SPIE Conf. Adv. Sig. Proc. VIII, San Diego, 22-24 July, 1998, pp.2–13

Blind channel identification and extraction of more sources than sensors

P. Comon[b a]

[a] I3S–CNRS, Algorithmes/Euclide, 2000 route des Lucioles, Sophia-Antipolis, F-06410 Biot, France

[b] Eurecom Institute, Mobile Communications Dept, BP193, F-06904 Sophia-Antipolis, France

ABSTRACT

It is often admitted that a static system with more inputs (sources) than outputs (sensors, or channels) cannot be *blindly* identified, that is, identified only from the observation of its outputs, and without any a priori knowledge on the source statistics but their independence. By resorting to High-Order Statistics, it turns out that static MIMO systems with fewer outputs than inputs *can be* identified, as demonstrated in the present paper.

The principle, already described in a recent rather theoretical paper, had not yet been applied to a concrete blind identification problem. Here, in order to demonstrate its feasibility, the procedure is detailed in the case of a 2-sensor 3-source mixture; a numerical algorithm is devised, that blindly identifies a 3-input 2-output mixture. Computer results show its behavior as a function of the data length when sources are QPSK-modulated signals, widely used in digital communications.

Then another algorithm is proposed to *extract* the 3 sources from the 2 observations, once the mixture has been identified. Contrary to the first algorithm, this one assumes that the sources have a known discrete distribution. Computer experiments are run in the case of three BPSK sources in presence of Gaussian noise.

Keywords: High-Order Statistics (HOS), Source Separation, Downlink Communications, User Extraction, Cumulant Tensor, Binary Quantics, Multiway array, Independent Component Analysis (ICA), Multiple Inputs Multiple Outputs (MIMO) static linear systems

1. INTRODUCTION

The principles on which the *blind identification* algorithm is based are related to rather unknown and forbidding canonical decompositions of homogeneous polynomials. Some of these results need to be introduced comprehensively. But algebra has been kept to a minimum, and a tradeoff between compactness and readability of the notation has been sought.

1.1. Notation

Throughout this paper, boldface letters, like $\mathbf{u}$, will denote single way arrays, *i.e.* vectors. Its *dimension* is the range of variation of its index. Capital letters, like G, will denote arrays with more than one way, *i.e.* matrices or many-way arrays. The *order* of an array will refer to the number of its ways. The entries of an array of order d are accessed via d indices, say $i_1..i_d$, every index i_a ranging from 1 to n_a. The integer n_a is one of the d *dimensions* of the array. For instance, a matrix is a 2-way array (order 2), and has thus 2 dimensions. A product between two arrays can be done in several manners. It is convenient to define two particular products.

Given two arrays, $A = \{A_{ij..\ell}\}$ and $B = \{B_{i'j'..\ell'}\}$ of orders d_A and d_B respectively, having the same first dimension, define the contraction product as:

$$(A \bullet B)_{j..\ell j'..\ell'} = \sum_{i=1}^{n_1} A_{ij..\ell} B_{ij'..\ell'}$$

Other author information: comon@unice.fr comon@eurecom.fr

This product yields an array of order $d_A + d_B - 2$. Next, the outer product yields an array of order $d_A + d_B$ and is simply defined as:

$$(A \circ B)_{ij..\ell i'j'..\ell'} = A_{ij..\ell}\, B_{i'j'..\ell'}$$

Examples. (i) Rank-one matrices are of the form $\mathbf{u} \circ \mathbf{v}$. (ii) The usual product between two matrices would be written $A\,B = A^T \bullet B$. (iii) The scalar product between two vectors can be written $\mathbf{u} \bullet \mathbf{v}$. In the remaining, $(^T)$ will stand for transposition, $(^*)$ for complex conjugation, and $(^\dagger)$ for transposition and conjugation (*i.e.* Hermitian transposition).

