A Generalization of Joint-Diagonalization Criteria for Source Separation

Eric Moreau, Member, IEEE

Abstract—In the field of blind source separation, joint-diagonalization-based approaches constitute an important framework, leading to useful algorithms such as the popular joint approximate diagonalization of eigenmatrices (JADE) and simultaneous third-order tensor diagonalization (STOTD) algorithms. However, they are often restricted to the case of cumulants of order four. In this paper, we extend the results leading to JADE and STOTD to cumulants of any order greater than or equal to three by exhibiting a new family of contrast functions that constitutes then a unified framework for the above known results. This also leads us to generalize some links between contrast functions and joint-diagonalization criteria on which these algorithms are based. In turn, one contrast of the new family allows us to show that a function recently proposed as a separation criterion is also a contrast. Moreover, for the two generalized JADE and STOTD contrasts, the analytical optimal solution in the case of two sources is derived and shown to keep the same simple expression, whatever the cumulant order. Finally, some computer simulations illustrate the potential advantage one can take by considering statistics of different orders for the joint-diagonalization of cumulant matrices.

Index Terms—Blind source separation, contrast function, high order statistics, independent component analysis, joint-diagonalization.

I. INTRODUCTION

The problem of source separation has found numerous solutions in the past decade [1]–[22]. Beginning with the originate works of Hérault and Jutten (see [10] and [12] and references therein), who have proposed an adaptive (on-line) algorithm, two of the first most important contributions are provided by Comon [5] and Cardoso and Souloumiac [2]. These later solutions are block (off-line) algorithms that are both closely related to contrast functions (also simply called contrasts). Such contrasts were introduced and defined in [5] and have recently found a generalization in [16]. The algorithm proposed in [2] is called joint approximate diagonalization of eigen-matrices (JADE), whereas the algorithm presented in [5] is called independent component analysis (ICA).

The JADE algorithm is derived using fourth-order cumulants, and JADE’s underlying contrast is thus given with fourth-order cumulants as opposed to ICA, which remains available whatever the order of cumulants is since it is greater than or equal to three. The fourth-order JADE contrast has also found interesting interpretations in terms of a joint-diagonalization criterion [2] and of a least-squares criterion [23]. The link with a joint-diagonalization criterion is the key for the derivation of the practical JADE algorithm.

More recently, two new algorithms optimizing contrast functions were proposed [7], [8], [17]. The one developed in [17] and called one stage iterative block separation (OSIBS) is derived in order to easily and directly maximize any kind of contrast using a gradient iterative procedure. The one proposed in [7] and [8] follows the same spirit as JADE in the sense that it is based on simultaneous diagonalization of third-order tensors built from fourth-order cumulants. In the following, we call it simultaneous third-order tensor diagonalization (STOTD), as proposed in [8].

Again, this latter algorithm is closely related to the maximization of a contrast function involving fourth-order cumulants.

In this paper, we are mainly interested in generalizing the underlying contrasts of JADE and STOTD and their links with joint-diagonalization criteria involving cumulants of any order greater than or equal to three. The main interest is then to be able to choose the cumulants order or to combine statistical information of different orders. Moreover, for both cumulant matrices and cumulant third-order tensors, the analytical solution in the case of two sources is derived. In particular, these solutions are shown to keep the same expression, whatever the order of the considered cumulants. As exemplified by the computer simulations, the use of third-order cumulant or the use of both third- and fourth-order cumulants can yield an algorithm with better performances.

The paper is organized as follows. First, in Section II, we state the problem formulation with the necessary assumptions regarding the input source signals and the mixing matrix. In Section III, we first give some reminders about contrast functions, and then, we generalize the underlying contrasts of JADE and STOTD, putting them together in a common family. In Section IV, after recalling the definition of a joint diagonalizer of a given matrix set, we show how the new “any order” JADE contrast can also be interpreted as a joint-diagonalization criterion of some cumulants matrices. We also show that the new “any order” STOTD contrast can be interpreted as a joint-diagonalization criterion of some cumulant third-order tensors. In Section V, we derive the analytical optimal solution of the new generalized JADE and STOTD contrasts. Finally, in Section VI, we propose some computer simulations to illustrate the potential interest one can find in joint-diagonalizing cumulant matrices of different orders and of combined orders.
II. Problem Formulation

Signals emitted from different sources are observed thanks to

\[ x(n) = Ga(n) + b(n) \]  

(1)

where \( n \in \mathbb{Z} \) is the discrete time; \( a(n) \) is the \((N, 1)\) vector of \( N \neq 2 \) unobservable real input signals \( a_i(n), i \in \{1, \ldots, N\} \), called sources; \( x(n) \) is the \((N, 1)\) vector of observed signals \( x_i(n); i \in \{1, \ldots, N\} \); \( b(n) \) is the \((N, 1)\) vector of additive noises; and \( G \) is the \((N, N)\) square mixing matrix assumed invertible. For clarity, in this paper, we restrict our attention to the case of real signals and mixtures.

Further, the following assumptions are considered.

A1a) “Independence” The sources \( a_i(n), i \in \{1, \ldots, N\} \), are zero-mean, unit power and statistically mutually independent.

A1b) “Stationarity” \( a(n) \) is a random vector stationary up to order under consideration, i.e., \( \forall i, a_i(n) \) the cumulant \( \text{Cum}(a_i(n), \ldots, a_i(n)) \) is an independent function of \( n \) denoted by \( C_{\text{R}}(a_i) \); moreover, at most, one of the cumulants \( C_{\text{R}}(a_i), i \in \{1, \ldots, N\} \), is null.

A1c) “Gaussian noise” \( b(n) \) is a Gaussian zero-mean random vector, statistically independent of the sources vector and whose components are mutually independent from each other.

It is important now to introduce the notion of white vectors [2], [5] because of its use as a first transformation in the JADE and STOTD algorithms. A vector \( z \) is said to be (spatially) white if its covariance matrix \( R_z = E[zz^T] \) equals the identity. The first second-order transformation is then defined as a whitening of the noiseless part of the observation vector \( x \). This is done by applying a whitening matrix \( B \) in such a way that \( BB^T = V \), where \( V \) is a unitary matrix, i.e., \( VV^T = I \). Hence, after the whitening transformation, the new “observed” vector reads

\[ z(n) = Bx(n) = V a(n) + b(n) \]  

(2)

where \( b(n) = Bl(n) \) is a linear transformation of a Gaussian vector and, thus, is also a Gaussian vector.

The blind source separation problem consists now of estimating a unitary matrix \( H \) in such a way that the vector

\[ y(n) = Hs(n) \]  

(3)

restores one of the different (possibly noisy) sources on each of its different components.

Because the sources are inobservable and the mixture is unknown, the exact power and order of each sources cannot be recovered. This is why the separation is said to be achieved when the global unitary matrix \( S \) defined by

\[ S \overset{\text{def}}{=} HV \]  

(4)

reads

\[ S = DP \]  

(5)

where \( D \) is an invertible diagonal matrix (here with unit modulus components) corresponding to arbitrary attenuations for the restored sources and \( P \) a permutation matrix corresponding to an arbitrary order of restitution. According to (3), (1), and (4), the output vector can be written as

\[ y(n) = Sa(n) + Hb(n). \]  

(6)

Because of the stationarity assumption, the explicit dependence of sources, observations, and output vectors with the discrete time \( n \) will be now omitted whenever no confusion is possible. Moreover, as we only consider cumulants of order greater than or equal to three in the next sections, we can consider a noise-free model in (6) thanks to the well-known properties of cumulants with respect to Gaussianity and independence.

