

PROCEEDINGS
SPIE—The International Society for Optical Engineering

Advanced Signal-Processing Algorithms, Architectures, and Implementations

Franklin T. Luk
Chair/Editor

10–12 July 1990
San Diego, California

Sponsored by
SPIE—The International Society for Optical Engineering

Published by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA



Volume 1348

SPIE (Society of Photo-Optical Instrumentation Engineers) is a nonprofit society dedicated to the advancement of optical and optoelectronic applied science and technology.

Eigenvalue Decomposition of a Cumulant Tensor with Applications

P. Comon

*THOMSON SINTRA, BP 53,
Cagnes/Mer Cedex, 06801 France*

and

J.F. Cardoso

*ENST, 46 rue Barrault,
Paris Cedex, 75634 France*

ABSTRACT

The so-called Independent Component Analysis (ICA) raises new numerical problems of particular nature. In fact, contrary to the Principal Component Analysis (PCA), ICA requires the resorting to statistics of order higher than 2, involving necessarily objects with more than two indices, such as the fourth-order cumulant tensor. ICA computation may be related to the diagonalization of the cumulant tensor, with some particularities stemming from the symmetries it enjoys. Two algorithms are proposed. The first connects the problem to the computation of eigenpairs of an hermitian operator defined on the space of square matrices. The second algorithm reduces the complexity by making use of the information redundancy inherent in the cumulant tensor; its convergence properties are much alike those of the Jacobi algorithm for EVD calculation.

Key words: Identification, Cumulant, Principal Component, Independent Component, Mixture, Contrast

1. INTRODUCTION

Given any element u of an Hilbert space H , the quantity u^* will denote the dual of u . If u and v are two vectors, then their scalar product may be written $\langle u, v \rangle = u^*v$. Let A be a linear operator on H ; the notation A^* should not be confused with the adjointed operator denoted A^H , which is defined as $\langle Au, v \rangle = \langle u, A^H v \rangle$, $\forall u, v \in H$. For the sake of clarity, we shall respect in the paper the following notations: uppercases denote matrices, tensors, or operators, boldfaced lowercases indicate vectors, and plain lowercases indicate scalars.

1.1. Cumulants

Cumulants are objects extensively used in statistics that serve to approximate a joint probability density function by an expansion up to a given order [1]. The covariance of a N-variate random variable is a table of two indices, each index varying in $\{1, \dots, N\}$. Likewise, a fourth-order cumulant is a table with four indices, and allows for a

complementary description of the statistical properties of the random variable considered. Let H be the Hilbert space of zero-mean random variables with values in C^N and finite moments up to order 4. For any pair of elements $\{x, y\}$ of H the cross-covariance matrix is given by $C = E\{xy^*\}$, and the scalar product by $\langle x, y \rangle = \text{trace}(C)$. Moreover, any element x of H admits a fourth-order cumulant whose components are defined as [3]:

$$Q_{hijk} = \text{cum}\{x_h, x_i^*, x_j^*, x_k\} = M_{hijk} - M_{hi}M_{jk}^* - M_{hj}M_{ik}^* - \tilde{M}_{hk}\tilde{M}_{ij}^*, \quad (1)$$

where $M_{hijk} = E\{x_h x_i^* x_j^* x_k\}$, $M_{ij} = E\{x_i x_j^*\}$, and $\tilde{M}_{ij} = E\{x_i x_j\}$, $h, i, j, k \in \{1, \dots, N\}$.

Here, superscript $(*)$ reduces to a mere complex conjugation. In subsequent sections we shall make use of two basic properties satisfied by cumulants, recalled below.

Multilinearity: If $y = Ax$, then

$$\text{cum}\{y_n, y_p^*, y_q^*, y_r\} = \sum_{h,i,j,k=1}^N A_{nh} A_{pi}^* A_{qj}^* A_{rk} Q_{hijk}. \quad (2)$$

Additivity: If x and y are statistically independent, and if $z = x + y$, then

$$\text{cum}\{z_n, z_p^*, z_q^*, z_r\} = \text{cum}\{x_n, x_p^*, x_q^*, x_r\} + \text{cum}\{y_n, y_p^*, y_q^*, y_r\}. \quad (3)$$

Note that the latter property is not satisfied by moments. Cumulants enjoy the required properties that authorize the denomination of tensors [2]. Nevertheless, since it is not essential for our purposes, covariant and contravariant indices will not be distinguished by their position, though all the operations executed on tensors in the sequel will still respect the rules of tensor calculus. In particular, Einstein notation will not be utilized.

