplication of $\mathbf{E} \cdot \mathbf{\Lambda}^{(1/2)}$ by $\mathbf{Q}$ yields a matrix of which the columns are Kronecker products. The Kronecker structure of $\mathbf{a}_r \otimes \mathbf{a}_r^*$ corresponds to the rank-1 structure of $\mathrm{unvec}(\mathbf{a}_r \otimes \mathbf{a}_r^*) = \mathbf{a}_r \mathbf{a}_r^H$. This may be exploited. What we need is a tool that allows us to distinguish between Hermitean matrices that are at most rank-1 and Hermitean matrices of which the rank is greater than one. Such a "rank-1 detecting device" is introduced in the following theorem.

Theorem 3: Consider the mapping $\Phi : (\mathbf{X}, \mathbf{Y}) \in \mathbb{H}^{J \times J} \times \mathbb{H}^{J \times J} \rightarrow \Phi(\mathbf{X}, \mathbf{Y}) \in \mathbb{C}^{J \times J \times J \times J}$ defined by

$$(\Phi(\mathbf{X}, \mathbf{Y}))_{ijkl} = x_{ij} y_{kl}^* + y_{ij} x_{kl}^* - x_{ik} y_{jl}^* - y_{ik} x_{jl}^*. \tag{9}$$

Then we have that $\Phi(\mathbf{X}, \mathbf{X}) = 0$ if and only if $\mathbf{X}$ is at most rank-1.

Define matrix $\mathbf{H} = \mathbf{E} \cdot \mathbf{\Lambda}^{(1/2)}$, Hermitean matrices $\mathbf{H}_s = \mathrm{unvec}(\mathbf{h}_s)$ and fourth-order tensors $\mathcal{P}_{st} = \Phi(\mathbf{H}_s, \mathbf{H}_t)$. Now, let $\mathbf{D}$ be any diagonal matrix and let $\mathbf{W} = \mathbf{Q} \cdot \mathbf{D} \cdot \mathbf{Q}^T$. Then, using the bilinearity of Φ, its rank-one detecting feature, and (8), it is readily found that $\sum_{s,t=1}^{R} (\mathbf{W})_{st} \mathcal{P}_{st} = 0$. This suggests to determine a matrix $\mathbf{W}$ from the latter equation, and find $\mathbf{Q}$ as its eigenmatrix. More specifically, we have the following theorem.

Theorem 4: Assume that the tensors $\Phi(\mathbf{a}_u \mathbf{a}_u^H, \mathbf{a}_v \mathbf{a}_v^H), 1 \leqslant u < v \leqslant R$, are linearly independent. Then there exist precisely R linearly independent real symmetric matrices $\mathbf{W}_r \in \mathbb{R}^{R \times R}$ that satisfy

$$\sum_{s,t=1}^{R} (\mathbf{W}_r)_{st} \mathcal{P}_{st} = 0. \tag{10}$$

These matrices have $\mathbf{Q}$ as a common eigenmatrix, i.e.,

$$\begin{aligned} \mathbf{W}_1 &= \mathbf{Q} \cdot \mathbf{D}_1 \cdot \mathbf{Q}^T \\ &\vdots \\ \mathbf{W}_R &= \mathbf{Q} \cdot \mathbf{D}_R \cdot \mathbf{Q}^T \end{aligned} \tag{11}$$

in which $\mathbf{D}_1, \ldots, \mathbf{D}_R \in \mathbb{R}^{R \times R}$ are diagonal.

We can now proceed as follows. Given $\mathcal{P}_{st}, 1 \leqslant s, t \leqslant R, R$ linearly independent matrices $\mathbf{W}_r$ are computed from (10), which is just a homogeneous set of linear equations. Then the matrix $\mathbf{Q}$ follows from the simultaneous EVD (11).

TABLE I
FOOBI ALGORITHM

1) Estimate the quadricovariance $\mathcal{C}^{\mathbf{x}}$ of the data. $\mathbf{C}^{\mathbf{x}} = \mathrm{mat}(\mathcal{C}^{\mathbf{x}})$.
2) Compute the EVD $\mathbf{C}^{\mathbf{x}} = \mathbf{E} \cdot \mathbf{\Lambda} \cdot \mathbf{E}^H$. $\mathbf{H} = \mathbf{E} \cdot \mathbf{\Lambda}^{\frac{1}{2}}$. The number of sources R equals $\mathrm{rank}(\mathbf{C}^{\mathbf{x}})$. Normalize the eigenvectors such that $\mathbf{H}_r = \mathrm{unvec}(\mathbf{h}_r)$, $1 \leqslant r \leqslant R$, is Hermitean.
3) Compute $\mathcal{P}_{st} = \Phi(\mathbf{H}_s, \mathbf{H}_t)$, $1 \leqslant s \leqslant t \leqslant R$. Stack the results in $\mathbf{P}$.
4) Compute the R right singular vectors $\mathbf{u}_r$ of $\mathbf{P}$ that correspond to the smallest singular values. Stack these vectors in upper triangular matrices $\mathbf{U}_r$. Compute $\mathbf{W}_r = \frac{1}{2}(\mathbf{U}_r + \mathbf{U}_r^T)$.
5) Compute orthogonal $\mathbf{Q}$ that best simultaneously diagonalizes the matrices $\mathbf{W}_r$.
6) Compute $\mathbf{F} = \mathbf{E} \cdot \mathbf{\Lambda}^{\frac{1}{2}} \cdot \mathbf{Q}$.
7) Estimate mixing vector $\mathbf{a}_r$ as the dominant left singular vector of $\mathrm{unvec}(\mathbf{f}_r)$, $r = 1, \ldots, R$.

In practice, we work with noisy cumulant estimates, such that (10) will only approximately be satisfied. The matrices $\mathbf{W}_r$ are then determined as follows. Due to the symmetry of $\mathbf{W}_r$, and the fact that $\mathcal{P}_{st} = \mathcal{P}_{ts}, s, t = 1, \ldots, R$, (10) can be written as

$$\sum_{s=1}^{R} (\mathbf{W}_r)_{ss} \mathcal{P}_{ss} + 2 \sum_{\substack{s,t=1 \\ s<t}}^{R} (\mathbf{W}_r)_{st} \mathcal{P}_{st} = 0. \tag{12}$$

In the usual form of a set of homogeneous linear equations, we have

$$\begin{aligned} \mathbf{P} \cdot ((\mathbf{W}_r)_{1,1}, \ldots, (\mathbf{W}_r)_{R,R}, \\ 2(\mathbf{W}_r)_{1,2}, 2(\mathbf{W}_r)_{1,3}, \ldots, 2(\mathbf{W}_r)_{R-1,R})^T = 0 \end{aligned} \tag{13}$$

in which the coefficient matrix $\mathbf{P} \in \mathbb{C}^{J^4 \times R(R+1)/2}$ is given by

$$\begin{aligned} \mathbf{P} = (\mathrm{vec}(\mathcal{P}_{11}), \ldots, \mathrm{vec}(\mathcal{P}_{RR}), \\ \mathrm{vec}(\mathcal{P}_{12}), \mathrm{vec}(\mathcal{P}_{13}), \ldots, \mathrm{vec}(\mathcal{P}_{R-1,R})). \end{aligned} \tag{14}$$

The least-squares solution of (13) consists of the R right singular vectors $\mathbf{u}_r$ of $\mathbf{P}$ that correspond to the smallest singular values. After stacking these vectors in upper triangular matrices $\mathbf{U}_r \in \mathbb{R}^{R \times R}$, in the manner suggested by (13), the matrices $\mathbf{W}_r$ are obtained as $\mathbf{W}_r = (1/2)(\mathbf{U}_r + \mathbf{U}_r^T)$. The following theorem guarantees that the vectors $\mathbf{u}_r$ are real, even in the case of noisy cumulant estimates.