Let us define some notations that will be useful in the following. Let \( \mathcal{A} \) be the set of random vectors satisfying assumptions A1a) and A1b). Let \( \mathcal{U} \) be the set of unitary matrices. The subset of \( \mathcal{U} \) of matrices \( S \) of the form (5) is denoted by \( \mathcal{P} \), and the subset of \( \mathcal{P} \) of diagonal matrices is denoted by \( \mathcal{D} \). Finally, the set of random vector \( y \) built from (6), where \( a \in \mathcal{A} \) and \( S \in \mathcal{U} \), is denoted by \( \mathcal{Y} \).

III. Generalization and New Contrasts

In this section, we are mainly interested in generalizing the JADE and STOTD contrasts. First, however, we recall some contrasts.

A. Recalls

A contrast is usually a function of the output of the separating system. As defined in [5], its (global) maximization arguments yield a separating solution, i.e., a matrix \( H \) such that the global matrix \( S \) can be factored as in (5).

Originally [5], a contrast is imposed to be a symmetric and scale invariant function in order not to discriminate one possible separating solution. Hence, the (global) maximization of a contrast is a necessary and sufficient condition for source separation, and then, contrast functions can be seen as measures of independence (at least for mixtures of a vector of independent sources). Recently, it was shown that the symmetry property of contrast is rather restrictive and not necessary. That is why a generalized definition was given in [16] and is now recalled for the readers convenience.

Definition 1: A contrast on \( \mathcal{Y} \) is a multivariate mapping \( I(\cdot) \) from \( \mathcal{Y} \) to the real set which satisfies the following three requirements.

\[ \text{R1.} \quad \forall y \in \mathcal{Y}, \quad \forall D \in \mathcal{D}, \quad I(Dy) = I(y). \]

\[ \text{R2.} \quad \forall a \in \mathcal{A}, \quad \forall S \in \mathcal{U}, \quad I(Sa) \leq I(a). \]

\[ \text{R3.} \quad \forall a \in \mathcal{A}, \quad \forall S \in \mathcal{U}, \quad I(Sa) = I(a) \Rightarrow S \in \mathcal{P}. \]

According to this definition, the maximization of a contrast (generally nonsymmetric) becomes a sufficient condition for source separation. Now, because all the contrasts considered in the following are defined on \( \mathcal{Y} \), then for simplicity in the text, we only say “contrast” rather than “contrast on \( \mathcal{Y} \)”.

Historically, one of the first contrasts can be found in [5]. It reads

\[ I_R(y) = \sum_{i=1}^{N} |C_{R[y]}|^2, \quad R = 3, 4, \ldots \]  

(7)
and is called the ICA contrast in the following. It is also shown in [5] that the function

\[ B_R(\mathbf{y}) = \sum_{i_1, \ldots, i_R = 1}^{N} \left( \text{Cum}[y_{i_1}, \ldots, y_{i_R}] \right)^2 \]  

(8)

is invariant under unitary transform on \( \mathbf{y} \), i.e., if \( UU^T = I \), then \( B_R(\mathbf{Uy}) = B_R(\mathbf{y}) \). This result leads to another interpretation of contrast \( T_R^0(\mathbf{y}) \). The maximization of \( T_R^0(\mathbf{y}) \) is thus equivalent to the minimization of the sum of all the squared cross-cumulants of the same order \( R \), which can even be interpreted as an independence measure (at order \( R \)).

Later, it was shown in [13] and [14] that the square absolute values of cumulants in (7) can be dropped, leading to the contrast

\[ T_R^0(\mathbf{y}) = \sum_{i_1, \ldots, i_R = 1}^{N} |C_R[y_{i_1}]|, \quad R = 3, 4, \ldots \]  

(9)

These two above contrasts \( T_R^0(\mathbf{y}) \) and \( T_R^0(\mathbf{y}) \), have recently found a useful generalization \(^1\) in [16] as

\[ T_R(\mathbf{y}) = \sum_{i_1, \ldots, i_R = 1}^{N} \gamma_i f(\text{Cum}[y_{i_1}])), \quad R \geq 3 \]

where \( f(\cdot) \) is a strictly increasing convex function and where the real numbers \( \gamma_i, i \in \{1, \ldots, N\} \) are assumed to satisfy \( \gamma_1 \geq \cdots \geq \gamma_N > 0 \). Thus, considering \( \forall i, \gamma_i = 1 \), the original contrast in (7) is recovered, and if \( f(u) = u^2 \), then we have the contrast in (9). This function \( T_R(\mathbf{y}) \) provides a parameterized family of contrasts, which allows a more flexible way to use them. For example, in [16], the optimal coefficients [minimizing the mean square error of the estimated rotation angle, cf. (34)] was derived in a simple case.

One can notice that there are no cross-cumulants in the above contrasts. In [2] and [7], two contrasts involving both cross-cumulants and auto-cumulants have been proposed. The first one [2] is called the “JADE contrast” and reads

\[ \mathcal{J}(\mathbf{y}) = \sum_{i_1 \neq j_1, i_2 \neq j_2}^{N} \left( \text{Cum}[y_{i_1}, y_{j_1}, y_{i_2}, y_{j_2}] \right)^2 \]  

(10)

whereas the second one [7], [8] is called the “STOTD contrast” and reads

\[ \mathcal{S}(\mathbf{y}) = \sum_{i_1 \neq j_1, i_2 \neq j_2}^{N} \left( \text{Cum}[y_{i_1}, y_{j_1}, y_{i_2}, y_{j_2}] \right)^2 \]  

(11)

B. More General Family of Contrasts Involving Cross-Cumulants

Contrary to \( T_R(\mathbf{y}) \) in (7), the contrasts \( \mathcal{J}(\mathbf{y}) \) in (10) and \( \mathcal{S}(\mathbf{y}) \) in (11) “only” rely on fourth-order cumulants. We now generalize these functions to cumulants of any order greater than or equal to three, including them in a common generalized family of contrasts. This is given by the following proposition.

\(^1\)This generalization to nonsymmetric contrasts cannot be integrated in the framework presented in this paper. Here, we are mainly concerned with links between contrasts and quadratic joint-diagonalization criteria that are symmetric.

**Proposition 1:** Let \( R, R_1, R_2, \) and \( R_3 \) be four integers such that \( R \geq 3, 2 \leq R_1 \leq R, 0 \leq R_2 \leq R - R_1 \) and \( R_3 = R - R_1 - R_2 \), using the notation

\[ C_R^{R_1, R_2}[:, \mathbf{y}] = \text{Cum}_{R_1 \times R_2 \times R_3 \text{ terms}} \left[ y_{i_1}, \ldots, y_{i_2}, \ldots, y_{i_j}, \ldots, y_{i_R} \right] \]

(12)

The function

\[ C_R^{R_1, R_2}(\mathbf{y}) = \sum_{i_1, \ldots, i_R = 1}^{N} \left( C_R^{R_1, R_2}[:, \mathbf{y}] \right)^2 \]

(13)

is a contrast in \( \mathcal{V}_w \), i.e., for white vectors \( \mathbf{y} \).

