TENSOR DIAGONALIZATION, A USEFUL TOOL IN SIGNAL PROCESSING

Pierre COMON

Thomson-Sintra, B.P.157, F-06903 Sophia-Antipolis cedex, France
I3S – CNRS, Sophia-Antipolis, F-06560 Valbonne, France, comon@alto.unice.fr

ABSTRACT

Tensors appear more and more often in signal processing problems, and especially spatial processing, which typically involves multichannel modeling. Even if it is not always obvious that tensor algebra is the best framework to address a problem, there are cases where no choice is left. Blind identification of multichannel non monic MA models is given as an illustrating example of this claim.

Key words. Tensor, High-Order Statistics, Identification, Non monic MA model, Array processing, Statistical independence, Information, Parallel algorithms.

1. INTRODUCTION

Many techniques in signal processing are based on the knowledge of the probability density function (pdf) of observations, possibly conditionnally to some unknown parameters. One may think for instance of the likelihood function. The Gaussian approximation has been assumed for a long time because (i) second-order moments were sufficient to solve most problems of interest, and (ii) moments of higher order were computationally heavy to handle. Today several problems have been pointed out that are not solvable under the Gaussian approximation, and in addition, the increase in computational power allows the use of more fancy approximations with the help of Higher-Order Statistics (HOS). The survey paper [15] points out some advantages in using HOS in signal processing and automatic control. See also the indroductory paper [9].

1.1. Tensors

A tensor of order r is an object defined in a N-dimensional coordinates system by a table with r indices, $g_{i_1,..i_r}, 1 \leq i_k \leq N$, that follows a particular transformation formula if the coordinates system is changed. If the system of coordinates is changed so that any vector u is transformed into a vector $U = Au$, where A is a $N \times N$ matrix, then the tensor is transformed into:

$$g_{i_1,..i_r} \to G_{i_1,..i_r} = \sum_{j_1,..j_r} A_{i_1j_1}..A_{i_rj_r} g_{j_1,..j_r} \qquad (1)$$

This property is often referred to as the *multilinearity property* of tensors. Such a tensor will be referred to as a $(N;r)$-tensor, in short.

1.2. Cumulants

Expansions of pdf about some given family of densities lead to objects that are called cumulants. See for instance [11] [12] for a general framework on density expansions. In general, cumulants (by default) are associated with expansions about the Gaussian pdf. If a random variable of dimension N admits finite moments up to order r, then its cumulants of order r exist. In other words, define the second characteristic function of X as $\Psi_x(v) = \log E\{\, e^{\jmath v^T X}\,\}$, where $\jmath = \sqrt{-1}$; this function always exists in a neighborhood of the origin. Then cumulants are coefficients of $\jmath^r\, v_1..v_r$ in the Taylor expansion of $\Psi_x(v)$ about the origin. An alternative way is to define them as a function of moments [3]:

$$\mathcal{C}\{X_1,..X_r\} = \sum (-1)^{p-1}\,(p-1)!\,\left[\prod_{i=1}^{p} E\{\prod_{j\in\nu_i} X_j\}\right], \qquad (2)$$

where the summation extends over all partitions $(\nu_1,..\nu_p)$ of $(1,..r)$, $1 \leq p \leq r$. Simple expressions can be derived for cumulants of moderate orders [14]. Such cumulants will be denoted in short $\mathcal{C}_{i_1,..i_r}\{X\}$ in the present paper, where indices i_k may be distinct or not.

Due to the fact that $\Psi_{AX}(v) = \Psi_X(A^\dagger v)$, cumulants inherit the multilinearity property and may be considered as tensors [14]. For any random variable X with values in $\mathbb{R}^N$, the indices $i_1,..i_r$ can be permuted in any manner without changing the value of the cumulants $\mathcal{C}_{i_1,..i_r}\{X\}$. Therefore, cumulants are *symmetric* tensors.

In practice, it is more convenient to work with so-called standardized data, because of numerical conditioning, among other reasons. For scalar random variables, this operation merely reduces to centering and dividing by the standard deviation. In the vector case, the standardization operation consists of centering, reducing the data by projection onto its range space, and rotating the subspace onto the eigenvectors coordinates system.

More precisely, assume T realizations $x(t)$ of a random variable X are available in the form of a $N \times T$ data matrix, $x(1:T)$. Denote the SVD of the centered data as $x(1:T) - mean(x) = U\Sigma V^T$, where U is $N \times \rho$ full column rank, $\rho \leq N$, and Σ is square invertible. Then standardized data are defined by $\tilde{x}(1:T) = \sqrt{T}\,V^T$, and have zero mean and unit covariance. Cumulants of standardized variables are called *standardized cumulants*.