Theorem 5: The right singular vectors of the matrix $\mathbf{P}$ in (14) are real.

After computation of the matrices $\mathbf{W}_r$, the common eigenmatrix $\mathbf{Q}$ in (11) can be obtained by means of the Jacobi algorithm developed in [7] and [8]. Multiplication by $\mathbf{E} \cdot \mathbf{\Lambda}^{(1/2)}$, as in (8), yields a matrix $\mathbf{F} \in \mathbb{C}^{J^2 \times R}$ of which the columns $\mathbf{f}_r$ are theoretically proportional to $\mathbf{a}_r \otimes \mathbf{a}_r^*$. In practice, we estimate $\mathbf{a}_r$ from the best rank-1 approximation of $\mathrm{unvec}(\mathbf{f}_r)$. The overall *Fourth-Order-Only Blind Identification (FOOBI)* algorithm is outlined in Table I.

Remark 1: In the derivation above, we have assumed that all sources are super-Gaussian. If all sources are sub-Gaussian, i.e., $\kappa_r < 0, r = 1, \ldots, R$, then we simply process $-\mathcal{C}^{\mathbf{x}}$. In case not all kurtosis values have the same sign, $\mathbf{C}^{\mathbf{x}}$ is indefinite. The

derivation then still applies, with the exception that $\mathbf{Q}$ is J-orthogonal [41] instead of orthogonal. This implies that, for the simultaneous diagonalization (11), a variant of the algorithm in [7] and [8] has to be worked out, that involves J-orthogonal matrices. We can also work as follows. Instead of imposing J-orthogonality, we simply start from

$$\mathbf{A} \odot \mathbf{A}^* = \mathbf{E} \cdot \mathbf{Q} \tag{15}$$

with $\mathbf{Q}$ some real nonsingular matrix. The procedure is essentially the same, but the simultaneous diagonalization in (11) now involves a real nonsingular matrix. For algorithms for this type of simultaneous diagonalization we refer to [20], [40], [42], and the references therein.

The condition on $\Phi(\mathbf{a}_u\mathbf{a}_u^H, \mathbf{a}_v\mathbf{a}_v^H)$ in Theorem 4 yields an upper bound on the number of sources FOOBI can handle. We call a property "generic" when it holds with probability one when the entries of the mixing matrix are sampled from continuous probability density functions. We have the following theorem.

Theorem 6: In the complex case, linear independence of $\Phi(\mathbf{a}_u\mathbf{a}_u^H, \mathbf{a}_v\mathbf{a}_v^H), 1 \leqslant u < v \leqslant R$, is generically guaranteed if $R(R-1) \leqslant J^2(J-1)^2/2$. In the real case, R is bounded as follows:

J	2	3	4	5	6	7	8	9
$R_{\max}$	2	4	6	10	15	20	26	33

Remark 2: In [11], [12], and [25] it is explained that in antenna array applications, the characteristics of the antennas and the geometry of the array may induce a structure in the entries of the higher order cumulant that limits the number of sources that can effectively be dealt with. Such a structure is neglected in Theorem 6. As a result, the number of sources that can be allowed is bounded by the minimum of: 1) the number of sources in Theorem 6 and 2) the maximal number of virtual sensors (VSs), derived in [11], [12], and [25].

III. FOOBI-2 Algorithm

Like in the previous section, we start from the EVD (7). Generically, as long as $R \leqslant J^2$ (complex case) or $R \leqslant J(J+1)/2$ (real case), the number of sources corresponds to the rank of $\mathbf{C}^{\mathbf{x}}$. In this section we assume that all sources are super-Gaussian. (If all sources are sub-Gaussian, then we process $-\mathcal{C}^{\mathbf{x}}$ instead of $\mathcal{C}^{\mathbf{x}}$.) This means that Theorem 2 still applies. We now introduce a new rank-1 detecting device.

Theorem 7: Consider the mapping $\Psi : (\mathbf{X}, \mathbf{Y}) \in \mathbb{H}^{J\times J} \times \mathbb{H}^{J\times J} \rightarrow \Psi(\mathbf{X}, \mathbf{Y}) \in \mathbb{H}^{J\times J}$ defined by

$$\Psi(\mathbf{X}, \mathbf{Y}) = \mathbf{XY} - \operatorname{trace}(\mathbf{X})\mathbf{Y} + \mathbf{YX} - \operatorname{trace}(\mathbf{Y})\mathbf{X}. \tag{16}$$

Then we have that $\Psi(\mathbf{X}, \mathbf{X}) = 0$ if and only if $\mathbf{X}$ is at most rank-1.

Let $\mathbf{H} = \mathbf{E} \cdot \mathbf{\Lambda}^{(1/2)}, \mathbf{H}_s = \operatorname{mat}(\mathbf{h}_s)$ and define symmetric matrices $\mathbf{B}_{ij} \in \mathbb{C}^{R\times R}$ by

$$(\mathbf{B}_{ij})_{st} = (\Psi(\mathbf{H}_s, \mathbf{H}_t))_{ij}, \quad 1 \leqslant i, j \leqslant J, \quad 1 \leqslant s, t \leqslant R.$$

The following theorem suggests a new algorithm for the computation of $\mathbf{Q}$.