1.2. Problem statement

Let s and e be two zero-mean random variables of dimension N with finite moments up to order 4. Assume that the components of s are statistically independent and define

$$y = \sum_{p=1}^N a_p s_p + e. \quad (4)$$

In this linear model, y is the observation, e stands for a disturbing measurement noise, and a_p are deterministic unknown vectors of unit norm. Like in the Principal Component Analysis (PCA), the goal is to recover vectors a_p based on repeated observations of y , under the single assumption that the components s_p are mutually independent.

Consider first the noiseless case. In matrix form, relation (4) may be written $y = As$. The computation of the output covariance matrix yields

$$R = A C A^*, \quad (5)$$

where $C_{pq} = E\{s_p s_q^*\} = \text{var}\{s_p\} \delta_{pq}$, since the components of s are uncorrelated by hypothesis. This relation means that the matrix R can be diagonalized by a congruent transformation. Note that such a diagonalization is not unique. Indeed let $R = LL^*$ be the lower Cholesky factorization of R ; then any matrix $B = LU$, where U is unitary, satisfies $R = BB^*$. Now let Δ be the diagonal matrix formed by the inverses of the norm of each column of B . Then the matrix $A' = B\Delta$ is also a candidate for the factorization (5). As a particular case, we find the PCA providing us

with a unitary matrix A and a matrix C formed of the eigenvalues of R . This shows that a second-order description of vector y is not sufficient in order to identify matrix A , and justifies the resorting to higher-order statistics [4].

1.3. Standardization

Standardization consists precisely of filtering the observation y by the inverse of a Cholesky factor of R . The terminology of "whitening" is also sometimes used. In practice, this amounts to solve for $\bar{y}$ the equation $L \bar{y} = y$. If matrix A is written as $A = LU\Delta$ in accordance with the previous section, the standardization allows to estimate the triangular part L and to deal henceforth with the unitary part U . In this first preprocessing we get rid of a piece of the problem, and we are left with the identification of a unitary matrix, U , from the observation:

$$\bar{y} = U s. \quad (6)$$

The final solution will be given in a last step as $A = LU\Delta$, as described previously. In the rest of the paper, we shall assume that model (6) is satisfied and deal exclusively with standardized variables, unless it is otherwise specified. The bar above standardized variables will be omitted for the sake of lightness when no confusion is possible.

2. FIRST APPROACH

2.1. The cumulant tensor as a linear operator

The $N \times N \times N \times N$ cumulant tensor, Q , can be seen as a linear operator Ω defined on the Hilbert space of $N \times N$ matrices, provided with the multiplication rule

$$\Omega(M)_{pq} = \sum_{i,j=1}^N Q_{pqij} M_{ij}. \quad (7)$$

Note that according to this rule, Ω satisfies $\langle \Omega M, N \rangle = \langle M, \Omega N \rangle$. Such an operator inherits the symmetry properties of cumulants, the most important being hermicity since we have from (1): $Q_{ijpq} = Q_{pqij}^*$. Hence Ω admits a spectral decomposition with real eigenvalues and orthonormal eigenmatrices [7]:

$$\Omega = \sum_{n=1}^N \lambda(n) M(n) M(n)^*. \quad (8)$$

Moreover, the cumulant tensor also enjoys by construction

$$Q_{ijpq} = Q_{jiqp}^*. \quad (9a)$$

It is easily seen [7] that this property allows eigenmatrices to be themselves chosen hermitian. The phase indetermination, inherent in eigenpair computation of usual (two indices) hermitian operators, can be changed into a sign indetermination in the case of four indices operators possessing the symmetries (9a).

The whole set of symmetry operations that leaves fourth order cumulants invariant can be generated by considering a last symmetry property:

$$Q_{ijpq} = Q_{ipjq} \quad (9b)$$

It is important to insist that the individual terms $\lambda(n) M(n) M(n)^*$ in (8) do not satisfy such a property (only their

sum does) unless the N-variate observed process satisfies the model (6). As a matter of fact, it may be shown that if model (6) is assumed, then all eigenmatrices have a rank one and therefore satisfy (9b) individually. This will become clearer in the next section.