**Proof:** A proof is given in Appendix A.

Now, because \( B_R(\mathbf{y}) \) in (8) is invariant under unitary transform of its argument vector \( \mathbf{y} \), then the maximization of \( C_R^{R_1, R_2}(\mathbf{y}) \) is equivalent to the minimization of the sum of the squared cross-cumulants of order \( R \) with their \( R_1 \) first indexes or their \( R_2 \) following indexes different.

By definition, if \( R_2 = 0 \) or \( R_3 = 0 \), no corresponding additional terms are considered in the cumulant in (12). Hence, in (13), if \( R_2 = 0 \) (or resp. \( R_3 = 0 \)), no sum over \( i_2 \) (or resp. \( j_1, \ldots, j_{R_3} \)) is considered. Now, let us remark that if \( R_2 = R_3 = 0 \), then \( C_R^{0,0}(\mathbf{y}) = T_R^0(\mathbf{y}) \), which is the ICA’s contrast. Moreover, considering \( R_2 = 0 \), if \( R_1 = 2 \), then \( C_R^{2,0}(\mathbf{y}) = \mathcal{J}(\mathbf{y}) \), which is the JADE’s contrast, whereas if \( R_3 = 3 \), then \( C_R^{3,0}(\mathbf{y}) = \mathcal{S}(\mathbf{y}) \), which is the STOTD’s contrast. Thus, the result stated in Proposition 1 gives a more general framework for contrasts, proving that \( T_R^0(\mathbf{y}) \), \( \mathcal{J}(\mathbf{y}) \), and \( \mathcal{S}(\mathbf{y}) \), independently proposed and with rather different proofs, belong, in fact, to the same family.

Concerning fourth-order cumulants only, let us notice that contrasts \( T_R^0(\mathbf{y}) \) in (7), \( \mathcal{J}(\mathbf{y}) \) in (10), and \( \mathcal{S}(\mathbf{y}) \) in (11) do not recover all possible contrasts in the new family. Indeed, there is another contrast that is in fact given by \( C_4^{2,2}(\mathbf{y}) \) as

\[ C_4^{2,2}(\mathbf{y}) = \sum_{i_1, i_2 = 1}^{N} \left( \text{Cum}[y_{i_1}, y_{i_2}, y_{i_1}, y_{i_2}] \right)^2 \]

Now, we can remark that it is always possible to write\(^2\)

\[ C_4^{2,2}(\mathbf{y}) = \phi_{\text{SH}}(\mathbf{y}) = B_4(\mathbf{y}) = B_4(\mathbf{a}) \]

(14)

where \( \phi_{\text{SH}}(\mathbf{y}) = \sum_{i,j,k,l} \text{Cum}[y_i, y_j, y_k, y_l] \). The above notation means that the sum is over the four indexes \( i, j, k, l \) such that \( i \neq j \neq k \neq l \). Hence, the sum of \( C_4^{2,2}(\mathbf{y}) \) and \( \phi_{\text{SH}}(\mathbf{y}) \) is a constant since \( B_4(\mathbf{a}) \) no longer depends on the separating matrix \( \mathbf{H} \) we are looking for. Then, the maximization of \( C_4^{2,2}(\mathbf{y}) \) is equivalent to the minimization of \( \phi_{\text{SH}}(\mathbf{y}) \), which thus implies that \( \phi_{\text{SH}}(\mathbf{y}) \) is a contrast. The function \( \phi_{\text{SH}}(\mathbf{y}) \) (with the same notation) is given in [3], where it is derived as an “approximation of the mutual information contrast.” The above derivations show that the minimization of this approximation \( \phi_{\text{SH}}(\mathbf{y}) \) is in fact a necessary and sufficient condition for the separation of sources.

Finally, let us point out that the family of contrasts described by \( C_R^{R_1, R_2}(\mathbf{y}) \) in (13) does not involve all classical contrasts.

\(^2\)The notation \( \phi_{\text{SH}}(\mathbf{y}) \) is kept from [3] to allow easy reference.
e.g., $I_R^H(y)$ in (9) does not belong to this family. Moreover, it does not contain all cross-cumulant based contrasts as recently shown in [18]. The main interest in considering $G^{R_1,R_2}_R(y)$ is its links with quadratic joint-diagonalization criteria, as shown in the following section.

IV. LINKS WITH QUADRATIC JOINT-DIAGONALIZATION

A. Joint-Diagonalization of Matrices

As pointed out in [2], one of the main interests in considering contrast $J_R(y)$ is its close relation with a joint-diagonalization criterion of a matrices set. Such joint-diagonalization is defined according to the following definition.

Definition 2: Let us consider a set, which is denoted by $M$, of $M$ square $(N, N)$ matrices $M(m), m = 1, \cdots, M$. A joint-diagonalizer $H$ of this set is a unitary matrix that maximizes the function

$$D(H, M) = \sum_{m=1}^{M} \left| \text{diag}(HM(m)H^T) \right|^2$$

where diag($\cdot$) stands for the vector built from the diagonal components of the matrix argument, and $|\cdot|$ stands for the Euclidean norm.

It is easily seen that the above function can be written as

$$D(H, M) = \sum_{i=1}^{N} \left( \sum_{m=1}^{M} |M_{i,i}^H(m)|^2 \right)$$

where

$$M_{i,i}^H(m) = \sum_{n_1,n_2} H_{i,n_1}H_{i,n_2}M_{n_1,n_2}(m).$$

In the following, we show that a link also exists for the any order generalization of $J_R(y)$ that is $G_R^2(y)$ [cf. (13)]. To simplify, this latter function will be now denoted by $J_R(y) = G_R^2(y)$. Before establishing this link, however, we need to introduce some notations. First, let us consider the linear operator $Q_{\leq R}$ that maps $R-2$ order tensors $O^{R-2}(r) = (O_{i_3 \cdots i_R}(r)), r = 1, \cdots, N^{R-2}$ onto matrices $L(r) = (L_{i_1,i_2}(r)), r = 1, \cdots, N^{R-2}$. It is written as

$$L(r) = Q_{\leq R}O^{R-2}(r),$$

and it is defined component-wise according to

$$L_{i_1,i_2}(r) = \sum_{i_3,\cdots, i_R} \text{Cum}(x_{i_3}, \cdots, x_{i_R})O^{R-2}(i_3, \cdots, i_R),$$

Let us also denote

$$M_R^2(O^{R-2}) = \{L(r); r = 1, \cdots, N^{R-2}\}$$

as the set of matrices defined in (18) and (19) for all values of $r$. The link between the contrast $J_R(\cdot)$ and a joint-diagonalization criterion $D(\cdot, \cdot)$ is now stated thanks to the following proposition.

Proposition 2: Letting $R \geq 3$ for any set

$$\{O^{R-2}(r) = (O_{i_3 \cdots i_R}(r)), r = 1, \cdots, N^{R-2}\}$$

of orthonormal $(N, \cdots, N)$ tensors\(^3\) if $H$ is a unitary matrix, we have

$$D(H, M_R^2(O^{R-2})) = J_R(Hx).$$

Proof: This is a particular case of Proposition 4 with $H_1 = 2$. Thus, see the proof of Proposition 4.