High-Order Statistics (HOS) are more and more widely used in various areas including signal processing and automatic control, as shows the present literature for the last few years. In situations where variables are multidimensional, it is useful to point out that cumulants may be considered as tensors. Nevertheless, very few tools are at disposal to manipulate tensors; if we may resort to Cholesky or Eigen-Value decompositions in case of symmetric matrices, there are unfortunately no such decompositions available for completely symmetric tensors. Some directions are proposed in this paper in sections 3 and 4.

2. APPLICATIONS

2.1. Non monic MA models

As a working example, consider the non monic multichannel MA model:

$$y(t) = \sum_{k=0}^{q} B_k\, w(t-k) + v(t), \qquad (3)$$

where the order q is assumed known, $v(t)$ is a nuisance noise, $w(t)$ is a spatially and temporally white noise, i.e., $w_i(t)$ and $w_j(t')$ are statistically independent, $\forall i, j \in \{1, ..N\}$, $i \neq j$, and $\forall t, t'$, $t \neq t'$. Assume all matrices B_k are unknown, $0 \leq k \leq q$, and matrix B_0 is invertible.

The problem of identifying the matrices B_k from the observations of system outputs only can be easily decomposed into two subproblems, as briefly recalled below. Just consider the (temporally) white process $x(t) = B_0 w(t)$, and denote $\bar{B}_k = B_k B_0^{-1}$, for $k > 0$. Then the model can obviously be rewritten as:

$$y(t) = \sum_{k=0}^{q} \bar{B}_k\, x(t-k) + v(t),\ x(t) = B_0 w(t). \qquad (4)$$

This model is monic, since $\bar{B}_0$ is now the identity matrix. Many algorithms have been proposed to date for the so-called *blind identification* of monic MA models. See for instance [19] [6] [18]. To give the flavour of these approaches in a few lines, consider the scalar case (N=1) and suppose it is wished to use only fourth order output cumulants. The idea is to remark that output cumulants are linked to each other through the linear system:

$$\mathcal{C}_{0ikq}\{y\}\, B_j = \mathcal{C}_{0jkq}\{y\}\, B_i, \qquad (5)$$

for any triplet of indices (i, j, k) such that $0 \leq i < j \leq q$ and $0 \leq k \leq q$. The coefficients $B_i, i > 0$, can thus be obtained just by solving an overdetermined $q(q+1)^2/2$ by q linear system. The idea extends to the multichannel case to the price of some complication in the notation [6], that we do not want to introduce here.

Of course, there are many other approaches to monic MA blind identification [15]. However, our goal is not to focus on this well known problem, but to spend some time on the identification of the first matrix coefficient, B_0.

2.2. Narrow band array processing

There are many cases in spatial signal processing where it is sought to obtain a diagonal tensor after linear transforms, e.g., seismics, interception problems in Sonar, blind estimation of radiating sources [5] [7] [17] [18], even if it is not stated explicitly in those terms. In fact, the problem in antenna array processing is to recover source signals that are statistically independent. Generally, a linear model is assumed, so that the problem amounts to identifying a linear system with mutually independent inputs.

First, independent sources should at least have a diagonal covariance matrix. But there are infinitely many congruent transforms that diagonalize a given covariance matrix; additional constraints are thus necessary to determine the transform uniquely. Bienvenu and Kopp [1], and independently Schmidt [16], proposed in the so-called MUSIC algorithm to fix this indetermination by using the knowledge of the array: the *direction of arrival* vectors should lie on some *array manifold* in the absence of noise.

Clearly, if the array manifold is not available, the MUSIC approach cannot succeed, and it is necessary to resort to HOS. The idea is now that independent sources should also have diagonal cumulant tensors, for all orders. By seeking for the linear transform that diagonalizes one or several tensors, one may identify the linear model, and eventually the source signals themselves. Nevertheless, it is not possible to recover the *directions of arrival* without some knowledge on the array manifold. For reasons of space, the application of tensor diagonalization to array processing is not developed further in this paper.

3. DIAGONALIZATION ISSUES

Consider in this section the following algebraic problem. Given a tensor $t_{ij..k}$ of order r, $2 \leq r \leq 4$, is it possible to find a linear change of coordinates defined by an invertible matrix, A, such that the tensor takes a diagonal form ?