2.2. Identification via Eigenvalue decomposition

Based on model (6) and using properties (2) and (3), the particular form of the cumulant tensor Q can be derived:

$$\Omega = \sum_{p=1}^N \kappa_p \mathbf{a}_p \mathbf{a}_p^* \mathbf{a}_p^* \mathbf{a}_p^* ; \quad \kappa_p = \text{cum}\{s_p, s_p^*, s_p^*, s_p\}. \quad (10)$$

Since the $\mathbf{a}_p$ have been orthogonalized in the standardization process, matrices $\mathbf{a}_p \mathbf{a}_p^*$ form an orthonormal set so that equation (10) also appears as a decomposition of tensor Ω into orthonormal matrices. Owing to the uniqueness of the spectral decomposition, expressions (8) and (10) must then be identical. Provided the source kurtosis κ_p (which appear as eigenvalues in equation 10) are all distinct, eigenmatrices associated with non-zero eigenvalues are uniquely determined, and are therefore necessarily of the form $\mathbf{a}_p \mathbf{a}_p^*$. Thus, eigenmatrices must be of rank one, and eigenvalues must coincide with κ_p . Hence the computation of eigenmatrices of Q provides the component vectors $\mathbf{a}_p$. This identification theorem is the backbone of the first approach.

When noise is present or when the cumulants are not estimated with sufficient accuracy, there are more than N significant eigenmatrices which are additionally no longer of rank-one. However the N dominant eigenmatrices can be retained, and their dominant eigenvector may still be used as an estimate of the true vector $\mathbf{a}_p$.

2.3. Complexity and computational aspects

The cumulant tensor can be stored in a table with two indices in an obvious manner, and the computation of eigenmatrices just amounts to compute eigenpairs of a $N^2 \times N^2$ hermitian matrix. Nevertheless, attention should be paid to the basis in which matrices are expressed. In fact, if computed in the canonical basis $\{E(i,j) = \mathbf{e}(i)\mathbf{e}(j)^*, \mathbf{e}(i)_p = \delta_{ip}\}$, eigenmatrices obtained would not be hermitian in general, due to numerical errors. Moreover, this basis contains N^2 terms and requires a decomposition with complex coefficients; in other words, it spans a linear space of same dimension as $(\mathbb{R}^2)^{N^2}$: this involves unnecessary storage requirements, since hermitian matrices form a linear space on $\mathbb{R}$ of dimension N^2 . Therefore, for both reasons of computational precision and storage, we recommend to use the following particular basis. It is formed of real symmetric matrices $\{S(i,j), 1 \leq i \leq j \leq N\}$ defined as $S(i,i) = \mathbf{e}(i)\mathbf{e}(i)^*$, $S(i,j) = (\mathbf{e}(i)\mathbf{e}(j)^* + \mathbf{e}(j)\mathbf{e}(i)^*)/\sqrt{2}$, and imaginary skew-symmetric matrices $\{S(i,j), 1 \leq j < i \leq N\}$ defined as $S(i,j) = (\mathbf{e}(i)\mathbf{e}(j)^* - \mathbf{e}(j)\mathbf{e}(i)^*)/\sqrt{-2}$.

We shall count only the number of multiplications, regardless of their nature (i.e. real or complex), which is of course a rough approximation of the computational burden. The complexity of such an algorithm is of order $O(4N^6/3)$ if all eigenpairs are required, However, formula (10) shows that in general, only N eigenvalues are significant, and only the N corresponding eigenmatrices are of interest. Hence the complexity can be decreased to $O(\alpha N^5)$, by resorting to a Lanczos algorithm for instance.

On the other hand, we should not hush up the complexity of the computation of Q itself. Suppose we have at our

disposal T independent measurements of $y(t)$. The components of Q may be computed by an expression of the form below, where a and b are coefficients chosen in order to obtain an unbiased estimator [1],

$$Q_{hijk} = aM_{hijk} - bM_{hi}M_{jk}^* - bM_{hj}M_{ik}^* - b\tilde{M}_{hk}\tilde{M}_{ij}^*, \quad (11)$$

and where M_{hijk} and M_{hi} stand now for sample moments of the form

$$M_{hijk} = \frac{1}{T} \sum_{t=1}^T y_h(t) y_i(t)^* y_j(t)^* y_k(t), \quad M_{jk} = \frac{1}{T} \sum_{t=1}^T y_j(t) y_k(t)^*, \quad \text{and} \quad \tilde{M}_{hk} = \frac{1}{T} \sum_{t=1}^T y_h(t) y_k(t).$$

For each t , the pairwise products of $y_i(t)$'s can be first computed; this needs $O(N^2)$ flops. Then all monomials can be obtained, and accumulated in the sample moments; this complexity is dominated by the computation of the fourth-order moment, which is of order $O(3TN^4/4)$ since there are $N^2(N+1)^2/4$ distinct elements in table Q .

So, the overall computational burden involved by the estimation of Q requires $O(3TN^4/4)$ flops. In general, T is much larger than N , so that this complexity also dominates the computation of eigenmatrices evaluated above. One of the reasons for looking for another alternative is the large complexity of this estimation.