Hence, if we use any orthonormal set of $R-2$ order tensors for building matrices of linear combination of $R$-order cumulants as in (19), then the joint-diagonalization (in the sense of Definition 2) of all these matrices is equivalent to the maximization of the contrast $J_R(\cdot)$. It is thus a necessary and sufficient condition for separation.

A consequence of the above proposition is given by the following.

Proposition 3: Let $R \geq 3$ if $G_R^2$ is the set of $M = N^{R-2}$ matrices

$$C(i_3, \cdots, i_R) = C(i_3, i_2(i_3, \cdots, i_R))$$

defined as

$$C_{i_1,i_2} (i_3, \cdots, i_R) = \text{Cum}(x_{i_1}, x_{i_2}, x_{i_3}, \cdots, x_{i_R}).$$

Then, if $H$ is a unitary matrix, we have

$$D(H, C_R^2) = J_R(Hx).$$

Proof: This is a particular case of Proposition 5 with $H_1 = 2$. Thus, see the proof of Proposition 5.

Proposition 3 considers a particular choice of the orthonormal set of $(R-2)$-order tensors under consideration in Proposition 2. This set corresponds to the canonical basis of $(R-2)$-order tensors, where each tensor has one and only one nonzero component that is 1. This canonical basis leads to matrices in (19) whose components are only one cumulant of order $R$; see (22) and (23). Thus, the joint-diagonalization of these particular “simple” matrices is again equivalent to the maximization of the contrast $J_R(\cdot)$ and then is a necessary and sufficient condition for separation.

In Propositions 2 and 3, the number of matrices to be joint-diagonalized increases exponentially with the order of the considered cumulants. For cumulants of order 3, $N$ matrices are needed, whereas $N^2$ matrices are required when fourth-order cumulants are considered. In [2], it is shown that in the case of fourth-order cumulants, the number of matrices that have to be effectively joint-diagonalized can be reduced to $N^2$ thanks to an appropriate eigendecomposition of the linear operator $Q_{\leq R}$ in (18). This is also the case with our generalized result, i.e., it can be easily shown that the joint-diagonalization of $N$ matrices is sufficient (and necessary), whatever the order of cumulants. This latter result again requires a particular orthonormal set of $(R-2)$-order tensors corresponding to a “diagonalization” (or an “eigen-decomposition”) of the linear operator $Q_{\leq R}$. However, such a diagonalization task leads to an increasing computational cost, and moreover, it only holds when the model of

\(^3\)We say that the set $\{O^{R-2}(r) = (O_{i_3 \cdots i_R}(r)), r = 1, \cdots, N^{R-2}\}$ is an orthonormal set of $R-2$ order tensors if $\forall r_1, r_2, \sum_{i_3,\cdots, i_R} O^{R-2}(i_3, \cdots, i_R)O^{R-2}(i_1, \cdots, i_R) = \delta_{r_1,r_2}$.}
mixture is exactly true, i.e., $x = Ga$ is explicitly needed. Thus, in practice, if one is not sure that the available data conform the initial model (e.g., existence of small delays or nonlinearities in the model), then the above decomposition is not really a very good idea.

However, for computation load, if one needs to reduce the number of matrices to be joint-diagonalized, the symmetric properties of cumulants can be advantageous used [3]. This can be realized thanks to a good choice of an orthonormal basis in Proposition 2, rather than the canonical one (corresponding to a maximal number of matrices to be joint-diagonalized in Proposition 3) or the “diagonalization” one (corresponding to the minimal number of matrices to be joint-diagonalized but requiring an exact model). For example, in the case of fourth-order cumulant ($R = 4$), one can consider the following orthonormal basis:

$$\begin{align*}
O_i^2(\xi_3, i_4) &= M_i(\xi_3, i_4), \quad \text{if } i_4 = i_3 \\
&= M_3(\xi_3, i_4), \quad \text{if } i_3 < i_4 \\
&= M_4(\xi_3, i_4), \quad \text{if } i_3 > i_4
\end{align*}$$

where

$$\begin{align*}
M_i(\xi_3, i_4) &= 1^{i_4}(1^{i_4})^T \\
M_3(\xi_3, i_4) &= \frac{1}{\sqrt{2}} (1^{i_4}(1^{i_4})^T + 1^{i_3}(1^{i_3})^T) \\
M_4(\xi_3, i_4) &= \frac{1}{\sqrt{2}} (1^{i_4}(1^{i_4})^T - 1^{i_3}(1^{i_3})^T)
\end{align*}$$

with $(\cdot)^T$ denotes the transposition operator and where $1^k$ denotes the $(N, 1)$ vector with 1 in the $k$th position and 0 elsewhere, i.e., $(1^1) = \delta_{1,j}$.

Now, because cumulants are symmetric, one easily shows that $Q_i[1^{i_4}(1^{i_4})^T] = Q_3[1^{i_3}(1^{i_3})^T]$ and then, $Q_i[M_3(\xi_3, i_4)] = 0$. Hence, the set of $N^2$ matrices $N^2(O^2)$ under consideration in Proposition 2 can be reduced to the set of $N(N + 1)/2$ matrices $\{Q_3[M_i(\xi_3, i_4)]; Q_3[M_3(\xi_3, i_4)]\}$, the other $N(N - 1)/2$ ones being identically zero.

### B. Joint-Diagonalization of Tensors

In [7], the function $S(y)$ is linked to a joint-diagonalization criterion of a set of third-order tensors. An identical link also exists for the any order generalization of $S(y)$ given in (13) and referred to as $G_R^{3.0}(y)$. However, let us remark that the order of the tensors to be joint-diagonalized is two when it corresponds to contrast $G_R^{3.0}(y)$ and three when it corresponds to contrast $G_R^{2.0}(y)$. Hence, it corresponds to the value of $R_3$ for the contrast. We now propose a generalization of all these results.

Let us first define the joint-diagonalization of $R_1$-order tensors by a unitary matrix as the following.

**Definition 3:** Consider a set of $T$ square $(N, \cdots, N) R_1$ order tensors $T_{R_1}(m), m = 1, \cdots, T$ denoted by $T$. A joint-diagonalizer of this set is a unitary matrix that maximizes the function

$$D(H, T) = \sum_{m=1}^{T} \left( \sum_{i=1}^{N} T_{R_1}^H(i,m) T_{R_1}^H(i,m) \right)^2$$

where

$$T_{R_1}^H(i,m) = \sum_{n_1, \cdots, n_{R_1}} H_{i,n_1, \cdots, n_{R_1}} T_{n_1, \cdots, n_{R_1}}^H(m).$$

We also need some notations. With $R_3 \leq R$, let us consider the linear operator $R_{R_3}^R$ that maps $(R - R_3)$-order tensors $O_{R_3-R_1}(r) = (O_{R_3-R_1}^{R_3-R_1}(r), r = 1, \cdots, N^{R_3-R_1}$ onto $R_1$-order tensors $L_{R_1}^R(r) = (L_{R_1}^{R_1}(r), r = 1, \cdots, N^{R_3-R_1}$ and it is defined component-wise according to

$$L_{R_1}^{R_1}(r) = R_{R_3}^R[O_{R_3-R_1}^R(r), r = 1, \cdots, N^{R_3-R_1}$$

By definition, if $R_3 = R$, there is no linear operator, and then,

$$L_{R_1}^{R_1}(1) = L_{R_1}^{R_1} = \text{Cum}[x_{i_1}, \cdots, x_{i_R}]$$

We also denote

$$T_{R_3}^R(O_{R_3-R_1}^R) = (L_{R_1}^{R_1}(r); r = 1, \cdots, N^{R_3-R_1}$$

as the set of $R_1$-order tensors defined in (27) and (28) for all values of $r$.