TABLE II
FOOBI-2 Algorithm

1) Estimate the quadricovariance $\mathcal{C}^{\mathbf{x}}$ of the data. $\mathbf{C}^{\mathbf{x}} = \operatorname{mat}(\mathcal{C}^{\mathbf{x}})$.
2) Compute the EVD $\mathbf{C}^{\mathbf{x}} = \mathbf{E} \cdot \mathbf{\Lambda} \cdot \mathbf{E}^H$. $\mathbf{H} = \mathbf{E} \cdot \mathbf{\Lambda}^{\frac{1}{2}}$. The number of sources R equals $\operatorname{rank}(\mathbf{C}^{\mathbf{x}})$. Normalize the eigenvectors such that $\mathbf{H}_r = \operatorname{unvec}(\mathbf{h}_r)$, $1 \leqslant r \leqslant R$, is Hermitean.
3) Compute $\Psi(\mathbf{H}_s, \mathbf{H}_t)$, $1 \leqslant s \leqslant t \leqslant R$. Stack the results in $\mathbf{B}_{ij}$, $1 \leqslant i \leqslant j \leqslant J$.
4) Compute orthogonal $\mathbf{Q}$ that best simultaneously off-diagonalizes the matrices $\operatorname{Real}(\mathbf{B}_{ij})$ and $\operatorname{Imag}(\mathbf{B}_{ij})$.
5) Compute $\mathbf{F} = \mathbf{E} \cdot \mathbf{\Lambda}^{\frac{1}{2}} \cdot \mathbf{Q}$.
6) Estimate mixing vector $\mathbf{a}_r$ as the dominant left singular vector of $\operatorname{mat}(\mathbf{f}_r)$, $r = 1, \ldots, R$.

Theorem 8: The matrix $\mathbf{Q}$ in (8) satisfies

$$\begin{aligned} \operatorname{diag}(\mathbf{Q}^T \cdot \operatorname{Real}(\mathbf{B}_{ij}) \cdot \mathbf{Q}) = 0, \quad 1 \leqslant i \leqslant j \leqslant J \\ \operatorname{diag}(\mathbf{Q}^T \cdot \operatorname{Imag}(\mathbf{B}_{ij}) \cdot \mathbf{Q}) = 0, \quad 1 \leqslant i < j \leqslant J. \end{aligned} \tag{17}$$

This theorem shows that the matrix $\mathbf{Q}$ can be computed by means of simultaneous off-diagonalization of $J(J+1)/2$ (real case) or J^2 (complex case) real symmetric matrices. The simultaneous off-diagonalization can be realized by means of a simple variant of the Jacobi algorithm derived in [7], [8]. It suffices to chose in each step the Jacobi rotation that minimizes (instead of maximizes) the sum of the squared diagonal entries. Simultaneous off-diagonalization also appeared in [4]. The FOOBI-2 algorithm is summarized in Table II.

The iteration that forms the core of FOOBI-2 (step 4 in Algorithm II) is computationally more expensive than FOOBI's core iteration (step 5 in Algorithm I), because the simultaneous off-diagonalization involves more matrices. On the other hand, FOOBI requires the computation of part of the SVD of the $(J^2 \times R(R+1)/2)$ matrix $\mathbf{P}$ (step 4 in Algorithm I). Also, FOOBI-2 is less restrictive in terms of the number of sources that can be allowed. It only requires that $R \leqslant J^2$ (complex case) or $R \leqslant J(J+1)/2$ (real case), provided decomposition (2) is unique. Consequently, we investigate under which conditions generic uniqueness holds. (We notice that in nongeneric cases Remark 2 still holds.)

In [16] it is stated that a decomposition in rank-1 terms is generically unique when the number of parameters in the decomposition is strictly smaller than the number of distinct tensor entries. When both numbers are equal, then generically only a finite number of decompositions are possible. In the complex case, the total number of distinct real and distinct imaginary parts of the entries of a generic $(J \times J \times J \times J)$ tensor $\mathcal{T}$ satisfying the symmetries $t_{klij} = t^*_{ijkl}, t_{jilk} = t^*_{ijkl}$, and $t_{ljki} = t_{ijkl}$ is given by

$$\begin{aligned} D(J) = 6\binom{J}{4} + 4\binom{J}{1}\binom{J-1}{2} + 3\binom{J}{2} \\ + 2\binom{J}{1}\binom{J-1}{1} + \binom{J}{1} \end{aligned} \tag{18}$$

where we assume that $\binom{J}{n} = 0$ when $n > J$. On the other hand, the number of distinct real parameters in the decomposition $t_{ijkl} = \sum_{r=1}^{R} \kappa_r a_{ir} a^*_{jr} a^*_{kr} a_{lr}$, with κ_r real, is equal to

$2JR$. The maximal rank for which the decomposition is unique is then given by R_{uc} in the following table:

J	2	3	4	5	6	7	8	9
R_{uc}	2	5	12	22	36	55	80	112
R_{ur}	2	4	8	13	20	29	41	54

Note that R_{uc} can be greater than J^2. In the real case, the number of distinct entries of a generic symmetric $(J\times J\times J\times J)$ tensor is equal to $J(J+1)(J+2)(J+3)/4!$, while the number of parameters in the decomposition equals JR. The maximal rank for which the decomposition is unique is then given by R_{ur} in the table above. Note that R_{ur} can be greater than $J(J+1)/2$.

IV. Simulations

In the first simulation, R narrow-band sources are received by a uniform circular array (UCA) of $J=4$ identical sensors of radius R_a. We assume free-space propagation. This means that the entries of the mixing matrix before normalization are given by

$$\tilde{a}_{jr} = \exp(2\pi\jmath(x_j\cos(\theta_r)\cos(\phi_r) + y_j\cos(\theta_r)\sin(\phi_r)))$$

where $x_j = (R_a/\lambda)\cos(2\pi(j-1)/J), y_j = (R_a/\lambda)\sin(2\pi(j-1)/J)$ and $\jmath = \sqrt{-1}$. We have $R_a/\lambda = 0.55$. The mixing matrix $\mathbf{A}$ is obtained by dividing the columns of $\tilde{\mathbf{A}}$ by their Frobenius norm. We consider two cases: $R=5$ and $R=6$. These values of R are not greater than the number of fourth-order VSs of the UCA [11], [12], [25]. The directions-of-arrival (DOAs) of the sources are given by $\theta_1 = 3\pi/10, \theta_2 = 3\pi/10, \theta_3 = 2\pi/5, \theta_4 = 0, \theta_5 = \pi/10, \theta_6 = 3\pi/5$ and $\phi_1 = 7\pi/10, \phi_2 = 9\pi/10, \phi_3 = 3\pi/5, \phi_4 = 4\pi/5, \phi_5 = 3\pi/5, \phi_6 = \pi/5$. In the case $R=5$, we consider the first five DOAs. The sources are unit-variance QAM4 in baseband, which means that they take their values equally likely in the set $\{\pm 1/\sqrt{2} \pm \jmath/\sqrt{2}\}$. Additive zero-mean complex Gaussian noise is added to the data. The mixing matrix is estimated by means of: 1) the FOOBI algorithm; 2) the FOOBI-2 algorithm; and 3) the BIRTH algorithm [2] (or 6-BIOME1 algorithm, in the terminology of [3]), which uses the sixth-order cumulant of the observations. (We note that the 6-BIOME3 algorithm is somewhat more accurate than 6-BIOME1, at the expense of a higher computation cost [3].) The precision is measured in terms of the mean relative error $E\{\|\mathbf{A}-\hat{\mathbf{A}}\|/\|\mathbf{A}\|\}$, in which the norm is the Frobenius norm and in which $\hat{\mathbf{A}}$ represents the optimally ordered and scaled estimate of $\mathbf{A}$. We conduct Monte Carlo experiments consisting of 100 runs.