3. SECOND APPROACH

It can be shown that cumulants of order k are statistics that allow the identification of polynomial transforms of degree $(k-1)$. Here, we are interested in the identification of a *linear* transformation. Obviously the linear character of the transform leaves much less degrees of freedom than the non-linear one, and it turns out that there indeed exists a redundancy in the tensor table Q . This reveals the existence of cheaper identification algorithms [4], as will be demonstrated.

3.1. Mutual and pairwise independence

The theorem given below proves the sufficiency of pairwise cumulants in table Q for the blind identification problem, but does not give any constructive algorithm. The algorithmic issue is postponed to section 3.3.

Theorem (13)

Let s be a N -dimensional vector whose components s_i are statistically independent random variables with a marginal probability density function (pdf) not reduced to a point-like mass; let A be a $N \times N$ regular matrix with unit-norm columns, and z the random vector $z = As$. Then the three following conditions are equivalent:

- (i) The components z_i of z are pairwise independent
- (ii) The components z_i of z are mutually independent
- (iii) The matrix A is a permutation.

Proof: Implications (iii) $\Rightarrow$ (ii) and (ii) $\Rightarrow$ (i) are quite obvious. We shall prove the last one, namely (i) $\Rightarrow$ (iii). We reproduce the proof derived in [6]. Denote Φ_z the characteristic function of z and Φ_s the one of s . Mutual independence of s_i yields

$$\Phi_s(u_1, \dots, u_N) = \prod_i \Phi_s(0, \dots, 0, u_i, 0, \dots, 0).$$

On the other hand $z = A s$ allows to write

$$\Phi_z(v_1, \dots, v_N) = \Phi_s(a_1^* v, \dots, a_N^* v),$$

where a_i is the i^{th} column of A ; then independence of the components s_i implies

$$\Phi_z(v_1, \dots, v_N) = \prod_i \Phi_s(0, \dots, 0, a_i^* v, 0, \dots, 0). \quad (14)$$

Now, by writing the independence of z_1 and z_2 we get

$$\Phi_z(v_1, v_2, 0, \dots) = \Phi_z(v_1, 0, 0, \dots) \Phi_z(0, v_2, 0, \dots) \Rightarrow \Phi_z(v_1, v_2, 0, \dots) = \Phi_s(a_{11}^* v_1, \dots, a_{1N}^* v_1) \Phi_s(a_{21}^* v_2, \dots, a_{2N}^* v_2).$$

Each of the two terms in the latter equation can split into the product of marginal characteristic functions. Compatibility of this relation with expression (14) necessitates

$$\Phi_s(\dots, 0, a_{1i}^* v_1 + a_{2i}^* v_2, 0, \dots) = \Phi_s(\dots, 0, a_{1i}^* v_1, 0, \dots) \Phi_s(\dots, 0, a_{2i}^* v_2, 0, \dots),$$

for any $i \in \{1, \dots, N\}$. Now the only variables whose characteristic function satisfies $\varphi(u+v) = \varphi(u)\varphi(v)$ are quasi-deterministic variables (i.e., variables having a pdf reduced to a point-like mass), and it is excluded by hypothesis. Thus it is necessary that either $a_{1i}^* v_1$ be null, or $a_{2i}^* v_2$. This result holds true for any pair (v_1, v_2) , which yields:

$$\{a_{1i} = 0 \text{ or } a_{2i} = 0\}, \forall i \in \{1, \dots, N\}.$$

By reasoning in a similar manner for all other pairs, one would obtain:

$$\{a_{ji} = 0 \text{ or } a_{ki} = 0\}, \forall i, j, k \in \{1, \dots, N\}, j \neq k.$$

But the regularity condition on A does not allow A to have a zero row or column; the entries that can be zero are therefore $a_{\sigma(i)i}$, where σ is any permutation. The normalization of the columns of A implies eventually that A is a permutation. $\square$

3.2. Contrast functions and ICA

Definition (15)

Let E be a set of variables with values in $\mathbb{C}^N$, stable by regular linear transformations. A mapping Ψ from E to $\mathbb{R}$ is named "contrast" if it satisfies the three following conditions [6]:

- Ψ depends only on the joint probability density function of the random variable considered,
- Ψ is invariant by scale changes: $\Psi(\Lambda x) = \Psi(x)$, $\forall \Lambda$ regular diagonal matrix,
- $\forall x \in E$, and $\forall A$ $N \times N$ matrix, if the components x_i of x are statistically independent, then $\Psi(Ax) \leq \Psi(x)$.