Now, the equivalence between the contrast $G_R^{R_3}(y)$ and a joint-diagonalization criterion of $R_1$ order tensors set is given by the following proposition.

**Proposition 4:** Letting $R \geq 3$ and $2 \leq R_3 \leq R$ for any set

$$\{O_{R_3-R_1}^{R_3-R_1}(r); r = 1, \cdots, N^{R_3-R_1}$$

of orthonormal $(N, \cdots, N)$ tensors, if $H$ is a unitary matrix,

$$D(H, T_{R_3}^R(O_{R_3-R_1}^R)) = G_R^{R_3}(Hx).$$

**Proof:** A proof is given in Appendix B.

This is a generalization of the result in Proposition 2 for the case $R_3 \geq 2$. Hence, if we use any orthonormal set of $(R - R_3)$-order tensors for building $R_1$-order tensors of linear combinations of $R$-order cumulants as in (28), then the joint-diagonalization (in the sense of Definition 3) of all these $R_1$-order tensors is equivalent to the maximization of the contrast $G_R^{R_3}(\cdot)$. It is thus a necessary and sufficient condition for separation.

A direct consequence of the above proposition is given by the following.

**Proposition 5:** Letting $R \geq 3$ and $2 \leq R_3 \leq R$, if $C_R^{R_3} = \{i_{R_3+1}, \cdots, i_R\}$ defined as

$$C_R^{R_3}(i_{R_3+1}, \cdots, i_R) = \text{Cum}[x_{i_1}, x_{i_2}, x_{i_3}, \cdots, x_{i_R}]$$

then if $H$ is a unitary matrix, we have

$$D(H, C_R^{R_3}) = G_R^{R_3}(Hx).$$

(32)
Proposition 5, which generalizes Proposition 3, considers a particular choice of the orthonormal set of $(R - R_3)$-order tensors under consideration in Proposition 4. This set corresponds to the canonical basis of $(R - R_3)$-order tensors. This canonical basis leads to $R_1$-order tensors in (28) whose components are only one cumulant of order $R_1$; see (31). Thus, the joint-diagonalization of these particular “simple” $R_1$-order tensors is again equivalent to the maximization of the contrast $J_{R_1}$ and is then a necessary and sufficient condition for separation.

For $R_1 = 3$, Proposition 5 is a generalization of one result in [8] to cumulants of any order greater than or equal to three. Proposition 4 is a general result that allows us to consider different orthonormal basis of the set of $(R - R_3)$-order tensors rather than the canonical one, as in Proposition 5. As exemplified in the above Section IV-A, this can be useful for the reduction of the number of tensors to be joint-diagonalized.

C. Combining Cumulants of Mixed Order

According to the results in Sections IV-A and B, we can now choose the order of cumulants (greater than or equal to three) for the joint-diagonalization of matrices or more generally tensors. In particular, third-order cumulants can be used, leading to the joint-diagonalization of $N$ matrices or to the diagonalization of only one tensor of order three. However, even if it is sufficient to joint-diagonalize tensors of cumulants of a given order, one can find interest in combining cumulants of different orders. For example, this can lead to algorithms that are more robust w.r.t. the statistics of sources. For example, one can combine third- and fourth-order cumulants. If third-order (resp. fourth-order) cumulants of the unknown sources vanish, then the other fourth-order (resp. third-order) ones can be directly used.

V. GENERALIZATION OF THE ALGORITHMS

The JADE and STOTD algorithms are both based on Jacobi optimization. This means that the maximization of the criterion under consideration is realized through a sequence of plane (or Givens) rotations, as initiated in [5]. Each plane rotation works on a pair of the output vector $\mathbf{g}(n)$ and one “sweep” or iteration consists of processing the outputs through all the $N(N - 1)/2$ possible pairs. Hence, the $N$-dimensional problem is reduced to $N(N - 1)/2$ 2-D problems. One of the main advantages is that the 2-D problem is simpler and often admits an analytical solution. Thus, let us now consider the only 2-D problem where a plane rotation has to be determined. In the following, we parameterize it as

$$H = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \quad (34)$$

First, let us consider the joint-diagonalization of matrices and then the maximization of $J_{R_1}$. For $N = 2$, it is easily seen that $J_{R_1}$ can be written as the quadratic form

$$J_{R_1} = \mathbf{u}_0^T A_R \mathbf{u}_0 \quad (35)$$

where $\mathbf{u}_0 = (\cos 2\theta \sin 2\theta)$ and where $A_R = (A_{R_{i,j}})$ is a $(2, 2)$ real symmetric matrix defined according to

$$A_{R_{i,j}} = t_{R_{i,j}}$$

where, using (23) and defining $i = (i_3, \ldots, i_R)$

$$t_{R_{i,j}} = \sum_i (C_{1,i}(i))^2 + (C_{2,i}(i))^2$$

$$t_{R_{i,j}} = \sum_i (C_{1,i}(i))^2$$

$$t_{R_{i,j}} = \sum_i (C_{1,i}(i), C_{2,i}(i))$$

$$t_{R_{i,j}} = \sum_i (C_{1,i}(i), C_{2,i}(i))$$

Hence, the value of $\theta$ maximizing $J_{R_1}$ can now be easily found by computing the normalized (unit-norm) eigenvector of $A_R$ associated with the largest eigenvalue. The sign of this vector can be fixed by restricting $\theta$ between $-\pi/4$ and $\pi/4$.

Now, let us consider the joint-diagonalization of third-order tensors that corresponds to the maximization of $G_{R_1}^0$. This is written as $S_R^0(\mathbf{y}) = G_{R_1}^0(\mathbf{y})$ in the following. Using the same derivations as in [7], it is easily seen that for $N = 2$, $S_R$ can also be written as a quadratic form as

$$S_R(\mathbf{y}) = \mathbf{u}_0^T B_R \mathbf{u}_0$$

Proof: A proof is easily derived because the tensors to be joint-diagonalized are considered additively.