1.2. Array decomposition

Now imagine data samples are stored in a $d-$way array, $\{G_{ij..\ell}\}$. This array is decomposable[1] if it is equal to the outer product of d vectors: $G = \mathbf{u} \circ \mathbf{v} \circ \cdots \circ \mathbf{w}$. A general array is a superposition of decomposable arrays, and a common problem in data analysis is precisely to determine those constituting factors.[1,2] In order to fix the ideas, let's concentrate on the fourth order case, and consider the array $\{G_{ijk\ell}\}$. The problem consists of finding a family of vector quadruples, $(\mathbf{t}(p), \mathbf{u}(p), \mathbf{v}(p), \mathbf{w}(p))$, such that $G = \sum_{p=1}^{\omega} \mathbf{t}(p) \circ \mathbf{u}(p) \circ \mathbf{v}(p) \circ \mathbf{w}(p)$. Clearly, three of the four factors can be determined only up to a constant multiplicative factor. It is thus legitimate to assume that these vectors have unit norm, without restricting generality, so that the model to identify is eventually:

$$G = \sum_{p=1}^{\omega} \gamma(p)\; \mathbf{t}(p) \circ \mathbf{u}(p) \circ \mathbf{v}(p) \circ \mathbf{w}(p) \tag{1}$$

where $\gamma(p)$ are unknown scalars, $1 \leq p \leq \omega$, and ω is minimal. For the moment, we shall refer to this equation as the Canonical Decomposition (CAND) of G. Note that this problem deflates to the standard factor analysis in the case of 2-way arrays.[3] In the latter case, it now well known that the minimal value of ω allowing such a canonical decomposition equals the *rank* of the matrix, and that the factors can be obtained by Singular Value Decomposition (SVD). However, the uniqueness is obtained to the price of imposing the additional constraint of orthogonality among each of the 2 families of ω vectors. This constraint is not mandatory at higher orders,[4] as will be pointed out subsequently.

1.3. Motivation

The motivation to go to higher order arrays is three-fold. First, as just said, the uniqueness is obtained with the help of an arbitrary constraint of orthogonality at order 2. This constraint may not be in accordance with the actual structure of the data. Second, data are often arranged in many-way arrays; the reduction to a 2-way array represents a loss of information. Third, the number of factors that can be identified is limited to the rank of the data matrix, itself bounded by the smallest dimension. Yet, there may very well be more factors than the smallest dimension.

Another field in data analysis where CAND is useful is learning theory. More precisely, if a probability density function (pdf) needs to be computed from a finite set of data in dimension n, then the number of samples required to reach a given relative precision grows exponentially with n. In practice, this makes it impossible to estimate pdf's in dimensions as small as 9 or 10.[5] One solution consists of splitting first the data into two subspaces of smaller dimension, so that the pdf rewrites as well as possible as $f(\mathbf{u}, \mathbf{v}) \approx g(\mathbf{u})\, h(\mathbf{v})$. In other words, the data are almost statistically independent in the two subspaces.[6] The number of samples required in each of the two subspaces is thus smaller by several orders of magnitude.

Beside data analysis, there are a number of fields where CAND is very useful, including Radar,[7] Acoustics, Complexity,[1] or Digital communications, and particularly in military applications such as interception and classification of radiating sources. In fact, the estimation of the sources must be carried out *blindly* in many military applications, that is, with very few a priori knowledge on the expected sources. In the civil domain, mobile communications need channel identification or equalization, which is carried out with the help of known sequences, emitted with this purpose. Since a small part of the source signal is known, one often refers to *semi-blind equalization*[8] of the channel. Constructive algorithms executing semi-blind equalization are based on principles of both blind and clear-sighted equalization, hence the interest in blind approaches. In these problems, the statistics of the data are put in a $d-$way array instead of the data themselves.

To be more explicit, assume the data are measured on n *sensors*, and that they satisfy the following $n-$*channel* linear model:

$$\mathbf{y} \approx A\,\mathbf{x} \tag{2}$$

where the components of $\mathbf{x}$ (referred to as *sources*) are statistically independent, and where A is a $n \times \omega$ unknown full rank matrix. The problem of estimating the sources from measurements of $\mathbf{y}$, solely based on independence between the x_i's is addressed in the literature under the terminology of Independent Component Analysis (ICA).[9,10] It may or may not conduct a prior *blind* identification of the mixing matrix A. In the former instance, it can be seen as a *symmetric* version of CAND. In the framework of ICA, the originality of the present work lies in the fact that the number of sources, ω, may be larger than the number of sensors, n. Let's now be more explicit.