Lemma (16)

Let U be a unitary matrix, and ${}^{2r}U$ the matrix obtained by raising the modulus of each entry to the power $2r$:

${}^{2r}U_{ij} = |U_{ij}|^{2r}$, $r \in \mathbb{N}^*$. Then the spectral radius of the matrix M , $M = {}^{2r}U^n {}^{2r}U$, is smaller than (or equal to) 1.

See [6] for a proof.

Theorem

(17)

Let $z \in E$, and denote $\Gamma_{hijk}(z) = \text{cum}\{\bar{z}_h, \bar{z}_i^*, \bar{z}_j^*, \bar{z}_k\}$, where $\bar{z}$ denotes the standardized variable $L^{-1}z$. The mapping from E to R^+ defined as

$$\Psi(z_1, \dots, z_N) = \sum_i |\Gamma_{iiii}(z)|^2$$

is a contrast.

Proof

The first condition requested by (15) is a direct consequence of the definition of cumulants. Invariance by scale change results from the whitening operation (6). Now let us look at the third condition. Let A be a regular square matrix and LU its LQ factorization (transposition of the QR factorization) where U is unitary and L triangular, and let x be a random variable with mutually independent components. Denote Q_{abcd} the standardized cumulants of the variable $y = Ax$, and K_{iiii} those of x . Then from (2) we have

$$Q_{abcd} = \sum_i U_{ai} U_{bi}^* U_{ci}^* U_{di} K_{iiii}.$$

Note that L does not appear in this expression since we are dealing with standardized cumulants. Then from lemma (16) we get immediately

$$\sum_{i,j} \sum_a |U_{ai}|^4 |U_{aj}|^4 K_{iiii} K_{jjjj}^* \leq \sum_k |K_{kkkk}|^2$$

since it may be also written $g^* M g \leq g^* g$. Thus $\Psi(y) \leq \Psi(x)$. More precisely, if U is a permutation, or if the K_{iiii} 's are all null (this can occur if x is Gaussian for instance), the equality $\Psi(y) = \Psi(x)$ holds. $\square$

Definition

(18)

Let y be a complex-valued zero-mean random variable of dimension N , and R its covariance matrix. The Independent Component Analysis (ICA) consists of searching a matrix A such that:

- A is a $N \times N$ regular matrix,
- the columns of A are of unit norm,
- $R = A \Delta^2 A^*$; with $\Delta = \text{Diag}\{\Delta_{ii}\}$, $\Delta_{ii} \geq 0$, $1 \leq i \leq M$,
- the matrix Δ^2 is the covariance of a random variable z whose M components are "the most independent possible", according to the maximization of a contrast function. Incidentally, $y = Az$ with these notations.

The concept of ICA is hence linked to a contrast function when the matrix A does not satisfy almost surely $y = Az$. Note that if it does, all contrast functions give the same answer $A = \Lambda P$, where Λ is a regular diagonal matrix with unit modulus entries and P is a permutation. We shall come back to this in a moment. The difference between ICA and PCA appears now clearly in this definition. In fact, PCA is obtained by replacing the last requirement of independence by the orthogonality of the columns of A .

3.3. Algorithms for $N = 2$

Theorem (13) suggests us to look at the blind identification problem in the case $N=2$ first. Assume the observation satisfies the noisy model (4) with $N=2$, and that it has already components uncorrelated at order 2. We are looking for a unitary transform F such that the variable $z = F y$ has the most independent components in the sense of the contrast (15). Denote F as :

$$F = c \begin{pmatrix} 1 & \theta \\ -\theta^* & 1 \end{pmatrix}, c = 1/\sqrt{1+|\theta|^2}.$$

From the multilinearity property (2), the cumulants of z can be expressed in terms of those of y [6]:

$$\begin{aligned} \Gamma_{1111}/c^4 &= Q_{2222}|\theta|^4 + 4\text{Re}\{Q_{2122}\theta\}|\theta|^2 + 4Q_{1122}|\theta|^2 + 2\text{Re}\{Q_{2112}\theta^2\} + 4\text{Re}\{Q_{1112}\theta\} + Q_{1111}, \\ \Gamma_{2222}/c^4 &= Q_{1111}|\theta|^4 - 4\text{Re}\{Q_{1112}\theta\}|\theta|^2 + 4Q_{1122}|\theta|^2 + 2\text{Re}\{Q_{2112}\theta^2\} - 4\text{Re}\{Q_{2122}\theta\} + Q_{2222}, \\ 1/c^4 &= [1+|\theta|^2]^2. \end{aligned} \quad (19)$$