Now, since it is well known that a non-negative linear combination of contrasts is also a contrast, then the joint-diagonalization of tensors of cumulants of mixed orders is again a necessary and sufficient condition for separation.
with
\[
\begin{align*}
B_{R_1,1} &= t'_{R_1} \\
B_{R_1,2} &= \frac{3}{2} t'_{R_2} \\
B_{R_2,2} &= \frac{1}{4} t'_{R_1} + \frac{1}{2} t'_{R_2} + \frac{3}{2} t'_{R_3}
\end{align*}
\]
where, using (31) with \(R_1 = 3\) and defining \(\mathbf{r}' = (i_4, \ldots, i_R)\)
\[
\begin{align*}
t'_{R_1,1} &= \sum_{\mathbf{r}} (C_{1,1,1}(\mathbf{r}'))^2 + (C_{3,2,2}(\mathbf{r}'))^2 \\
t'_{R_1,2} &= \sum_{\mathbf{r}} (C_{3,1,1,2}(\mathbf{r}'))^2 + (C_{3,2,2}(\mathbf{r}'))^2 \\
t'_{R_2,1} &= \sum_{\mathbf{r}} C_{3,1,1,2}(\mathbf{r}') C_{3,2,2}(\mathbf{r}') + C_{1,1,1,2}(\mathbf{r}') C_{3,2,2}(\mathbf{r}') \\
t'_{R_2,2} &= \sum_{\mathbf{r}} C_{3,1,1,2}(\mathbf{r}') C_{3,2,2}(\mathbf{r}') - C_{3,2,2}(\mathbf{r}') C_{3,2,2}(\mathbf{r}').
\end{align*}
\]

Let us first notice that the expression of the contrasts \(J_S(\mathbf{y})\) and \(J_R(\mathbf{y})\) have exactly the same quadratic form, whatever the cumulant order in the case of two sources. This is a remarkable fact that will simplify a lot the use of cumulants of different orders in the algorithms. The optimal value of \(\theta\) can be found exactly as in the matrix case considered above, but an analytical optimal value of \(\eta_{x+y}\) can be directly derived from (35) or (36).
Indeed, one finds after some simple algebra that for a symmetric matrix \(\mathbf{Z} = (Z_{n,i})\)
\[
\begin{align*}
\mathbf{w}^T \mathbf{Z} \mathbf{w} &= D + \cos(4(\theta - \alpha))
\end{align*}
\]
where \(D = (Z_{1,1} + Z_{2,2})/2\) is a constant term since it does not depend on \(\theta\), and \(E\) is a non-negative constant term \(E = \sqrt{((Z_{1,1} - Z_{2,2})/2)^2 + Z_{1,2}^2}\). The angle \(\alpha\) can be determined
\[
\begin{align*}
\alpha &= \frac{1}{4} \arctan(Z_{1,2}, \frac{1}{2}(Z_{1,1} - Z_{2,2}))
\end{align*}
\]
where the value of \(\arctan(y, x)\) is, by definition, the unique angle \(\beta \in (-\pi, \pi]\) for which \(\cos(\beta) = (x/(x^2 + y^2)^{1/2})\) and \(\sin(\beta) = (y/(x^2 + y^2)^{1/2})\). Now, recalling that we are looking for the value of \(\theta\) denoted \(\theta_{opt}\), which maximizes a given contrast written as \(\mathbf{w}^T \mathbf{Z} \mathbf{w}\) in (37). Since \(D\) and \(E\) are constant and \(E\) is non-negative, the maximum value of the left term in (37) corresponds to the maximum value of the cosine, that is 1, yielding
\[
\theta_{opt} = \alpha.
\]

Let us remark that following the idea of Proposition 6 matrices or third-order tensors of cumulants of different orders can be considered very easily. Indeed, let us consider for example matrices of cumulants of orders \(S_1, \ldots, S_m, m \in \mathbb{N}^a\) such that \(3 \leq S_1 < \cdots < S_m\). Then, according to (35)
\[
\sum_{i=1}^{m} \gamma_i J_S(\mathbf{y}) = \mathbf{w}^T \sum_{i=1}^{m} \gamma_i A_{S_i} \mathbf{w}
\]
where the \(\gamma_i\)'s are real non-negative constants with at least one nonzero. Hence, it is sufficient to consider the matrix \(\mathbf{Z} = \sum_{i=1}^{m} \gamma_i \mathbf{A}_{S_i}\) for the determination of the value of \(\theta_{opt}\) in (38) and (39). Notice that the above development can also be done with the contrast \(S_R(\mathbf{y})\) because it takes exactly the same expression in the case of two sources, as given in (36). Finally, for the same reason, one can also combine linearly contrasts \(J_S(\mathbf{y})\) and \(J_R(\mathbf{y})\) for different values of \(S\) and \(R\). Such a generalized algorithm for joint-diagonalization of cumulant matrices is called eJADE for “extended JADE” when considering the original implementation of JADE [24] and adding directly the new matrices to be joint-diagonalized. However, even if the above derivations are realized for the goal of a joint-diagonalization task, we have for \(N = 2\) the optimal solutions of the contrasts under consideration. Thus, we can also use them in an implementation as ICA [25] for the sufficient task of maximization of the underlying contrast. In such a case, the algorithms are called ICaj and ICAs for, respectively, the one optimizing the generalized JADE’s contrast \(J_P(s)\) and the one optimizing the generalized STOTD’s contrast \(J_R(s)\).

VI. COMPUTER SIMULATIONS

In order to illustrate the potential usefulness of the above results, some computer simulations are now presented. We intend first to joint-diagonalize both third-order and/or fourth-order cumulant matrices in order to compare our results with the joint diagonalization of only fourth-order cumulant matrices. For the fourth-order case, we use the original JADE algorithm in its version 1.5 of December 1997 for real signals, whereas for the other cases, we consider eJADE in exactly the same conditions. We use both only third-order cumulant matrices whose corresponding algorithm is denoted eJADE(3) and third-plus-fourth-order cumulant matrices whose corresponding algorithm is denoted eJADE(3,4). Then, we intend to realize a first comparison between the ICaj and ICAs algorithms using the any order generalized contrasts with both third- and fourth-order cumulants. The corresponding algorithms are denoted by ICaj(3,4) and ICAs(3,4).

Now, for the illustration of the behavior of the different algorithms with respect to the source statistics, we have to build a signal, say \(a(n)\), with parameterized third- and fourth-order cumulants. Hence, for \(a(n)\), we consider a discrete i.i.d. signal called MS(\(\alpha\)) that takes its values in the set \(-1, 0, \alpha\) with the respective probability \(\{1/(1+\alpha), (\alpha - 1)/\alpha, 1/(\alpha(1 + \alpha))\}\). The real parameter \(\alpha\), called “cumulant parameter,” is such that \(\alpha \geq 1\). Hence, for a MS(\(\alpha\)) signal \(a(n)\), one easily has \(E[a] = 0, E[a^2] = 1\), and
\[
\begin{align*}
C_2[a] &= \alpha - 1 \\
C_4[a] &= \alpha^2 - \alpha - 2.
\end{align*}
\]

In Fig. 1, we plot the respective curves of \(C_2[a]\) (with a solid line) and \(C_4[a]\) (with a dashed line) w.r.t. the cumulant parameter \(\alpha\). For \(\alpha = 1\), we have a binary signal with a zero third-order cumulant and a fourth-order cumulant taking its minimum value, that is, \(-2\). For \(\alpha = 2\), the fourth-order cumulant is null, whereas the third-order one equals 1.
The performances of the algorithms are associated with an index/measure [14] defined on the global matrix $\mathbf{S} = (S_{i,j})$ according to

$$\tilde{\text{ind}}_\beta (\mathbf{S}) = \frac{1}{2} \left[ \sum_i \left( \sum_j \frac{|S_{i,j}|^\beta}{\max_\ell |S_{i,\ell}|^\beta} - 1 \right) + \sum_j \left( \sum_i \frac{|S_{i,j}|^\beta}{\max_\ell |S_{i,\ell}|^\beta} - 1 \right) \right] \quad (42)$$

where $\beta$ is a nonzero positive integer. Indeed, this non-negative index is zero if $\mathbf{S}$ satisfies (5), and a small value indicates the proximity to the desired solutions.

We consider two cases of sources.