Fig. 1 shows the accuracy as a function of the SNR, when 5000 samples are used. The FOOBI and FOOBI-2 curves practically coincide. BIRTH is less accurate. We have also compared to the AC-DC algorithm [42], applied to the dominant R Hermitean eigenmatrices of the fourth-order cumulant. This means that exactly the same statistics as in FOOBI and FOOBI-2 are used. However, AC-DC failed to reliably estimate the mixture.

Fig. 2 shows the accuracy as a function of the number of data samples T, for the case $R=5$. The SNR was taken equal to 16 dB. Again, the FOOBI and FOOBI-2 curves practically coincide and BIRTH is less accurate.

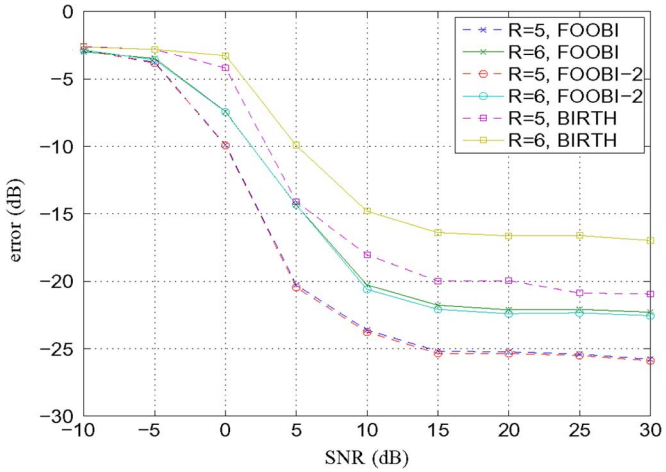

Fig. 1. Accuracy as a function of SNR in the first experiment ($J=4; R=5,6;5000$ samples).

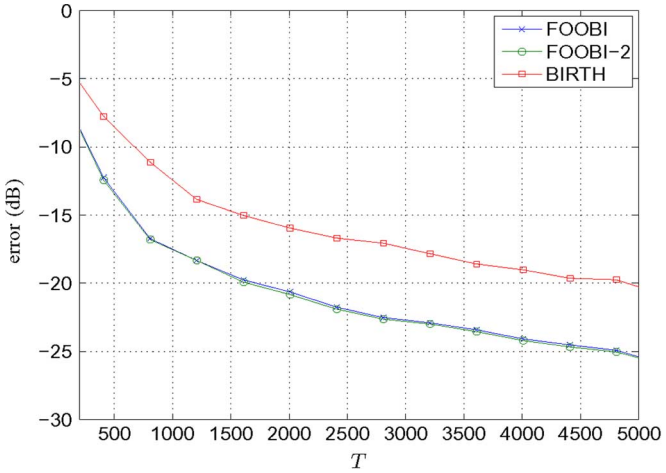

Fig. 2. Accuracy as a function of data length in the first experiment ($J=4; R=5;16$ dB).

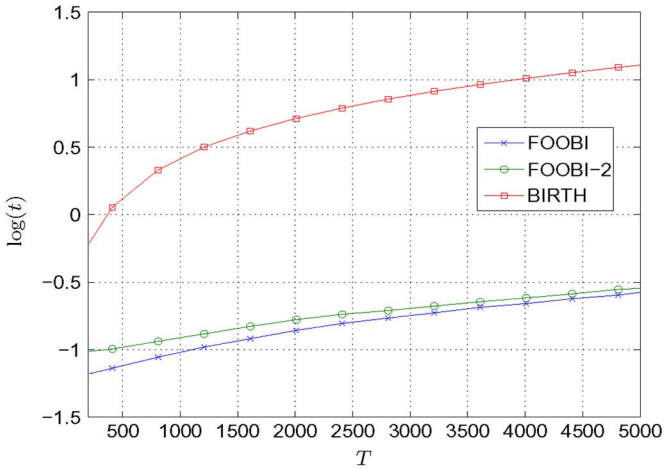

Fig. 3. Computation time as a function of data length in the first experiment ($J=4; R=5;16$ dB).

In Fig. 3 we compare the computational cost of the algorithms. FOOBI and FOOBI-2 are about equally expensive in this simulation. BIRTH is about a factor 40 more expensive than FOOBI and FOOBI-2. The reason is that BIRTH requires the

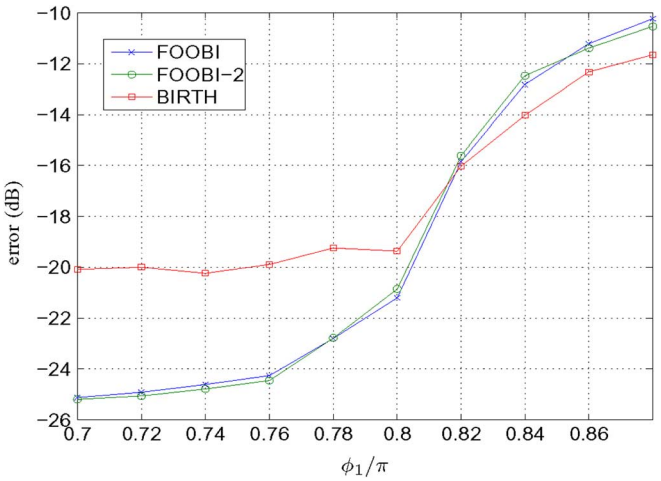

Fig. 4. Accuracy as a function of angle of first mixing vector ($J = 4; R = 5; 16$ dB).

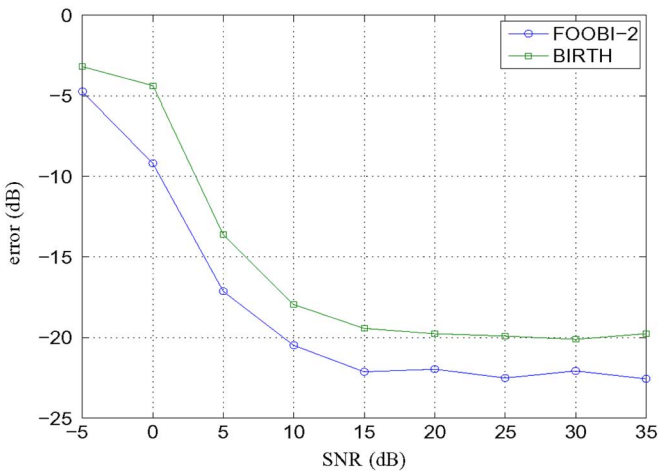

Fig. 5. Accuracy as a function of SNR in the second experiment ($J = 3; R = 5; 5000$ samples).