The contrast function $\Psi(\theta)$ is thus a ratio of two real polynomials of degree 8 in θ and θ^* . In order to find the best solution F , we must find the maxima of $\Psi(\theta)$ in the disk $\{|\theta| \leq 1\}$. In fact, other solutions lying outside this disk are directly obtainable from those lying inside since $\Psi(-1/\theta^*) = \Psi(\theta)$. Physically, this stems from the fact that if z is solution, then so is $z' = \Lambda P z$, where

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } \Lambda = e^{j\psi} \begin{pmatrix} e^{-j\psi} & 0 \\ 0 & -e^{j\psi} \end{pmatrix},$$

since Λ is indeed a diagonal matrix with unit modulus entries. Define the auxiliary variable $\xi = \theta - 1/\theta^*$. In this manner both the equivalent values of the tangent θ are obtained from ξ by solving the trinomial $\theta\theta^* - \xi\theta^* - 1 = 0$. Then the contrast $\Psi(\theta)$ can be expressed as a function of ξ only: $\Psi(\theta) = \psi(\xi)$, where $\psi(\xi)$ is a ratio of two polynomials of degree 4 in ξ and ξ^* . The change of variable allows first to get rid of the indetermination of θ , and also to divide the degree of the polynomials by half in the expression of $\psi(\xi)$. For the sake of lightness, we shall give the general explicit form of $\psi(\xi)$ in the real case only.

Real case

In the real case, all indices in cumulants can be permuted, so that they can always be sorted in increasing order. Equations (19) rewrite then

$$\begin{aligned} \Gamma_{1111}/c^4 &= Q_{2222}\theta^4 + 4Q_{1222}\theta^3 + 6Q_{1122}\theta^2 + 4Q_{1112}\theta + Q_{1111}, \\ \Gamma_{2222}/c^4 &= Q_{1111}\theta^4 - 4Q_{1112}\theta^3 + 6Q_{1122}\theta^2 - 4Q_{1222}\theta + Q_{2222}. \end{aligned} \quad (20)$$

And the contrast function can be expressed as:

$$\psi(\xi) = \left[\sum_{i=0}^4 b_i \xi^i \right] / [\xi^2 + 4]^2. \quad (21)$$

It can be shown that the stationary points of $\psi(\xi)$ are the roots of the polynomial of degree 4 below¹:

¹: the author thanks Denis Cren for calculating explicitly those terms.

$$\alpha(\xi) = \sum_{i=0}^4 c_i \xi^i, \text{ with} \quad (22)$$

$$c_4 = Q_{1111} Q_{1112} - Q_{2222} Q_{1222},$$

$$c_3 = Q_{1111}^2 + Q_{2222}^2 - 4(Q_{1112}^2 + Q_{1222}^2) - 3Q_{1122}(Q_{1111} + Q_{2222}),$$

$$c_2 = 3(Q_{1222} - Q_{1112})(Q_{1111} + Q_{2222} - 6Q_{1122}),$$

$$c_1 = 3(Q_{1111}^2 + Q_{2222}^2) - 2Q_{1111}Q_{2222} - 32Q_{1112}Q_{1222} - 36Q_{1122}^2,$$

$$c_0 = 4(Q_{1112}Q_{2222} - Q_{1222}Q_{1111} - 3Q_{1112}Q_{1111} + 3Q_{1222}Q_{2222} - 6Q_{1122}Q_{1112} + 6Q_{1122}Q_{1222}).$$

whereas we were expecting a polynomial of degree 5. Thus, there are in this case *explicit analytical solutions*, and at most two of them correspond to maxima of $\psi(\xi)$. Moreover, since there are in general only two real roots to polynomial $\alpha(\xi)$, the contrast $\psi(\xi)$ admits in general a single maximum.

Noiseless case

Let us go back to the complex case. There is another situation where the explicit solution turns out to be simple. In fact in the noiseless case, the relation $(\xi + \rho)^2 = 0$ is always satisfied at stationary points [4], where

$$\rho = \frac{Q_{1112} - Q_{1222}}{Q_{1122}}.$$

This provides us with two equivalent double roots defined by:

$$|\theta_k| = (-1)^k \frac{|\rho|}{2} + \sqrt{\frac{\rho\rho^*}{4} + 1}, \arg\{\theta_k\} = \arg\{\rho\} + (k-1)\pi, k \in \{0, 1\}. \quad (23)$$

Efficient algorithms for computing the maxima of $\psi(\xi)$ in the unit disk in the noisy complex case still remain to be found.