1) Three synthetic sources: The first two sources $a_1(n)$ and $a_2(n)$ are MS($\alpha$) signals, whereas the third one $a_3(n)$ is a Gaussian i.i.d. signal. Concerning the mixture, we use the following matrix:

$$\mathbf{G} = \begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.8 & 1 & 0.9 \\ 0.8 & 0.8 & 1 \end{pmatrix}, \quad (43)$$

2) Two real speech signals [26] are plotted in Fig. 2.

In all the experiments, we plot both the mean and the standard deviation (STD) of the estimated index $\tilde{\text{ind}}_\beta (\mathbf{S})$ over 500 Monte Carlo runs. In the first above case, the runs are realized on the three synthetic sources, whereas in the second case, they are realized on the $(2, 2)$ mixing matrix whose components are chosen randomly with an uniform law in the interval $[-1, 1]$.

In the simulations, we first compare JADE (joint-diagonalization of only fourth-order cumulant matrices) with eJADE(3) (joint-diagonalization of only third-order cumulant matrices) and eJADE(3,4) (joint-diagonalization of both third- and fourth-order cumulant matrices). JADE is represented by a solid line with circle (calculated) points and eJADE(3) is represented by a dotted line with plus (calculated) points, whereas eJADE(3,4) is represented by a dashdot line with star (calculated) points.

Second, we make a first comparison between the performances of ICAj(3,4) and ICAs(3,4). ICAj(3,4) is represented by a dashed line with star (calculated) points, whereas ICAs(3,4) is represented by a dotted line with plus (calculated) points.

**Experiment 1—Index versus Cumulant Parameter $\alpha$:** For case 1 of sources, we plot the mean and STD of the estimated index as a function of the cumulant parameter $\alpha$. The data number is held constant to $N_d = 400$.

FIG. 3. Mean and STD of the estimated index of JADE, eJADE(3), and eJADE(3,4) w.r.t. the cumulant parameter $\alpha$. The number of data $N_d = 400$, and there is no noise.
Fig. 4. Mean and STD of the estimated index of JADE, eJADE(3), and eJADE(3,4) w.r.t. the cumulant parameter $\alpha$. The number of data $N_d = 400$, and the noise power $P_n = 0.02$.

Fig. 5. Mean and STD of the estimated index of ICAJ(3,4) and ICAS(3,4) w.r.t. the cumulant parameter $\alpha$. The number of data $N_d = 400$, and there is no noise.

Fig. 6. Mean and STD of the estimated index of JADE, eJADE(3), and eJADE(3,4) w.r.t. the data number when $\alpha = 1.7$. The lower curves correspond to a noise power $P_n = 0$, whereas the upper ones correspond to $P_n = 0.02$.

Fig. 7. Mean and STD of the estimated index of JADE, eJADE(3), and eJADE(3,4) w.r.t. the data number when $\alpha = 2.5$. The lower curves correspond to a noise power $P_n = 0$, whereas the upper ones correspond to $P_n = 0.02$.

performances of eJADE(3,4) compared with JADE are not degraded when the third-order statistical information are insignificant, that is, in the neighborhood of $\alpha = 1$.

In Fig. 5, we observe that the performances of ICAJ(3,4) are a little bit better than the ICAS(3,4) ones in the case of sources with negative fourth-order cumulants ($\alpha \leq 2$), whereas for positive fourth-order sources cumulants ($\alpha \geq 2$), the conclusion is reversed.

**Experiment 2—Index versus Data Number:** Now, for case 1 of sources, the cumulant parameter is held constant, and we plot the mean and STD of the estimated index as a function of the data number. In Fig. 6, we use $\alpha = 1.7$ (this case corresponds to a negative fourth-order cumulant), whereas in Fig. 7, we use $\alpha = 2.5$ (positive fourth-order cumulant). The lower (resp. the upper) curves correspond to $P_n = 0$ (resp. $P_n = 0.02$).

Figs. 6 and 7 display that the improvement in performance of the new algorithms eJADE(3) and eJADE(3,4) is mainly important when the number of observations is moderate, that is, all around $N_d = 200, 400$.

**Experiment 3—Real Speech Signals:** In this case 2 of sources, we plot the mean and STD of the estimated index as a function of the data number taken as the first $N_d$ samples of each speech signal.

Fig. 8 shows that the joint-diagonalization of only third-order cumulant matrices with eJADE(3) can be sufficient for the separation of speech signals with, however, lower performances. This is not surprising because the third- and fourth-order cumulants (estimated over the whole signals) of the speech signals we use are, respectively, around 0.5 and 2.7. That is, loosely speaking, the probability density functions of these speech signals are “more super-Gaussian than nonsymmetrical.” On the other hand, the performances of the algorithm eJADE(3,4) using both third- and fourth-order cumulant matrices are a little bit better.
Fig. 8. Mean and STD of the estimated index of JADE, eJADE(3), and eJADE(3,4) w.r.t. the first $N_d$ samples of the speech signals.

Fig. 9. Mean and STD of the estimated index of ICAj(3,4) and ICAs(3,4) w.r.t. the first $N_d$ samples of the speech signals.

Fig. 9 illustrates the behavior of ICAj(3,4) and ICAs(3,4) with the same speech signals. It shows that in the present case, ICAj(3,4) has better performance in comparison with ICAs(3,4).

VII. CONCLUSION

In this paper, we have essentially proposed a generalization of some results leading to the JADE and STOTD algorithms that were exclusively based on fourth-order cumulants. The main point is that we generalize the contrast functions implicitly maximized by JADE and STOTD and establish their links with a joint-diagonalization criterion that is the basis of these algorithms. Moreover, we show that one can joint-diagonalize tensors of cumulants of any order (greater than or equal to three) and/or joint-diagonalize tensors of cumulants of mixed orders.

On the way, we also extend the family of contrast functions involving cross-cumulants. This new family gives a common framework for considering altogether the ICA’s contrast, the JADE’s contrast, and the STOTD’s contrast. In particular, considering fourth-order cumulants, these derivations allow us to show that a recent function given in [3] and derived as an “approximation of the mutual information contrast” is a contrast. The new contrast family also allows us to develop some equivalents between them and joint-diagonalization criteria.

We also show how any order cumulants and/or cumulants of mixed orders can be considered simply for the joint-diagonalization of matrices or third-order tensors, leading to what is called “extended algorithms.”

Finally, computer simulations are presented both to illustrate the above results and to compare the performances of JADE with its extended new version eJADE and to compare ICAj and ICA. For this task, we first define a discrete random signal called MS$(\alpha)$ whose cumulants are parameterized. In all the presented cases, curves show that eJADE is more robust than JADE with respect both to the statistics of the sources and to the number of available data. Moreover, it seems that the relative performances of ICAj and ICAs depend on the sign of the fourth-order cumulants of the sources.