Random variables x_i are statistically independent if all their cross-cumulants vanish.[11] For instance, at order 4, we should have that:

$$\mathrm{Cum}\{x_i, x_j, x_k, x_\ell\} = \gamma(i)\,\delta(i, j, k, \ell)$$

where δ is the Kronecker function, equal to 1 if all its arguments are equal and null otherwise. Yet, cumulants satisfy the multilinearity property,[11] so that (2) implies that:

$$G_{ijk\ell} = \sum_p \gamma(p)\, A_{ip}\, A_{jp}\, A_{kp}\, A_{\ell p} \tag{3}$$

if G denotes the fourth order cumulant tensor of $\mathbf{y}$. Now denoting by $\mathbf{a}(p)$ the pth column of A, it is easily seen that (3) can be written as a symmetric canonical decomposition:

$$G = \sum_{p=1}^{\omega} \gamma(p)\, \mathbf{a}(p) \circ \mathbf{a}(p) \circ \mathbf{a}(p) \circ \mathbf{a}(p) \tag{4}$$

This shows more explicitly that ICA is the *symmetric version* of CAND.

In the remaining, we shall be interested mainly in this symmetric decomposition, with $d = 4$, and with moderate values of the dimension n. In particular, in downlink mobile communications, it is realistic to assume that the receiver diversity will range between $n = 2$ and $n = 4$. In fact current equipment offers only $n = 1$, but it is reasonable to assume that $n = 2$ will be available in the near future, by exploiting either the spatial, sampling, or polarization diversities. In (4), $n-$dimensional vectors $\mathbf{a}(p)$ account for this diversity by the fact that they are not mutually collinear.

1.4. State of the art

In real world applications, it happens that the observation is dynamic, that is, $\mathbf{y}(t) = H(t) \star \mathbf{x}(t)$, where the mixture is characterized by a multichannel impulse response $H(t)$ rather than by a constant matrix A. In such cases, additional equations can be utilized, as demonstrated by Tong[12] or Regalia,[13] even if ICA is underlying in all multichannel blind equalization problems.[14] But the problem with fewer sensors than sources has been very little addressed and only very recently by the Signal Processing community, even in the static case (see for instance DeLathauwer[15] Tong[12] Cao[16] Comon[4] et alterae).

In this paper, we concentrate only on the static model (2), also encountered in factor analysis. In the latter field, the problem started to raise serious attention around 1970 (see in particular Caroll[17] and Kruskal[1]). With the terminology of this field, $\omega > n$ is translated into "more factors than subjects".

Bergman[18] has been probably the first to notice that the concept of *rank* is difficult to extend from matrices to higher order arrays. Caroll[17] provides the first *canonical decomposition* algorithm of a three-way array, later referred to as *Candecomp* model. Kruskal[1] conducts several years later a detailed analysis of uniqueness, and relates several definitions of rank. The algorithm *Candelinc*[2] was devised by Caroll and others in the eighties; it allows to compute a canonical decomposition subject to a priori linear constraints. Leurgans[19] derives sufficient identifiability conditions for the 3-way array decomposition; as opposed to Kruskal, his proof is constructive and yields a numerical algorithm running several matrix SVD's. A solid account on decompositions of 3-way arrays can also be found in DeLathauwer's PhD thesis[15]; an interesting tool defined therein is the *HOSVD*,[20] generalizing the concept of SVD to arrays of order 3, in a different manner compared to Caroll.

As shown in a recent work,[4] symmetric arrays of order d and dimension n can be **associated bijectively** to homogeneous polynomials of degree d in n variables. Based on this remark, decomposing a symmetric array is equivalent to decomposing a homogeneous polynomial into a sum of linear forms raised to the dth power (we shall go back to this in section 2). The advantage, compared to *Candecomp*, is that the symmetry is not broken. As a consequence, the problem can then be connected to early works in invariant theory[21] and multilinear algebra.[22] The first results go back to the beginning of the century with the works of Sylvester (see section 3.1) and Wakeford.[23] One can also mention, among others, the works of Rota[24] on binary quantics, and those of Reznick on quantics of even degree,[25] especially in the complex case.[26] Reznick introduced the concept of *width*, which corresponds to the rank in the case of matrices.