A necessary condition is that the number of free parameters be preserved. A symmetric tensor has $\binom{r}{N+r-1}$ independent parameters. In the matrix case, we thus have $N(N+1)/2$ parameters, which is smaller than the N^2 entries of A. But in the tensor case, it is clear that the number of free parameters would decrease: $O(N^r)$ in the original symmetric tensor, and N^2 in the decomposition. This immediate statement proves that only a small subset of symmetric tensors of order larger than 2 is (linearly) diagonalizable.

On the other hand, this statement does not prove anything if the transform A is allowed to map the N-dimensional space into a space of possibly larger dimension, say $P \geq N$. The decomposition can now be written as:

$$t_{ij..k} = \sum_{p=1}^{P} A_{ip} A_{jp} .. A_{kp} T_p, \qquad (6)$$

where T_p is a diagonal $(P; r)$-tensor, and A is a $N \times P$ matrix. The number of free parameters in decomposition (6) is now NP. In that case the question is to know (1) how to choose P, and (2) how to compute the P rows of the transform A. Thus there are two different approaches: one can look for an *exact* diagonal tensor decomposition, but in a space of larger dimension, or conversely look for an approximate tensor diagonalization in a space of same dimension. These two aspects are now commented.

3.1. Exact diagonalization

There is an obvious bijective relation between the set of symmetric $(N; r)$-tensors, $t_{ij..k}$, and the set of homogeneous polynomials of degree r in N variables. In fact, associate each tensor t with the polynomial $\pi(x, y, ..z) = \sum_i \sum_j .. \sum_k t_{ij..k}\, x_i y_j .. z_k$. This equivalence has been already utilized in many eighteenth century works.

Consequently, looking for a $N \times P$ matrix that maps the $(N; r)$-tensor into a diagonal $(P; r)$-tensor is equivalent

to looking for P linear forms such that the polynomial π is mapped to a sum of r-th powers of linear forms. The exact diagonalization problem exists in the theory of homogeneous polynomial forms, but very few results are known today, despite the long history of the subject.
In the case of binary forms (N=2) however, it is known that every binary form of even degree $2m$ can be put as a sum of m perfect powers, provided the determinant of some Hankel matrix built with the polynomial coefficients is null [13]. This determinant is often referred to as the Catalecticant. Moreover, a general binary form of degree $2m$ can be written as a sum of $m+1$ perfect powers. For instance, a binary quartic can still be expressed as a sum of 3 perfect powers, even if its catalecticant is non zero. There are also a number of results in the case of forms of odd degree. For instance, a theorem attributed to Sylvester claims that a binary form of degree $2m+1$ can be written as a sum of $m+1$ perfect powers of linear forms. These decompositions are generally not unique.
Now regarding forms with more than 2 variables, the results become much more confusing. In fact, it is not possible to state systematically what is the minimal mumber of forms, P, as a function of the dimension N and the degree r. If P is chosen just in order to satisfy the inequality $NP \geq \binom{r}{N+r-1}$, then P turns out to be too small in some cases. Actually, every case is particular. Additionally, assuming P is known, the calculation of the decomposition itself is very difficult to carry out. That's why this discussion is deferred to a companion paper [8], where the case of cubics ($r=3$) and quartics ($r=4$) will be mainly addressed.

3.2. Approximate diagonalization

In this section, only invertible transforms are considered (i.e. $P=N$). As pointed out previously, tensors of order r and dimension N cannot generically be exactly diagonalized. Thus, there is first a need to define in what respect the *approximation* will be understood. Carl Jacobi introduced in 1846 a criterion dedicated to matrices. In order to diagonalize a symmetric matrix by an orthogonal change of coordinates, he proposed to minimize the sum of squares of non-diagonal entries. Denote Q the matrix defining the transformation, and g a given symmetric matrix. Since the Froebenius norm of the transformed matrix, $G = QgQ^\dagger$ remains the same, one can alternatively maximize the sum of squares of diagonal entries. This defines an optimization criterion for second order symmetric tensors:

$$\Upsilon_2(Q;g) = \sum_{n=1}^{N} G_{nn}^2. \tag{7}$$

For tensors of higher order, we define the following criteria, that will find steady justifications in the next section:

$$\Upsilon_3(Q;g) = \sum_{n=1}^{N} G_{nnn}^2,\ \Upsilon_4(Q;g) = \sum_{n=1}^{N} G_{nnnn}^2. \tag{8}$$

It must be kept in mind that, because of the multilinearity property, these criteria are implicit functions of the matrix Q:

$$G_{ijk} = \sum_{mno} Q_{im}Q_{jn}Q_{ko}g_{mno}, \tag{9}$$

$$G_{hijk} = \sum_{lmno} Q_{hl}Q_{im}Q_{jn}Q_{ko}g_{lmno}. \tag{10}$$

In the remaining sections, the optimization criteria will be justified, and the approximate diagonalization will be described in detail.