APPENDIX A

PROOF OF PROPOSITION 1

Let us first consider the function $G^2_{R,0}(y)$ denoted by $\mathcal{J}_R(y) \equiv G^2_{R,0}(y)$ and written as

$$\mathcal{J}_R(y) = \sum_{\ell_1, \ldots, \ell_{R-1}, \ell_1} \left( \text{Cum}[y_{\ell_1}, y_{\ell_1}, y_{\ell_2}, \ldots, y_{\ell_{R-1}}] \right)^2 .$$

With $\mathcal{S} = (S_{\ell_1, j})$, according to (6), the multilinearity of cumulants and the independence of sources, we have

$$\text{Cum}[y_{\ell_1}, y_{\ell_1}, y_{\ell_2}, \ldots, y_{\ell_{R-1}}] = \sum_{\ell} S^2_{\ell_2, \ell_2, \ell_2, \ldots, \ell_{R-1}, \ell_1} C_R[a_{\ell_1}] C_R[a_{\ell_2}] .$$

Hence

$$\mathcal{J}_R(y) = \sum_{\ell_1, \ldots, \ell_1} \sum_{\ell_2} \sum_{\ell_1, \ell_1, \ell_2, \ell_2, \ldots, \ell_{R-1}} S^2_{\ell_1, \ell_1, \ell_2, \ell_2, \ldots, \ell_{R-1}, \ell_1} C_R[a_{\ell_1}] C_R[a_{\ell_2}] .$$

Now, because $\mathcal{S}$ is a unitary matrix, then $\forall \ell_1, \ell_2, \sum_{m} S_{m, \ell_1, \ell_2} = \delta_{\ell_1, \ell_2}$, where $\delta_{i, j} = 1$ if $i = j$ and 0 otherwise. Thus, from (46) and because $\forall \ell_1, \sum_{\ell_1} S^2_{\ell_1, \ell_1, \ell_1, \ell_1} \leq \sum_{\ell_1} S^2_{\ell_1, \ell_1, \ell_1, \ell_1} = 1$, we then have

$$\mathcal{J}_R(y) = \sum_{\ell_1} \left( \sum_{\ell_1} S^2_{\ell_1, \ell_1, \ell_1, \ell_1} \right) \left( C_R[a_{\ell_1}] \right)^2 \leq \sum_{\ell_1} \left( C_R[a_{\ell_1}] \right)^2 = \mathcal{J}_R(\alpha).$$

It is easily shown that the equality in (47) holds if and only if $\mathcal{S}$ satisfies (5). Moreover, $\mathcal{J}_R(y)$ is invariant when $y_{\ell_1}(n), i =$
1, \ldots, N are multiplied by a factor of modulus 1. Thus, $J_R(y)$ is a contrast function.

Now, letting us consider $G_{R_1, R_2}(y)$, we easily have

$$G_{R_1, R_2}(y) \leq J_R(y) \leq J_R(a) = G_{R_1, R_2}(a). \tag{48}$$

The equality in the above inequalities holds iff $S$ is of the form (5); thus, $G_{R_1, R_2}(y)$ is a contrast.

**APPENDIX B**

**PROOF OF PROPOSITION 4**

Denoting $D = D(H, T_{R}^{R'}(O_{R-R'})^{T})$ and using (25), (26) and (29), we have

$$D = \sum_r \sum_{k_1} \cdots \sum_{m_{R_1}} \sum_{n_{R_1}} H_{i_1, m_1} \cdots H_{i_1, m_{R_1}} L_{m_1, \ldots, m_{R_1}, n_{R_1}}^{R_1}(r) \sum_{m_{R_2}} \sum_{n_{R_2}} H_{i_2, m_{R_2}} L_{m_{R_2}, \ldots, m_{R_1}, n_{R_1}}^{R_2}(r).$$

Now, using (28) and noting

$$\text{Cum}_k[m] = \text{Cum}[x_{k_1}, \ldots, x_{k_R}]$$

where $k = (k_1, \ldots, k_R)$ and $k = m$ or $n$, we have

$$D = \sum_r \sum_{k_1} \cdots \sum_{m_{R_1}} \sum_{n_{R_1}} H_{i_1, m_1} \cdots H_{i_1, m_{R_1}} \sum_{m_{R_2}} \sum_{n_{R_2}} H_{i_2, m_{R_2}} L_{m_{R_2}, \ldots, m_{R_1}, n_{R_1}}^{R_2}(r) \cdot \text{Cum}_k[m] \cdot \text{Cum}_k[n] \cdot \text{Cum}_k[m] \cdot \text{Cum}_k[n]. \tag{49}$$

On the other hand, we have

$$\text{Cum}^{R_1}_{R_2}(y) = \sum_{m_{R_1}} H_{i_1, m_{R_1}} \sum_{n_{R_1}} H_{i_1, n_{R_1}} \sum_{m_{R_2}} H_{i_2, m_{R_2}} L_{m_{R_2}, \ldots, m_{R_1}, n_{R_1}}^{R_2}(r) \cdot \text{Cum}_k[m] \cdot \text{Cum}_k[n]. \tag{50}$$

Finally, comparing (49) and (50), we have the proposed result

$$D(H, T_{R}^{R'}(O_{R-R'})^{T}) = \text{Cum}^{R_1}_{R_2}(Hx). \tag{51}$$

**APPENDIX C**

**PROOF OF PROPOSITION 5**

Let $1^j$ denote the $(N, 1)$ vector with 1 in the $i$th position and 0 elsewhere, i.e., $1^j_i = \delta_{i,j}$. Let $O_{R}^{R'}(m_{R_1+1}, \ldots, m_{R_R}) = (1^{n_{R_1+1}} \cdots 1^{m_{R_R}})(m_{R_1+1} \cdots m_{R_R}) \in \{1, \ldots, N\}^{R-R_1}$ be $N^{R-R_1}$ tensors of order $R - R_1$. Denoting

$$R_m = (R_{x}^{R}(O_{c}^{R-R'})^{T})_{m_{R_1+1}, \ldots, m_{R_R}}$$

we then have

$$R_m = \sum_{m_{R_1+1}, \ldots, m_{R_R}} \text{Cum}(x_{m_{R_1+1}}, \ldots, x_{m_{R_1}+1}, \ldots, x_{n_{R_1}+1}) \cdot \text{Cum}_{R_{x}^{R}(O_{c}^{R-R'})}.$$ \tag{52}

Thus, $T_{R}^{R'}(O_{c}^{R-R'})$ clearly form an orthonormal set, then direct application of Proposition 4 yields

$$D(H, T_{R}^{R'}(O_{c}^{R-R'})) = \text{Cum}^{R_1}_{R_2}(Hx).$$

**ACKNOWLEDGMENT**

The author would like to acknowledge J.-F. Cardoso for stimulating discussions and for the preprint of paper [3]. He also wishes to thank L. De Lathauwer for his very interesting Ph.D. dissertation. Special thanks are due to N. Thirion-Moreau for insightful comments. Finally, he is grateful to the useful comments given by the reviewers.

**REFERENCES**


MOREAU: GENERALIZATION OF JOINT-DIAGONALIZATION CRITERIA 541


Eric Moreau (S’93–M’96) was born in Lille, France. He graduated from the Ecole Nationale Supérieure des Arts et Métiers (ENSAM), Paris, France, in 1989 and received the Agrégation de Physique degree from the Ecole Normale Supérieure de Cachan, Cachan, France, in 1990. He received the DEA degree in 1991 and the Ph.D. degree in 1995, both in the field of signal processing, from the University of Paris-Sud. Since 1995, he has been with the Système, Information et Signal Laboratory (SIS), Institut des Sciences de l’Ingénieur de Toulon et du Var (ISITV), La Valette, France, as an Assistant Professor. His main research interests are in statistical signal processing using high-order statistics.