2. POLYNOMIALS AND TENSOR ARRAYS

2.1. Linear spaces

A cumulant array may be considered as a *tensor*, because cumulants (as well as moments) satisfy the multilinearity property.[11] For simplicity, tensors will be subsequently considered as mere arrays. The set of symmetric arrays of order d and dimension n is a vector space of dimension:[4]

$$D(n;d) = \binom{n+d-1}{d} \tag{5}$$

Now, every symmetric array G of order d and dimension n can be associated with a homogeneous polynomial p of degree d in n variables as follows:

$$p(x_{i_1}, \cdots x_{i_n}) = \sum_{k_1..k_d=1}^{n} G_{k_1..k_d}\, x_{k_1} \cdots x_{k_d} \tag{6}$$

In the above expression, some terms appear several times. For instance at order 4 and in dimension 2:

$$p(x_1, x_2, x_3, x_4) = G_{1111}x_1^4 + 4\,G_{1112}x_1^3x_2 + 6\,G_{1122}x_1^2x_2^2 + 4\,G_{1222}x_1x_2^3 + G_{2222}x_2^4$$

and the total number of terms is $n^d = 16$. For this reason, a compact notation needs to be introduced. Denote $\mathbf{i}$ a vector of n integer indices, called *multi-index*. If $\mathbf{a}$ is a vector of size n, then the following conventions are assumed:

$$\mathbf{a}^{\mathbf{i}} = \prod_{k=1}^{n} a_k^{i_k} \quad ; \quad (\mathbf{i})! = \prod_{k=1}^{n} i_k! \quad ; \quad |\mathbf{i}| = \sum_{k=1}^{n} i_k \tag{7}$$

Next, the multinomial coefficient can be written as:

$$c(\mathbf{i}) = \frac{|\mathbf{i}|!}{(\mathbf{i})!} \tag{8}$$

For instance, in the binary case, $c(1111) = 1$, $c(1112) = 4$, and $c(1122) = 6$; notice that in the binary case, the multi-index $\mathbf{i}$ is necessary of the form $(i, d-i)$ so that one can denote $c(\mathbf{i}) \equiv c(i)$, with some abuse of notation. In addition, $c(i) = \binom{d}{i} = \frac{d!}{i!\,(d-i)!}$.

With this notation, any homogeneous polynomial of the form (6) can be written compactly without redundancy:

$$p(\mathbf{x}) = \sum_{|\mathbf{i}|=d} G_{\mathbf{i}}\, c(\mathbf{i})\, \mathbf{x}^{\mathbf{i}} \tag{9}$$

The linear space of homogeneous polynomials is of dimension $D(n;d)$ as stated above, and one can choose as basis the set of all monomials of degree d: $\mathcal{B}(n;d) = \{\mathbf{x}^{\mathbf{i}},\ |\mathbf{i}| = d\}$.

Now the last ingredient we need is a scalar product. It turns out that the choice we are going to make is crucial in the development of some properties. Let p and q be two homogeneous polynomials of degree d in n variables, $p(\mathbf{x}) = \sum_{|\mathbf{i}|=d} \gamma(\mathbf{i}, p)\, c(\mathbf{i})\, \mathbf{x}^{\mathbf{i}}$, and $q(\mathbf{x}) = \sum_{|\mathbf{i}|=d} \gamma(\mathbf{i}, q)\, c(\mathbf{i})\, \mathbf{x}^{\mathbf{i}}$. Then define their scalar product as:

$$\langle p, q \rangle = \sum_{|\mathbf{i}|=d} c(\mathbf{i})\, \gamma(\mathbf{i}, p)^*\, \gamma(\mathbf{i}, q) \tag{10}$$

This definition corresponds to the so-called *apolar* scalar product,[24] divided by $(d!)$. This choice may seem strange and arbitrary, but it is justified by the property below. Let $L(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$ be a linear form. Then, $L^d(\mathbf{x}) = \sum_{|\mathbf{j}|=d} c(\mathbf{j})\, \mathbf{a}^{\mathbf{j}}\, \mathbf{x}^{\mathbf{j}}$, and:

$$\langle p, L^d \rangle = p(\mathbf{a}^*)^* \tag{11}$$

To see this, simply notice that from (10) $\langle p, L^d \rangle = \sum_{|\mathbf{j}|=d} c(\mathbf{j})\, \gamma(\mathbf{j}, p)^*\, \mathbf{a}^{\mathbf{j}}$. This fundamental property (11) will very useful when deriving Sylvester's theorem in section 3.1.