4. BLIND IDENTIFICATION OF STATIC LINEAR SYSTEMS

4.1. Notation

Given realizations $y(t)$ of a random vector Y with values in $\mathbb{R}^N$, it is desired to estimate a matrix F such that $Y = FZ$, where Z is a random vector whose composents are statistically independent. In this framework, only independence up to order 3 or 4 will be required (that is, cross cumulants of components of Z are null at orders 2 and 3 or 4). Standardization already insures independence at order 2 (the covariance matrix is identity). Yet, standardized data are defined up to a multiplicative orthogonal matrix (in addition to scale and permutation transforms already pointed out). So there is clearly some degree of freedom left to improve statistical independence.
It can be sought for an orthogonal matrix Q such that higher-order correlations of $Z = QY$ are minimized (since the transform is now invertible, we also have

$$Y = Q^T Z \tag{11}$$

equivalently). For this purpose, define the third and fourth order sample cumulants of Z [11, 12]:

$$G_{ijk} = \frac{1}{T}\sum_{t=1}^{T} z_i(t)\,z_j(t)\,z_k(t), \tag{12}$$

$$G_{ijkl} = \frac{1}{T}\sum_{t=1}^{T} z_i(t)\,z_j(t)\,z_k(t)\,z_l(t) - \delta_{ij}\delta_{kl} - \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}, \tag{13}$$

where δ is the Kronecker symbol, and $z(t)$ realizations of Z. Denote g_{ijk} and g_{ijkl} the corresponding cumulants of Y defined in the same manner.

4.2. Definition of ICA

Let Y be a random vector with values in $\mathbb{R}^N$ admitting a probability density $p_Y(u)$, and assume T realizations $y(t)$ are observed (index t does not necessarily refer to time). The ICA of Y consists of searching for a $N\times\rho$ matrix, F, $N \geq \rho$, that minimizes the *statistical dependence* between the components of the $\rho \times 1$ vector, Z, defined by

$$Y = F\,Z, \tag{14}$$

in the sense of the maximization of a *contrast function*:

Definition 4.1 *Let $\mathcal{E}_N$ be the space of real random variables of dimension N admitting a density. A contrast is a mapping Ψ from the set of densities $\{p_Z, Z \in \mathcal{E}_N\}$ to $\mathbb{R}$ satisfying the 3 requirements:*

1. *$\Psi(p_Z)$ does not change if the components Z_i are permuted: $\Psi(p_{PZ}) = \Psi(p_Z)$, $\forall P$ permutation;*
2. *Ψ is invariant by scale change: $\Psi(p_{\Lambda Z}) = \Psi(p_Z)$, $\forall\Lambda$ diagonal invertible;*

3. *Ψ decreases by linear combination: if X has independent components, then $\Psi(p_{AX}) \leq \Psi(p_X)$, $\forall A$ invertible.*

For the sake of clarity, let us also give the definition below:

Definition 4.2 *The $\Psi-ICA$ of a random vector Y of size N with finite covariance matrix V_Y is the pair of matrices (F, Δ) such that: (i) matrix Δ is diagonal positive and F is $N \times \rho$ full column rank, (ii) the covariance of Y factorizes as $V_Y = F \Delta^2 F^T$, (iii) vector Y writes as $Y = F\,Z$, where Z is a $\rho \times 1$ random vector, $\rho \leq N$, with covariance Δ^2 and whose density maximizes a given contrast function, Ψ.*

If requirement (iii) in the definition above is replaced by $F^T F = I$, then we get the definition of Principal Component Analysis (PCA). Other links with PCA are pointed out in section 4.5.
Conversely, consider the static linear system:

$$Y = M\,X + W, \tag{15}$$

where vector X is the input vector and has independent components, and W stands for various nuisances (e.g. measurement noise). Then ICA can be viewed as a means to obtain estimates Z_n of the input components X_n, when only realizations of the output, $y(t) = Mx(t) + w(t)$, are observed. Note that the (only) key assumption that may be used in this problem is the statistical independence of the components X_n; for instance, statistics of the noise W are unknown.