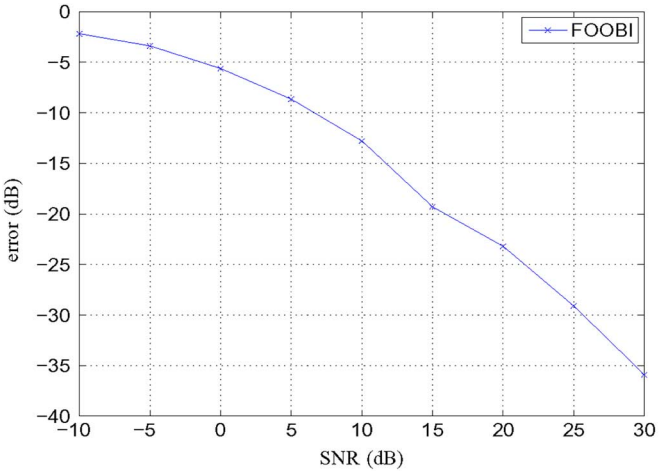

Fig. 6. Accuracy as a function of SNR in the third experiment ($J = 18, R = 25$).

estimation of $O(J^6)$ cumulant entries. The estimation of the sixth-order cumulant accounts for more than 90% of the total computational cost. The estimation of the fourth-order cumulant accounts for about 10% of the total FOOBI or FOOBI-2 cost when 200 samples are used; respectively, 70% when 5000 samples are used. The computation time varies little as a function of the SNR.

Fig. 4 shows the accuracy as a function of the condition of the problem, for the case $R = 5$. The SNR was taken equal to 16 dB, and 5000 samples were used. The figure shows what happens when ϕ_1 is varied from $7\pi/10$ to $0.89\pi/10$ (the latter value is very close to ϕ_2). Again, the FOOBI and FOOBI-2 curves practically coincide. FOOBI and FOOBI-2 are more accurate than BIRTH when the problem is well conditioned. On the other hand, when the two mixing vectors are very close, BIRTH is more reliable than FOOBI and FOOBI-2. The reason is that the vectors $\mathbf{a}_1 \otimes \mathbf{a}_1 \otimes \mathbf{a}_1^*$ and $\mathbf{a}_2 \otimes \mathbf{a}_2 \otimes \mathbf{a}_2^*$ are less close than the vectors $\mathbf{a}_1 \otimes \mathbf{a}_1^*$ and $\mathbf{a}_2 \otimes \mathbf{a}_2^*$, as explained in [12].

In the second simulation, $R = 5$ narrow-band sources are received by a UCA of $J = 3$ identical sensors. The number of sources is less than the number of fourth-order VSs for this array, which is equal to 6 [12]. However, the number of sources is above the FOOBI bound (Theorem 6). Consequently, the mixing matrix is only estimated by means of: 1) the FOOBI-2 algorithm and 2) the BIRTH algorithm. The DOAs of the sources are equal to the first five DOAs in the first experiment. All other parameters are as in the first experiment.

Fig. 5 shows the accuracy as a function of the SNR. FOOBI-2 turns out to be more accurate than BIRTH. Similar curves as in the first experiment have been obtained for the accuracy and the computational cost as a function of the number of samples.

Finally, we show the results of a simulation with entirely synthetic data. In this simulation, there are 18 observation channels and 25 sources. The sources have unit kurtosis. The entries of the mixing matrix are drawn from a zero-mean unit-variance complex Gaussian distribution. The columns are subsequently scaled to unit length. The noise-free cumulant is computed directly from (3). Whenever the condition number of $\mathbf{C}^{\mathbf{x}}$ is greater than 100, a new mixing matrix is generated, so that we do not consider severely ill-conditioned data. Additive zero-mean Gaussian noise is added directly on the cumulant. With the noise term represented by $\mathbf{C}^{\mathbf{n}} \in \mathbb{C}^{18^2 \times 18^2}$, the SNR is defined as $\|\mathbf{C}^{\mathbf{x}}\|/\|\mathbf{C}^{\mathbf{n}}\|$, in which the norms are Frobenius norms. The mixing matrix is estimated by means of the FOOBI algorithm. Fig. 6 shows the results of a Monte Carlo experiment consisting of 100 runs.

V. Conclusion

In this paper we have studied the estimation of the mixing matrix from the observed fourth-order cumulant tensor in underdetermined ICA. As long as the number of sources is less than the number of VSs of the antenna array (if the results of [11], [12], and [25] apply), it can be found as the rank of a matrix representation of the cumulant. Under a specific condition on the mixing vectors, allowing for a number of sources that increases quadratically with the number of observations, the noise-free solution may be found from an EVD. For noisy data we proposed the FOOBI algorithm, which computes the solution by means of a simultaneous matrix diagonalization. A second algorithm, called FOOBI-2, was based on a simultaneous off-diagonalization. FOOBI-2 is even less restrictive in the number of

sources than FOOBI. The algorithms are based on new results w.r.t. the decomposition of a fourth-order symmetric tensor in a sum of symmetric rank-1 terms. We have determined the maximum number of terms such that this decomposition is generically unique. Throughout the paper, both the real and the complex case have been addressed. The performance of the algorithms has been illustrated by means of simulations.

APPENDIX I
PROOFS

Theorem 1:

Proof: Due to the symmetry $t_{klij} = t^*_{ijkl}$, the matrix $\mathbf{T}$ is Hermitean. Hence, its EVD takes the form of (5), with $\mathbf{E}$ columnwise orthonormal and λ_r real. The tensor $\mathcal{T}$ can thus be decomposed as in (4), with $\mathbf{E}_r$ mutually orthonormal w.r.t. the scalar product of matrices, and λ_r real. Consider the tensor $\mathcal{S}$, defined by $s_{ijkl} = t_{jilk}$, and its matrix representation $\mathbf{S} = \mathrm{mat}(\mathcal{S})$. We have

$$s_{ijkl} = \sum_r^R \lambda_r (\mathbf{E}_r)_{ji} (\mathbf{E}_r)^*_{lk}$$
$$\mathbf{S} = \tilde{\mathbf{E}} \cdot \mathbf{\Lambda} \cdot \tilde{\mathbf{E}}^H \tag{19}$$

in which $\tilde{\mathbf{E}} = (\tilde{\mathbf{e}}_1, \ldots, \tilde{\mathbf{e}}_R)$ is defined by $\tilde{e}_{(j-1)J+i,r} = e_{(i-1)J+j,r}, i, j = 1, \ldots, J, r = 1, \ldots, R$. On the other hand, we have

$$t^*_{ijkl} = \sum_r^R \lambda_r (\mathbf{E}_r)^*_{ij} (\mathbf{E}_r)_{kl}$$
$$\mathbf{T}^* = \mathbf{E}^* \cdot \mathbf{\Lambda} \cdot \mathbf{E}^T. \tag{20}$$

Because of the symmetry $t_{jilk} = t^*_{ijkl}$, we have $\mathcal{S} = \mathcal{T}^*$ and $\mathbf{S} = \mathbf{T}^*$. In the case where all eigenvalues are different, projectors $\tilde{\mathbf{e}}_r \tilde{\mathbf{e}}_r^H$ and $\mathbf{e}^*_r \mathbf{e}^T_r$ corresponding to the same eigenvalue are equal. Hence

$$(\mathbf{E}_r)_{ji} (\mathbf{E}_r)^*_{lk} = (\mathbf{E}_r)^*_{ij} (\mathbf{E}_r)_{kl}. \tag{21}$$

If λ is a multiple eigenvalue, then the corresponding rank-1 projectors can be chosen equal. We conclude that the projectors satisfy the same symmetries as $\mathcal{T}$ itself.