2.2. Definitions of rank

We define the *array rank*[1] of a symmetric array (or just *rank* in short) as the minimal value of ω allowing to obtain the decomposition given in (4). Note that other terminologies such as *tensor rank*[15] or *polynomial width*[25] are also encountered. This definition will be used in sections 3.2 and 3.4.

But other definitions have been proposed,[1,18,27] and are related to matrix ranks. In order to extract a matrix slab from a many-way array (possibly not symmetric, but assumed square here for the sake of simplicity), one can for instance define[27] *mode$-k$ vectors*, $1 \leq k \leq d$. These vectors are obtained by letting index i_k vary, $1 \leq i_k \leq n$, the other $d-1$ indices being fixed. The *mode$-k$ rank* is defined as the rank of the set of all possible mode$-k$ vectors (there are n^{d-1} of them). For symmetric arrays, mode$-k$ ranks all coincide. For matrices, the mode-1 rank is the column rank, and the mode-2 rank is the row rank. *Mode$-k$ rank* and *array rank* are apparently not related to each other. But there exist close links between the High-Order SVD introduced by DeLathauwer[15,27] and mode$-k$ ranks.

2.3. Courant-Fisher characterization

This section can be skipped in a first reading. In fact, other possible tracks for handling canonical decompositions are examined, but will not be utilized in the rest of the paper.

One alternative approach that would allow to connect the problem to known results borrowed from linear algebra is to consider a symmetric array of order d, G, as a linear operator, $\mathcal{L}_\alpha$, mapping the set of arrays of order $d-\alpha$ onto the set of arrays of order α, $0 \leq \alpha < d$. This mapping would be indeed defined as:

$$\{\mathcal{L}_\alpha(B)\}_{i_1\cdots i_\alpha} = \sum_{i_{\alpha+1}\cdots i_d} G_{i_1\cdots i_\alpha \cdots i_d} \cdot B_{i_{\alpha+1}\cdots i_d} \tag{12}$$

which rewrites in compact form, with the notation introduced in section 1.1: $\mathcal{L}_\alpha(B) = G \underbrace{\bullet\bullet\cdots\bullet}_{d-\alpha\ times} B$

In order to fix the ideas without restricting much the generality, consider the case $d = 4$. Then, there are 4 linear operators that one can define:

$$\mathcal{L}_0(B) = G \bullet B \bullet B \bullet B \bullet B;\ \ \mathcal{L}_1(B) = G \bullet B \bullet B \bullet B;\ \ \mathcal{L}_2(B) = G \bullet B \bullet B;\ \ \mathcal{L}_3(B) = G \bullet B$$

For instance, $\mathcal{L}_2$ maps a matrix to a matrix. As a linear operator in finite dimension, $\mathcal{L}_2$ can be represented by a symmetric matrix, which can be diagonalized. Its eigenvalue decomposition can be written as $\mathcal{L}_2 = \sum_p \gamma(p)\, A(p) \circ A(p)$, where the $A(p)$'s can be called *eigenmatrices*. If uniqueness of this decomposition can be guaranteed, then from (4) we should have that $A(p) = \mathbf{a}(p) \circ \mathbf{a}(p)$, that is, the eigenmatrices should be of rank one. This resembles the approaches advocated by Cardoso.[29,30]

Now consider the following optimization criteria, when $d = 4$:

$$\begin{aligned}
\Omega_0^2 = |\mathcal{L}_0(\mathbf{v}\circ\mathbf{v}\circ\mathbf{v}\circ\mathbf{v})|^2 &= |G\bullet\mathbf{v}\bullet\mathbf{v}\bullet\mathbf{v}\bullet\mathbf{v}|^2\\
\Omega_1^2 = ||\mathcal{L}_1(\mathbf{v}\circ\mathbf{v}\circ\mathbf{v}) - \mu\,\mathbf{v}||^2 &= ||G\bullet\mathbf{v}\bullet\mathbf{v}\bullet\mathbf{v} - \mu\,\mathbf{v}||^2\\
\Omega_2^2 = ||\mathcal{L}_1(\mathbf{v}\circ\mathbf{v}) - \mu\,\mathbf{v}\circ\mathbf{v}||^2 &= ||G\bullet\mathbf{v}\bullet\mathbf{v} - \mu\,\mathbf{v}\circ\mathbf{v}||^2\\
\Omega_3^2 = ||\mathcal{L}_1(\mathbf{v}) - \mu\,\mathbf{v}\circ\mathbf{v}\circ\mathbf{v}||^2 &= ||G\bullet\mathbf{v} - \mu\,\mathbf{v}\circ\mathbf{v}\circ\mathbf{v}||^2\\
\Omega_4^2 &= ||G - \mu\,\mathbf{v}\circ\mathbf{v}\circ\mathbf{v}\circ\mathbf{v}||^2
\end{aligned}$$

where the Frobenius norm is assumed. By looking for stationary values of $(\mu, \mathbf{v})$ subject to the constraint $||\mathbf{v}||^2 = 1$, it is possible to derive characteristic equations. One can show that stationarity of Ω_m implies $\mu = G\bullet\mathbf{v}\bullet\mathbf{v}\bullet\mathbf{v}\bullet\mathbf{v}$, $\forall m$. On the other hand, $d\Omega_0 = \mathbf{0}$ or $d\Omega_4 = \mathbf{0}$ lead to $G\bullet\mathbf{v}\bullet\mathbf{v}\bullet\mathbf{v} = \lambda\mathbf{v}$, but $d\Omega_1 = \mathbf{0} \Rightarrow (3\,\mathbf{v}\bullet\mathbf{v}\bullet G\bullet G\bullet\mathbf{v}\bullet\mathbf{v} - 4\,\mu\, G\bullet\mathbf{v}\bullet\mathbf{v})\bullet\mathbf{v} = \lambda\mathbf{v}$, $d\Omega_2 = \mathbf{0} \Rightarrow (\mathbf{v}\bullet G\bullet G\bullet\mathbf{v} - 2\,G\bullet\mathbf{v}\bullet\mathbf{v})\bullet\mathbf{v} = \lambda\,\mathbf{v}$, and $d\Omega_3 = \mathbf{0} \Rightarrow (G\bullet\bullet\bullet G - 4\,\mu\, G\bullet\mathbf{v}\bullet\mathbf{v})\bullet\mathbf{v} = \lambda\,\mathbf{v}$.

Of course similar results[32] hold for $d = 3$. Inspired from these equations, one can build an iteration (a kind of Rayleigh quotient[31]) allowing to compute the vector $\mathbf{v}$ dominating the decomposition of G. Then, after subtracting its contribution, one could repeat the iteration as long as there is still a residual of significant norm. As a result, one would end up with the decomposition (4). Unfortunately, some algorithmic issues still remain to be addressed.

3. BLIND IDENTIFICATION ALGORITHM

The algorithm described in section 3.2.3 is based on an old result of multilinear algebra obtained by James Joseph Sylvester. In order to make it understandable, it is necessary to go into some mathematics, that have nevertheless been kept to a minimum.

3.1. Sylvester's theorem

A binary quantic $p(x,y)=\sum_{i=0}^{d}\gamma_i\,c(i)\,x^i\,y^{d-i}$ can be written as a sum of dth powers of r distinct linear forms:

$$p(x,y)=\sum_{j=1}^{r}\lambda_j\,(\alpha_j\,x+\beta_j\,y)^d,$$

if and only if (i) there exists a vector $\mathbf{g}$ of dimension $r+1$, with components g_ℓ, such that

$$\begin{bmatrix}\gamma_0 & \gamma_1 & \cdots & \gamma_r\\ \vdots & & & \vdots\\ \gamma_{d-r} & \cdots & \gamma_{d-1} & \gamma_d\end{bmatrix}\mathbf{g}=\mathbf{0}. \tag{13}$$

and (ii) the polynomial $q(x,y)\stackrel{\text{def}}{=}\sum_{\ell=0}^{r}g_\ell\,x^\ell\,y^{r-\ell}$ admits r distinct roots.