4.3. Identifiability

The first property that can be noticed, is that if a pair (F, Δ) is the $\Psi-$ICA of Y, then so is any pair of the form $(F\Lambda DP, P^T\Lambda^{-1}\Delta P)$, where Λ is a $\rho \times \rho$ invertible real positive diagonal scaling matrix, D is a $\rho \times \rho$ diagonal matrix with entries of unit modulus, and P is a $\rho \times \rho$ permutation. In other words, *as is the case for PCA*, solutions need to be considered modulo this equivalence relation.

Definition 4.3 *A contrast Ψ will be said discriminating over a set $\mathcal{E}$ if the equality $\Psi(p_{AX}) = \Psi(p_X)$ holds only when A is of the form ΛP, as soon as X is a random vector of $\mathcal{E}$ having independent components.*

Then we have the following identifiability theorem:

Proposition 4.4 *Let no noise be present in model (15), and define $Y = M\,X$ and $Y = F\,Z$, X being in some set $\mathcal{E}$ of $\mathcal{E}_N$, and having independent components and an invertible covariance. Then if Ψ is discriminating over $\mathcal{E}$, $\Psi(p_Z) = \Psi(p_X)$ if and only if $F = M\Lambda P$, where Λ is invertible diagonal and P is a permutation.*

In other words, the matrix M can be uniquely estimated by matrix F modulo the above mentioned equivalence relation. The identifiability theorem 4.4 needs some space to be proved, and it is referred to [7] for more details. Related results can also be found in [10].
Reminding that N random variables X_n are independent if and only if their joint pdf $p_X(u)$ is equal to the product of the marginal pdf's, $\prod_{n=1}^N p_{X_n}(u_n)$, a quite natural measure of independence is the distance between two such quantities. If the Kullback divergence is taken as a distance measure [2], then we obtain the average mutual information:

$$I(p_Z) = \int p_Z(u)\,\log\,\frac{p_Z(u)}{\prod_{n=1}^N p_{Z_n}(u_n)}\,du. \tag{16}$$

It can be shown that $\Psi_0 \stackrel{\text{def}}{=} -I(p_Z)$ is a contrast, and is discriminating over the set of random variables having at most one Gaussian component [7].

Proposition 4.5 *With the definitions given in section 1, it can be shown [7] that the functionals Ψ_r below are also contrasts, and are discriminating over the subset of $\mathcal{E}_N$ of random vectors with finite moments up to order r and with at most one component with null cumulant of order r:*

$$\Psi_3(p_Z) = \sum_i G_{iii}^2 \quad ; \quad \Psi_4(p_Z) = \sum_i G_{iiii}^2. \tag{17}$$

This last property gives steady foundations to the criteria suggested in section 3.

4.4. Numerical aspects

Since any orthogonal matrix can be decomposed into a product of $N(N-1)/2$ plane rotations and a diagonal matrix with entries of unit modulus, it seems natural to look first at the case of plane rotations. It turns out that in dimension $N = 2$, explicit expressions can be given for contrast functions and their maxima. Denote θ the tangent of the rotation angle of Q, and $\xi = \theta - 1/\theta$; then these contrasts are simple rational functions:

$$\psi_3(\theta; g) = (\theta + \frac{1}{\theta})^{-3} \sum_{i=1}^{3} a_i\,\left(\theta^i - (-\theta)^{-i}\right) \tag{18}$$

$$\psi_4(\xi; g) = (\xi^2 + 4)^{-2} \sum_{i=0}^{4} b_i\,\xi^i. \tag{19}$$

where coefficients a_i and b_i are given in appendix. If $N = 2$, it is consequently easy to find the absolute maximum of ψ_r. In fact, it can be shown that non trivial stationary points of ψ_r are given by the roots of polynomials of the form:

$$\omega_3(\xi; g) = d_2\,\xi^2 + d_1\,\xi - 4\,d_2, \quad \omega_4(\xi; g) = \sum_{i=0}^{4} c_i\,\xi^i. \tag{20}$$

Coefficients c_i and d_j are given in appendix.