3.4. Algorithm for $N > 2$

In the previous section, we used the contrast function as a tool for computing the ICA of a 2-variate observation. Here, we shall decompose the identification of the unknown $N \times N$ unitary matrix into a sequence of 2×2 Givens rotations. In fact, remind that any unitary matrix may be expressed (non uniquely) as the product of $N(N-1)/2$ Givens rotation with real cosine, and a diagonal matrix with unit modulus entries. The procedure looks very much alike the Jacobi algorithm for diagonalizing hermitian matrices. The purpose of this section is to show that the global contrast function increases each time a pair is processed.

Lemma

Define

$$\Omega(y) = \sum_{h,i,j,k} |Q_{hijk}|^2, \quad (24)$$

where Q denotes the tensor of the whitened variable $\bar{y}$. Then $\Omega(y)$ is constant under linear and regular transforms.

Proof

The proof is quite obvious. Since the variable considered is whitened, it suffices to prove the invariance under

unitary transforms. Let x be a standardized (whitened) random variable, and y be defined as $y = U x$ where U is unitary. Denote Q and K the fourth-order cumulant tensor of y and x , respectively. Then from (2):

$$Q_{ijkl} = \sum_{a,b,c,d} U_{ia} U_{jb}^* U_{kc}^* U_{ld} K_{abcd}.$$

It results immediately that:

$$\Omega(y) = \sum_{a,b,c,d} \sum_{e,f,g,h} \sum_{i,j,k,l} U_{ia} U_{jb}^* U_{kc}^* U_{ld} U_{ie}^* U_{jf} U_{kg} U_{lh}^* K_{abcd} K_{efgh}^*.$$

But $\sum_i U_{ia} U_{ie}^* = \delta_{ae}$ since $U U^* = I$. Consequently in the sum above, only the terms for which $a=e, b=f, c=g$, and $d=h$ subsist, yielding

$$\Omega(y) = \sum_{a,b,c,d} |K_{abcd}|^2,$$

which is the definition of $\Omega(x)$. This proves that $\Omega(y)$ is invariant. $\square$

This lemma, used in the proof of theorem (26), shows incidentally that the maximization of marginal cumulants is equivalent to the minimization of cross-cumulants.

Algorithm

(25)

Given a $N \times T$ data matrix, Y , the algorithm:

- computes the triangular matrix L such that $YY^* = LL^*$,
- compute $A := L^{-1}$ and $Y := A Y$,
- then executes an increasing number of sweeps, each defined as follows:
for $i = 1$ to N and for $j = i$ to N ,
 - compute the cumulants Q_{abcd} , where $a,b,c,d \in \{i,j\}$ as in (1)
 - compute the value of the tangent θ maximizing $\Psi_{ij}(\theta)$, and the corresponding Givens rotation, $F(i,j)$,
 - compute the new data matrix $Y := F(i,j) Y$,
 - accumulate $A := F(i,j) A$.

The algorithm terminates when all Givens rotations are equal to identity, up to a given precision level.

Theorem

(26)

As the number of sweeps tends to infinity, algorithm (25) converges.

Proof

We shall show that the contrast function is monotonically increasing and bounded above. When the pair (y_i, y_j) is processed, only the components (i,j) of y are affected by the transform. Consequently the marginal cumulants of the form Q_{pppp} where $p \notin \{i,j\}$ are not affected by this Givens rotation. Since the cumulants Q_{iiii} and Q_{jjjj} increase in modulus, the contrast function increases, by construction. It increases in the strict sense if the rotation $F(i,j)$ is different from identity. On the other hand from Lemma (24), the contrast function $\Psi(\theta)$ is bounded above by the fixed positive number $\Omega(y)$. $\square$

Note that this theorem does not give any idea about the speed of convergence, and we should better expect a number of sweeps of order N . This algorithm has shown excellent behavior on noisy measurements. However a more accurate

analysis of the convergence still needs to be completed, showing in particular that there is a unique maximum, and that it is attainable by such a relaxation scheme. These issues are also addressed in the convergence analysis of the standard Jacobi algorithm.

3.5. Complexity

In the real (resp. complex) case, processing one pair reduces to computing 5 (resp. 6) cumulants and to solving a polynomial equation. In the same manner as in section 2.3, we can notice that it is better to compute first 3 (resp. 3) pairwise products between the rows of Y , and then compute the 5 (resp. 6) cumulants, which amounts to $O(8T)$ flops (resp. $9T$). The computation of the roots of a polynomial represents a fixed cost, that we may assume negligible compared to T . Lastly, filtering the data requires then $O(4T)$ flops, and accumulation of the transform $O(N)$ flops.

Thus the overall complexity of running K sweeps in the real case (resp. complex) is of order $6KN^2T$ flops (resp. $13KN^2T/2$), since there are $N(N-1)/2$ pairs to process in each sweep. Accordingly, the complexity is again dominated by the calculation of cumulants themselves, and it is smaller than in the first approach by one order of magnitude if the number of sweeps, K , is of same order as N .