Now we show that this implies that the matrices $\mathbf{E}_r$ can always be normalized to Hermitean matrices. Note that the projector $\mathbf{e}_r \mathbf{e}_r^H$ does not change when $\mathbf{E}_r$ is multiplied by a unit-modulus scalar α_r. Let $\mathbf{E}'_r = \alpha_r \mathbf{E}_r$. If some diagonal entry of $\mathbf{E}_r$, say $(\mathbf{E}_r)_{pp}$, is nonzero, then we choose α_r such that $(\mathbf{E}'_r)_{pp}$ is real. Since we have $(\mathbf{E}'_r)_{ji} (\mathbf{E}'_r)^*_{pp} = (\mathbf{E}'_r)^*_{ij} (\mathbf{E}'_r)_{pp}$ for all $i, j, \mathbf{E}'_r$ is Hermitean. If all the diagonal entries of $\mathbf{E}_r$ are zero, then we proceed as follows. First notice that (21) implies that all $|(\mathbf{E}_r)_{ij}| = |(\mathbf{E}_r)_{ji}|$. If, say, $(\mathbf{E}_r)_{pq} \neq 0$, then we multiply $\mathbf{E}_r$ by α_r chosen such that $(\mathbf{E}'_r)_{pq} = (\mathbf{E}'_r)^*_{qp}$. Since we have $(\mathbf{E}'_r)_{ji} (\mathbf{E}'_r)^*_{qp} = (\mathbf{E}'_r)^*_{ij} (\mathbf{E}'_r)_{pq}, \mathbf{E}'_r$ is Hermitean. ∎

The computation of tensor decomposition (4) amounts to the computation of the classical matrix EVD in (5), in which the eigenvectors are normalized in order to make the matrices $\mathbf{E}_r$ Hermitean, as explained in the proof. We emphasize that these eigenmatrices are not Hermitean by default, as they may be multiplied by any unit-modulus scalar. Multiplication by j even yields anti-Hermitean eigenmatrices. We also mention that, in the complex case, the rank R can be as large as J^2, the maximal rank of $(J^2 \times J^2)$ matrices. The equivalent of Theorem 1 for real-valued tensors is simply obtained by dropping complex conjugations. The proof is trivial. Since the eigenmatrices are real symmetric here, R is bounded by $J(J+1)/2$, the dimension of the vector space of real symmetric matrices.

Theorem 2:

Proof: Both $(\mathbf{A} \odot \mathbf{A}^*) \cdot (\tilde{\mathbf{C}}^{\mathbf{s}})^{(1/2)}$ and $\mathbf{H} \cdot \mathbf{\Lambda}^{(1/2)}$ are square roots of the positive (semi)definite matrix $\mathbf{C}^{\mathbf{x}}$. Hence, they are related as in (8), with $\mathbf{Q}$ $(R \times R)$ unitary. We will now show that $\mathbf{Q}$ is in fact real. Consider the permutation matrix $\mathbf{\Pi} \in \mathbb{R}^{J^2 \times J^2}$, defined by

$$\begin{cases} (\mathbf{\Pi})_{(i-1)J+j,(j-1)J+i} = 1, & 1 \leqslant i, j \leqslant J \\ (\mathbf{\Pi})_{i,j} = 0, & \text{elsewhere.} \end{cases}$$

From the symmetry properties of the columns of $\mathbf{A} \odot \mathbf{A}^*$ and $\mathbf{E}$, we have

$$\begin{aligned} \mathbf{\Pi} \cdot (\mathbf{A} \odot \mathbf{A}^*) \cdot (\tilde{\mathbf{C}}^{\mathbf{s}})^{\frac{1}{2}} &= (\mathbf{A} \odot \mathbf{A}^*)^* \cdot (\tilde{\mathbf{C}}^{\mathbf{s}})^{\frac{1}{2}} \\ &= \mathbf{\Pi} \cdot \mathbf{E} \cdot \mathbf{\Lambda}^{\frac{1}{2}} \cdot \mathbf{Q} = \mathbf{E}^* \cdot \mathbf{\Lambda}^{\frac{1}{2}} \cdot \mathbf{Q}. \end{aligned} \tag{22}$$

Combination of (8) and (22) shows that $\mathbf{Q}$ is real. ∎

Theorem 3:

Proof: The "if" part is obvious. For the "only if" part, we start from

$$x_{ij} x^*_{kl} = x_{ik} x^*_{jl} \quad \forall i, j, k, l$$

which implies

$$\sum_{k,l=1}^J x_{ij} x^*_{kl} \delta_{kl} = \sum_{k,l=1}^J x_{ik} x^*_{jl} \delta_{kl} \quad \forall i, j.$$

The latter equation can be written in matrix terms as

$$\mathbf{X}\, \mathrm{trace}(\mathbf{X}) = \mathbf{X}^2. \tag{23}$$

The case $\mathrm{trace}(\mathbf{X}) = 0$ can be discarded, since it implies that $\mathbf{X}^2 = 0$ and hence $\mathbf{X} = 0$, since $\mathbf{X}$ is Hermitean. Dividing (23) by $\mathrm{trace}(\mathbf{X})$ shows that the unit trace Hermitean matrix $\tilde{\mathbf{X}} = \mathbf{X}/\mathrm{trace}(\mathbf{X})$ satisfies $\tilde{\mathbf{X}}^2 = \tilde{\mathbf{X}}$. Hence, $\tilde{\mathbf{X}}$ is an orthogonal projector. Moreover, since the dimension of the image space of an orthogonal projector is equal to its trace, the rank of $\tilde{\mathbf{X}}$ is equal to one. We conclude that $\mathbf{X}$ is rank-1. The theorem can also be proved in analogy with [21, Th. 2.1]. ∎

Theorem 4:

Proof: We first show that every real symmetric matrix $\mathbf{W}$ that satisfies (10), has $\mathbf{Q}$ as eigenmatrix. Due to the bilinearity of Φ, we have from (8)

$$\mathcal{P}_{st} = \sum_{u,v=1}^R (\mathbf{Q}^T)_{us} (\mathbf{Q}^T)_{vt} \kappa_u^{1/2} \kappa_v^{1/2} \Phi\left(\mathbf{a}_u \mathbf{a}_u^H, \mathbf{a}_v \mathbf{a}_v^H\right). \tag{24}$$

Substitution of (24) in (10) yields

$$\sum_{s,t=1}^R \sum_{u,v=1}^R (\mathbf{Q}^T)_{us} (\mathbf{Q}^T)_{vt} (\mathbf{W})_{st} \kappa_u^{1/2} \kappa_v^{1/2} \Phi\left(\mathbf{a}_u \mathbf{a}_u^H, \mathbf{a}_v \mathbf{a}_v^H\right) = 0.$$

According to Theorem 3, we have $\Phi\left(\mathbf{a}_u\mathbf{a}_u^H, \mathbf{a}_u\mathbf{a}_u^H\right) = 0, 1 \leqslant u \leqslant R$. Additionally taking into account the symmetry of $\mathbf{W}$ and the fact that Φ is symmetric in its arguments, we obtain

$$\sum_{s,t=1}^{R} \sum_{\substack{u,v=1 \\ u<v}}^{R} (\mathbf{Q}^T)_{us}(\mathbf{Q}^T)_{vt}(\mathbf{W})_{st}\kappa_u^{1/2} \times \kappa_v^{1/2}\Phi\left(\mathbf{a}_u\mathbf{a}_u^H, \mathbf{a}_v\mathbf{a}_v^H\right) = 0. \quad (25)$$

If the tensors $\Phi(\mathbf{a}_u\mathbf{a}_u^H, \mathbf{a}_v\mathbf{a}_v^H), 1 \leqslant u < v \leqslant R$, are linearly independent, then the coefficients in (25) have to be zero

$$\sum_{s,t=1}^{R} (\mathbf{Q}^T)_{us}(\mathbf{Q}^T)_{vt}(\mathbf{W})_{st} = d_{uv}\delta_{uv} \quad \forall u, v. \quad (26)$$

This can be written in a matrix format as

$$\mathbf{W} = \mathbf{Q} \cdot \mathbf{D} \cdot \mathbf{Q}^T \quad (27)$$

in which $\mathbf{D}$ is diagonal. Since linear independence of matrices $\mathbf{W}$ amounts to linear independence of diagonal matrices $\mathbf{D}$, at most R real symmetric matrices can satisfy (10). On the other hand, it is easy to verify that any real diagonal matrix $\mathbf{D}$ generates a real symmetric matrix $\mathbf{W}$ that does satisfy (10). This proves the theorem. ■

Theorem 5:

Proof: It suffices to prove that $\mathbf{P}^H\mathbf{P}$ is Hermitean. This can be done by computing its entries and taking into account that the matrices $\mathbf{H}_r$ are Hermitean. ■

Theorem 6:

Proof: The complex case is a technical variant of the proof of [21, Th. 2.5]. The real case is analyzed in [37]. An algorithm is described that allows to compute $R_{\max}$ for any given J. It is conjectured that in the real case the bound is of the form

$$\frac{R(R-1)}{2} \leqslant \frac{J(J-1)}{4}\left(\frac{J(J-1)}{2}+1\right) - \frac{J!}{(J-4)!4!}1_{\{J\geqslant 4\}},$$

where

$$1_{\{J\geqslant 4\}} = \begin{cases} 0, & \text{if } J < 4 \\ 1, & \text{if } J \geqslant 4 \end{cases}$$

■

Theorem 7:

Proof: It is easy to verify that $\Psi(\mathbf{X}, \mathbf{X}) = 0$ if $\mathbf{X}$ is rank-1. For the "only if" part, let the EVD of $\mathbf{X}$ be given by $\mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^H$. We have $\Psi(\mathbf{X}, \mathbf{X}) = 0$ if and only if

$$\begin{aligned} \mathbf{U} \cdot \mathbf{\Lambda}^2 \cdot \mathbf{U}^H &= \operatorname{trace}(\mathbf{\Lambda})\, \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^H \\ \mathbf{\Lambda}^2 &= \operatorname{trace}(\mathbf{\Lambda})\, \mathbf{\Lambda} \\ \mathbf{\Lambda} &= \operatorname{trace}(\mathbf{\Lambda})\, \mathbf{I}. \end{aligned}$$

Hence, at most one eigenvalue can be different from zero. ■

Theorem 8:

Proof: Due to the bilinearity of Ψ, we have

$$\sum_{s,t=1}^{R} (\mathbf{Q})_{sr}(\mathbf{Q})_{tr}\Psi(\mathbf{H}_s, \mathbf{H}_t) = 0 \quad 1 \leqslant r \leqslant R \quad (28)$$

This equation can be written in terms of $\mathbf{B}_{ij}$ as follows:

$$\operatorname{diag}(\mathbf{Q}^T \cdot \mathbf{B}_{ij} \cdot \mathbf{Q}) = 0 \quad 1 \leqslant i \leqslant j \leqslant J$$

This equation is equivalent with (17) because of the link between $\mathbf{B}_{ij}$ and $\mathbf{B}_{ji}$. ■

REFERENCES

[1] F. Abrard and Y. Deville, "A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources," *Signal Process.*, vol. 85, no. 7, pp. 1389–1403, Jul. 2005.

[2] L. Albera, A. Ferréol, P. Comon, and P. Chevalier, "Sixth order blind identification of underdetermined mixtures (BIRTH) of sources," in *Proc. 4th Int. Symp. Independent Component Analysis and Blind Signal Separation (ICA'03)*, Nara, Japan, Apr. 2003.

[3] L. Albera, A. Ferréol, P. Comon, and P. Chevalier, "Blind identification of overcomplete mixtures of sources (BIOME)," *Lin. Alg. Appl.*, vol. 391, pp. 1–30, Nov. 2004.

[4] A. Belouchrani, K. Abed-Meraim, M. G. Amin, and A. M. Zoubir, "Joint antidiagonalization for blind source separation," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP-01)*, Salt Lake City, UT, May 2001, vol. 5, pp. 2789–2792.

[5] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Process.*, vol. 81, no. 11, pp. 2353–2362, Nov. 2001.

[6] J.-F. Cardoso, "Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP-91)*, Toronto, ON, Canada, 1991, pp. 3109–3112.

[7] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elect. Eng., F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[8] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 17, no. 1, pp. 161–164, Jan. 1996.

[9] J. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition," *Psychometrika*, vol. 35, pp. 283–319, 1970.

[10] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," Dept. Statist., Stanford Univ., Stanford, CA, Tech. Rep. 30401, 1996.