4.5. The CM algorithm

Assume standardized data are available, $y_n(t), 1 \leq n \leq N, 1 \leq t \leq T$. The algorithm proposed proceeds (pairwise) as follows, for r equals either 3 or 4:

1. Initialize $F = U\,\Sigma\,/\,\sqrt{T}$, as defined in section 1, and $z(1:T) = y(1:T)$;

2. For *sweep* = 1 to S,

3. Sweep all the pairs (i, j) in a prescribed ordering, and for each pair, do:

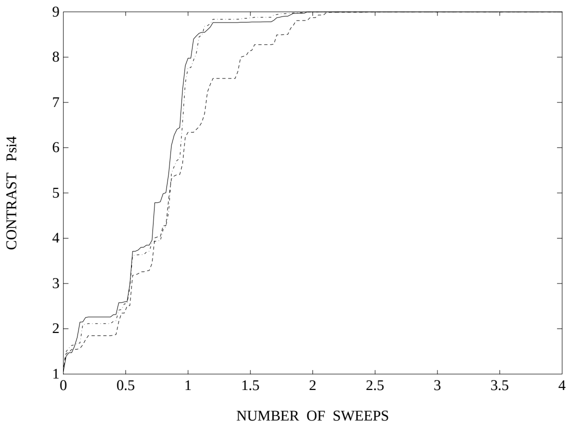

Figure 1: Variation of contrast ψ_4 for 3 different orderings and with input kurtosis [1 1 1 1 1 0 -1 -1 -1 -1].

(a) Compute the $r+1$ cumulants of order r of the pair (z_i, z_j), denoted here g, with the help of expressions similar to (12) or (13);

(b) Root the polynomial of degree r, $\omega_r(\xi)$, and retain the value of ξ yielding the largest $\psi_r(\xi)$;

(c) Compute the plane rotation $Q^{(i,j)}$ acting in the plane (i, j), with tangent θ defined by the root of: $\theta^2 - \xi\theta - 1 = 0$ in the interval $(-1,\ 1]$;

(d) Apply the plane rotation to rows i and j of $z(t)$ for every t: $z(t) \leftarrow Q^{(i,j)} z(t)$;

(e) Accumulate the transform: $F \leftarrow FQ^{(i,j)T}$;

4. Stop if $sweep = S$, or if all estimated angles have been small in the last sweep.

It is reasonable to take $S = 1 + floor(N^\alpha), \alpha \leq 1/2$. It can be checked out that the most computationally heavy step is 3a; the complete algorithm requires approximately $O(\frac{11}{2}\, S\, N^2\, T)$ operations for $r = 3$, and $O(6\, S\, N^2\, T)$ operations if $r = 4$. For instance, if $T = O(N^{3/2})$ and $S = O(\sqrt{N})$, the complexity is of order N^4 for $r = 4$. But much larger values of T can be envisaged.

Computer results

Simulations presented now were run with $N = 10$, the mixing matrix M was defined by $M_{ii} = 1$ and $M_{ij} = 2$ elsewhere. The contrast Ψ_4 was used, and fourth order input cumulants were those given in the figure captions. Figures 1 and 2 give the behaviour of contrast Ψ_4 as more and more pairs are processed. The 3 particular orderings tested have been performed by just swapping input cumulants, and the same cyclic ordering was performed afterwards. As expected, the speed of convergence depends on the ordering, but not the limit reached in these examples.

Limitations

The Jacobi algorithm was originally dedicated to the diagonalization of symmetric matrices by orthogonal change of coordinates. More precisely, given a matrix g with components g_{ij}, at each step of the algorithm, it is sought for an orthogonal matrix Q such that the criterion ψ_2 is maximized:

$$\psi_2(G) = \sum_i G_{ii}^2;\ G_{ij} = \sum_{p,q} Q_{ip} Q_{jq} g_{pq}. \tag{21}$$

Stationary points of ψ_2 can be shown to satisfy $G_{qq}G_{qr} = G_{rr}G_{qr}$, for any pair of indices $(q,r), q \neq r$. Next, the relation $d^2\psi_2 < 0 \Leftrightarrow G_{qr}^2 < (G_{qq} - G_{rr})^2$ proves (in an original and elegant manner) that the only maximum corresponds to $G_{qr} = 0$, whereas $G_{qq} = G_{rr}$ corresponds to a minimum. Other stationary points are then saddle points.
Similarly, one can look at relations characterizing local maxima of criteria Ψ_3 and Ψ_4 :

$$\begin{aligned} G_{qqq}G_{qqr} - G_{rrr}G_{qrr} &= 0 && (22)\\ 4G_{qqr}^2 + 4G_{qrr}^2 - (G_{qqq} - G_{qrr})^2 - (G_{rrr} - G_{qqr})^2 &< 0 && (23)\\ G_{qqqq}G_{qqqr} - G_{rrrr}G_{qrrr} &= 0 && (24)\\ 4G_{qqqr}^2 + 4G_{qrrr}^2 - (G_{qqqq} - \tfrac{3}{2}G_{qqrr})^2 & && \\ -(G_{rrrr} - \tfrac{3}{2}G_{qqrr})^2 &< 0 && (25) \end{aligned}$$

for any pair of indices $(p,q), p \neq q$. As a conclusion, contrary to Ψ_2 in the matrix case, Ψ_r might have theoretically spurious local maxima in the tensor case, $r > 2$, even if this has never been observed (see also computer experiments carried out in presence of noise [7]).