There exists another way of organizing the computations. In fact, if all cumulants of y are once for all computed as in section 2.3, the updating of pairwise cumulants after a Givens rotation can be done with the help of relation (2). The overall complexity of this scheme is $O(TN^4/8)+O(5KN^6/12)$ in the real case, and $O(3TN^4/4)+O(3KN^6)$ in the complex case. This procedure is more attractive than the above only for very large values of T and K , namely if KT is larger than $O(N^4)$.

4. APPLICATIONS

Some recent results are available in the literature that make use of cumulants to improve performance of existing algorithms, but the algorithms defined are very much akin to them and do not question in general the philosophy assumed. In contrast, the use of cumulants in our framework really sheds new light on some classical problems including data compression, equalization, detection, classification, localization and estimation. Let us pause a moment on these signal processing problems.

Data compression is the simplest example. It suffices to replace PCA by ICA when computing the dominant components; compressed data are then obtained by an oblique projection [6]. The advantage appears quite clearly when the background noise is strongly anisotropic.

Equalization consists essentially of separating random signals that have been linearly mixed in an unknown manner during their travel on a transmission line. This problem can be solved with the help of ICA when signals are unknown [4] [5], whereas it is usually addressed by exploiting known properties of the signals.

Detection of the number of significant signals present can be carried out by simply testing the diagonal elements of matrix Δ in the ICA expression [6], similarly to eigenvalue-based detection in the so-called high-resolution methods.

Bayesian supervised classification is based on the knowledge of the joint pdf of the data observed in each class. Yet, estimation of this pdf is possible only for long data records and limited number of classes. Here, ICA used as a preprocessing allows to approximate the joint pdf by the product of several marginal pdf [6], hence allowing

classification in much less restrictive cases (more classes, shorter records).

Localization of N sources from measurements on an antenna of N sensors can also be addressed with the help of ICA. With this intention, it suffices to pick up each column of matrix A in the ICA expression, which corresponds to a filter matched to each source direction [5]. After a mere parametric regression, the directions of arrival of impinging sources can be obtained directly, i.e., without exhaustive search like in the Music algorithm (no direct procedure was available to date, except with special antenna by using the so-called Esprit approach). Results of this paper show that the complexity of the ICA is of order $O(TN^3)$ regardless of the antenna, which is indeed polynomial. For 2-dimensional antennas this argument carries weight.

This survey of potential applications is not very thorough, but already reveals a wide field of possible research activities.

5. CONCLUDING REMARKS

Independent Component Analysis should attract more and more attention because of its many possible applications in signal processing and statistics. Its computation turns out to require the use of statistics of order higher than two, which are basically tensors. This makes it difficult to design efficient algorithms, because of large computational and storage requirements. The first approach presented uses a spectral decomposition of the cumulant tensor into eigenvalue and eigenmatrices, which is a general purpose tool with well known properties. Particularities of the first approach include the possibility to extract more signals than the observation dimension, N (this is not discussed in the paper for reasons of space). The second approach aims to decrease both storage and complexity by taking advantage of the redundancy inherent in the cumulant tensor. Because of its similarity to Jacobi's algorithm, it also provides an obvious parallel implementation. However, some theoretical issues are still left open, regarding speed of convergence for instance.

6. REFERENCES

- [1] M.G.KENDALL and A.STUART, *The Advanced Theory of Statistics*, vol.1, 1960, Griffin.
- [2] P.McCULLACH, *Tensor Methods in Statistics*, 1987, Chapman and Hall.
- [3] D.R.BRILLINGER, *Time Series, Data Analysis and Theory*, Holden Day, 1981.
- [4] P. COMON, "Separation of Sources using High-Order Cumulants", *SPIE Conference on Advanced Algorithms and Architectures for Signal Processing*, vol. Real-time Signal Processing XII, San Diego, aug 8-10, 1989, 170-181.
- [5] J.F.CARDOSO, "Source Separation using Higher-Order Moments", *Conference ICASSP*, Glasgow, 1989, 2109-2112.
- [6] P.COMON, "Independent Component Analysis and Blind Identification", *Traitement du Signal*, special issue *Non-Linear and Non-Gaussian*, fall 1990.
- [7] J.F.CARDOSO, "Eigenstructure of the Fourth-Order Cumulant Tensor with Application to the Blind Source Separation Problem", *Conference ICASSP*, Albuquerque, New Mexico, april 3-6, 1990.

Acknowledgment

This work has been supported in part by the Direction of Research, Studies and Techniques (DRET), Paris.