[11] P. Chevalier and A. Ferréol, "On the virtual array concept for the fourth-order direction finding problem," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2592–2595, Sep. 1999.

[12] P. Chevalier, L. Albera, A. Ferréol, and P. Comon, "On the virtual array concept for higher order array processing," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1254–1271, Apr. 2005.

[13] P. Comon, "Independent component analysis, A new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[14] P. Comon and B. Mourrain, "Decomposition of quantics in sums of powers of linear forms," *Signal Process.*, vol. 53, no. 2, pp. 93–107, Sep. 1996.

[15] P. Comon, "Blind identification and source separation in 2×3 underdetermined mixtures," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 11–22, Jan. 2004.

[16] P. Comon, "Canonical tensor decompositions," Lab. I3S, Sophia-Antipolis, France, Tech. Rep. RR-2004-17, Jun. 2004.

[17] P. Comon and M. Rajih, "Blind identification of under-determined mixtures based on the characteristic function," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'05)*, Philadelphia, PA, Mar. 18–23, 2005, vol. 4, pp. 1005–1008.

[18] L. De Lathauwer, P. Comon, and B. De Moor, "ICA algorithms for 3 sources and 2 sensors," in *Proc. 6th Signal Processing Workshop on Higher Order Statistics*, Caesarea, Israel, Jun. 14–16, 1999, pp. 116–120.

[19] L. De Lathauwer, B. De Moor, and J. Vandewalle, "An algebraic ICA algorithm for 3 sources and 2 sensors," in *Proc. 10th Eur. Signal Processing Conf. (EUSIPCO 2000)*, Tampere, Finland, Sep. 5–8, 2000.

[20] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 26, pp. 295–327, 2004.

[21] L. De Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 28, no. 3, pp. 642–666, 2006.

[22] L. De Lathauwer and J. Castaing, "Second-order blind identification of underdetermined mixtures," in *Independent Component Analysis and Blind Source Separation, Proc. 6th Int. Conf. Independent Component Analysis (ICA 2006)*. Berlin, Germany: Springer-Verlag, Mar. 5–8, 2006, vol. 3889, Lecture Notes in Computer Science, pp. 40–47.
[23] L. De Lathauwer and J. Castaing, "Independent component analysis based on simultaneous matrix diagonalization: The underdetermined case," Lab. ETIS, Cergy-Pontoise, France, Tech. Rep., submitted for publication to *IEEE Trans. Signal Process*.
[24] Y. Deville, M. Benali, and F. Abrard, "Differential source separation for underdetermined instantaneous or convolutive mixtures: Concept and algorithms," *Signal Process.*, vol. 84, no. 10, pp. 1759–1776, Oct. 2004.
[25] M. C. Dogan and J. M. Mendel, "Applications of cumulants to array processing—Part I: Aperture extension and array calibration," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1200–1216, May 1995.
[26] D. Erdogmus, L. Vielva, and J. C. Principe, "Nonparametric estimation and tracking of the mixing matrix for underdetermined blind source separation," in *Proc. ICA'01*, San Diego, CA, Dec. 2001, pp. 189–194.
[27] J. Eriksson and V. Koivunen, "Identifiability, separability, and uniqueness of linear ICA models," *IEEE Signal Process. Lett.*, vol. 11, no. 7, pp. 601–604, Jul. 2004.
[28] A. Ferréol, L. Albera, and P. Chevalier, "Fourth order blind identification of underdetermined mixtures of sources (FOBIUM)," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1640–1653, May 2005.
[29] P. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Trans. Neural Netw.*, vol. 16, no. 5, pp. 992–996, Jul. 2005.
[30] R. A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an 'explanatory' multi-mode factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.
[31] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.
[32] J. B. Kruskal, "Rank, decomposition, and uniqueness for 3-way and N-way arrays," in *Multiway Data Analysis*, R. Coppi and S. Bolasco, Eds. Amsterdam, The Netherlands: North-Holland: Elsevier Science, 1989.
[33] Y. Li, A. Cichocki, and S.-I. Amari, "Analysis of sparse representation and blind source separation," *Neural Comput.*, vol. 16, no. 6, pp. 1193–1234, Jun. 2004.
[34] M. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Comput.*, vol. 12, no. 2, pp. 337–365, Feb. 2000.
[35] C. R. Rao and S. Mitra, *Generalized Inverse of Matrices and Its Applications*. New York: Wiley, 1971.
[36] N. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *J. Chemometrics*, vol. 14, no. 3, pp. 229–239, May/Jun. 2000.
[37] A. Stegeman, J. M. F. ten Berge, and L. De Lathauwer, "Sufficient conditions for uniqueness in candecomp/parafac and indscal with random component matrices," *Psychometrika*, vol. 71, no. 2, pp. 219–229, Jun. 2006.
[38] A. Taleb, "An algorithm for the blind identification of N independent signals with 2 sensors," in *Proc. 16th Int. Symp. on Signal Processing and Its Applications (ISSPA'01)*, Kuala-Lumpur, Malaysia, Aug. 13–16, 2001, pp. 5–8.
[39] F. Theis, E. W. Lang, and C. G. Puntonet, "A geometric algorithm for overcomplete linear ICA," *Neurocomputing*, vol. 56, pp. 381–398, Jan. 2004.
[40] A.-J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1136–1155, May 1996.
[41] A.-J. van der Veen, "A Schur method for low-rank matrix approximation," *SIAM J. Matrix Anal. Appl.*, vol. 17, pp. 139–160, Jan. 1996.
[42] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1545–1553, Jul. 2002.

Lieven De Lathauwer (M'04–SM'06) was born in Aalst, Belgium, on November 10, 1969. He received the Master's degree in electro-mechanical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven (K.U. Leuven), Leuven, Belgium, in 1992 and 1997, respectively. His Ph.D. thesis concerned signal processing based on multilinear algebra.

He is currently with the Centre National de la Recherche Scientifique, Cergy-Pontoise, France. His research interests include linear and multilinear algebra, statistical signal and array processing, higher order statistics, independent component analysis, identification, blind identification, and equalization. He is Associate Editor of the *SIAM Journal on Matrix Analysis and Applications*.

Joséphine Castaing was born in Paris, France, in 1978. She received the Master's degree in signal processing and the Ph.D. degree in applied sciences from the University of Cergy-Pontoise, Cergy-Pontoise, France, in 2003 and 2006, respectively.

Her research interests are algebraic methods for source separation.

Jean-François Cardoso is currently Directeur de Recherche at the Centre National de la Recherche Scientifique, in the Signal and Image Processing Department, Ecole Nationale Supérieure des Télécommunications, Paris, France. His research area is statistical signal processing. Since 1989, he has been extensively working on all aspects of blind source separation and independent component analysis. Since 2001, he has collaborated closely with cosmologists for the statistical analysis of astronomic data.