5. CONCLUDING REMARKS

As tensors are more and more utilized in various areas, including signal processing and control, it would be useful to know what are the decompositions at disposal for completely symmetric tensors (e.g. Cholesky, Eigenvalue...) and what algorithms can be resorted to for their effective computation. The problem of tensor diagonalization can be addressed in various manners. It is clear that only a small subset of symmetric tensors can be exactly diagonalized by orthogonal change of coordinates. In fact, the number of free parameters is much smaller in a diagonalizable tensor. In particular, a diagonalizable tensor of order 4 and dimension 2 satisfies $(G_{1112} - G_{1222})^2 - G_{1122}(G_{1111} + G_{2222}) + 2G_{1122}^2 = 0$. More generally, entries of a (linearly) diagonalizable tensor lie on a N^2-dimensional manifold, as already pointed out.

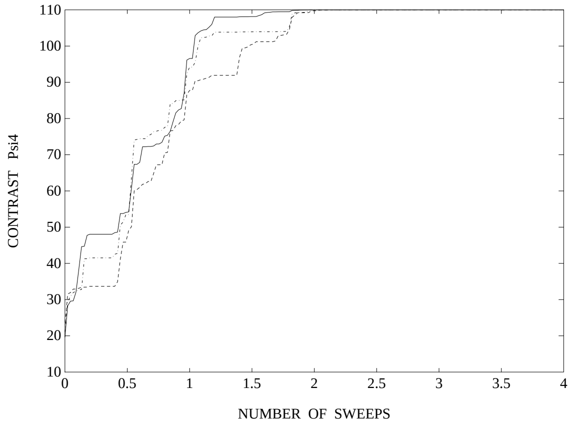

Figure 2: Variation of contrast ψ_4 for 3 different orderings and with input kurtosis [5 4 3 2 1 -1 -2 -3 -4 -5].

The algorithm proposed with $r = 3$ or 4, and $N \geq 2$, has been proved to converge to the absolute maximum of the criteria only for $N = 2$. Thus convergence needs to be studied more thoroughly. Besides this key remark, other issues that have been left aside for the moment include: (i) speed and memory issues, that can be probably addressed by designing appropriate sweeping strategies, (ii) the possibility of carrying out the diagonalization of a symmetric tensor without calculating its entries explicitly, as is done by the SVD for covariance matrices (an effort in this direction was made in the algorithm proposed in section 4.5), (iii) the exploitation of possible structures in tensors (e.g. sparse, banded, Toeplitz...).

A suboptimal algorithm has been proposed by Cardoso (see [5] or the proceedings of SPIE in 1990 pages 361–372), and considers tensors of order 4 and dimension N as linear operators acting on matrices of size N^2. The computation of the EVD of such operators gives a means to compute the ICA by resorting to standard reliable codes. But it has the inconvenience to break symmetry, and to be applicable only for even orders. In addition, an approximation must be made when the diagonalization is not exact, and and the approach then also lacks in convergence proof [4].

The principles used for handling matrices are not as effective as expected for handling tensors. Specific tools dedicated to tensors remain to be developed.

6. APPENDIX

Coefficients a_i, b_j, c_i and d_j are polynomial functions of the standardized cumulants of the observations, g_{ijk} and g_{ijkl}. For Ψ_3 and ω_3:

$$\begin{aligned}
a_3 &= g_{111}^2 + g_{222}^2, \\
a_2 &= 6\,(g_{122}\,g_{222} - g_{111}\,g_{112}), \\
a_1 &= 9\,(g_{122}^2 + g_{112}^2) + 6\,(g_{112}\,g_{222} + g_{111}\,g_{122}); \\
d_2 &= a_2/6 \;=\; g_{122}\,g_{222} - g_{111}\,g_{112}, \\
d_1 &= a_1/3 - a_3.
\end{aligned}$$

Next for Ψ_4 and ω_4, it is useful to define:

$$\begin{aligned}
t &= 16\,(g_{1112}^2 + g_{1222}^2), \\
u &= g_{1111} + g_{2222} - 6\,g_{1122}, \\
v &= 4\,(g_{1222} - g_{1112}), \\
w &= 6\,g_{1122}\,(g_{1111} + g_{2222}).
\end{aligned}$$

Then :

$$\begin{aligned}
b_4 &= g_{1111}^2 + g_{2222}^2, \\
b_3 &= -8\,(g_{1111}\,g_{1112} - g_{1222}\,g_{2222}), \\
b_2 &= 4\,b_4 + t + 2\,w, \\
b_1 &= 4\,b_3 - 2\,u\,v, \\
b_0 &= 2\,(b_4 + t + 2\,w + 36\,g_{1122}^2 + 2\,g_{1111}\,g_{2222} \\
&\quad + 32\,g_{1112}\,g_{1222}); \\
c_4 &= -b_3/8 \;=\; g_{1111}\,g_{1112} - g_{2222}\,g_{1222}, \\
c_3 &= 2\,b_4 - b_2/4 \;=\; b_4 - (t + 2\,w)/4, \\
c_2 &= 3\,b_3/2 - 3\,b_1/8 \;=\; 3\,u\,v/4, \\
c_1 &= b_2 - b_0/2, \\
c_0 &= b_1/2 \;=\; 2\,b_3 - \,u\,v.
\end{aligned}$$

7. REFERENCES

[1] G. BIENVENU and L. KOPP. Optimality of high-resolution array processing using the eigensystem approach. *IEEE Trans. ASSP*, 31(5):1235–1248, October 1983.

[2] R. E. BLAHUT. *Principles and Practice of Information Theory.* Addison-Wesley, 1987.

[3] D. R. BRILLINGER. *Time Series, Data Analysis and Theory.* Holden-Day, 1981.

[4] J. F. CARDOSO and A. SOULOUMIAC. Blind beamforming for non-Gaussian signals. *IEE Proceedings - Part F*, 140(6):362–370, December 1993. Special issue on Applications of High-Order Statistics.

[5] J.F. CARDOSO. Fourth-order cumulant structure forcing, application to blind array processing. In *Proc. IEEE SP Workshop on SSAP*, pages 136–139, Victoria, October 1992.

[6] P. COMON. MA identification using fourth order cumulants. *Signal Processing, Eurasip*, 26(3):381–388, 1992.

[7] P. COMON. Independent Component Analysis, a new concept ? *Signal Processing, Eurasip*, 36(3), 1994.

[8] P. COMON and B. MOURRAIN. A generalization of sylvester's theorem to quantics. In *SPIE conference on Advanced Signal Processing*, San Diego, July 24–29 1994.

[9] G. B. GIANNAKIS. Cumulants: A powerful tool in signal processing. *Proceedings of The IEEE*, 75(9):1333–1334, September 1987.

[10] A. M. KAGAN, Y. V. LINNIK, and C.R. RAO. *Characterization Problems in Mathematical Statistics.* Wiley, 1973.

[11] M. KENDALL and A. STUART. *The Advanced Theory of Statistics.* C. Griffin, 1977.

[12] S. KOTZ and N. L. JOHNSON. *Encyclopedia of Statistical Sciences.* Wiley, 1982.

[13] J. KUNG and G. ROTA. The invariant theory of binary forms. *Bulletin of the American Mathematical Association*, pages 27–85, 1984.

[14] P. McCULLAGH. *Tensor Methods in Statistics.* Monographs on Statistics and Applied Probability. Chapman and Hall, 1987.

[15] C. NIKIAS and J. MENDEL. Signal processing with higher-order spectra. *IEEE Signal Processing Magazine*, pages 10–37, July 1993.

[16] R. O. SCHMIDT. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antenna Propagation*, 34(3):276–280, March 1986.

[17] S. SHAMSUNDER and G. B. GIANNAKIS. Modeling of non-Gaussian array data using cumulants: DOA estimation of more sources with less sensors. *Signal Processing, Eurasip*, 30(3):279–297, February 1993.

[18] A. SWAMI, G. GIANNAKIS, and S. SHAMSUNDER. A unified approach to modeling multichannel ARMA processes using cumulants. *IEEE Trans. on SP*, 41, 1993.

[19] A. SWAMI and J. M. MENDEL. ARMA parameter estimation using only output cumulants. *IEEE Trans. on ASSP*, 38(7):1257–1265, July 1990.