Proof. Let's first prove the forward assertion. Define vector $\mathbf{g}$ via the coefficients of the polynomial $q(x,y)$:

$$q(x,y)\stackrel{\text{def}}{=}\prod_{j=1}^{r}(\beta_j^*\,x-\alpha_j^*\,y).$$

For any monomial $m(x,y)$ of degree $d-r$, we have $\langle m\,q,p\rangle=\sum_{j=1}^{r}\lambda_j\,\langle m\,q,(\alpha_j x+\beta_j y)^d\rangle$, by hypothesis on p. Next, from property (11), we have $\langle m\,q,p\rangle=\sum_{j=1}^{r}\lambda_j\,mq(\alpha_j^*,\beta_j^*)^*$. Yet, this scalar product is null since there is (at least) one factor in q vanishing at $x=\alpha_j^*$ $y=\beta_j^*$, for every j, $1\le j\le r$. This proves that $\langle m\,q,p\rangle=0$ for any monomial of degree $d-r$. In particular, it is true for the $d-r+1$ monomials $\{m_\mu(x,y)=x^\mu\,y^{d-r-\mu},\ 0\le\mu\le d-r\}$. And this is precisely what the compact relation (13) is accounting for, since it can be seen that $\langle m_\mu q,p\rangle=\sum_{\ell=0}^{r}g_\ell\,\gamma_{\ell+\mu}$. Lastly, the roots of $q(x,y)$ are distinct because the linear forms $(\alpha_j\,x+\beta_j\,y)$ are distinct. The reverse assertion, proved along the same lines, is the basis of the numerical algorithm developed in the next section.

3.2. Algorithm with two sensors and 3 sources

3.2.1. Generic rank and uniqueness

Sylvester's theorem is not only proving the existence of the r forms, but also gives a means to compute them. In fact, given the set of coefficients $\{\gamma_i\}$, it is always possible to find the vector $\mathbf{g}$ from (13), and then deduce the forms from the roots of the associated polynomial $q(x,y)$. More precisely, if r is unknown, one starts with a $d-1\times 2$ Hankel matrix, and one assumes $r=1$. If this matrix is full column rank, one goes to $r=2$ and test the rank of the $d-2\times 3$ Hankel matrix, and so forth. At some point, the number of columns exceeds the number of rows, and the algorithm stops. In general, this is what happens, and the *generic rank* r is obtained precisely at this stage, when $2r>d$.

For odd values of d, we have thus a generic rank of $r=\frac{d+1}{2}$, whereas for even values of d, $r=\frac{d}{2}+1$, generically. It is then clear that when d is even, there are at least two vectors satisfying (13), because the Hankel matrix is of size $\frac{d}{2}\times(\frac{d}{2}+2)$. To be more concrete, take as example $d=4$. The first Hankel matrix having more columns than rows is of size 2×4, and obviously has generically 2 vectors in its null space.

As a conclusion, when d is odd, there is generically a unique vector $\mathbf{g}$ satisfying (13), but there are two of them when d is even. In order to fix this indeterminacy, the idea proposed is to use another array, which should admit a related decomposition, as explained in the next section.

3.2.2. Choice of two cumulant arrays

We have seen why it is necessary to resort to orders higher than 2. Order 3 statistics have the great advantage that the uniqueness problem is much easier to fix, as emphasized earlier in this paper, leading to simpler constructive algorithms. Unfortunately, they often yield ill-conditioned problems, in particular when sources are symmetrically distributed about the origin. For these reasons, only 4th order statistics will be considered, even if the decomposition problem is much harder.

As in (2), denote $\mathbf{y}$ the random variable representing the $n-$sensor observation. The data we are considering here belong to the field of complex numbers. Thus there are 3 distinct 4th order cumulants that can be defined, namely:

$$G_{ijk\ell} = \mathrm{Cum}\{y_i, y_j, y_k, y_\ell\};\ \tilde{G}_{ijk\ell} = \mathrm{Cum}\{y_i, y_j, y_k^*, y_\ell^*\};\ \tilde{\tilde{G}}_{ijk\ell} = \mathrm{Cum}\{y_i, y_j, y_k, y_\ell^*\} \tag{14}$$

Again because of conditioning, the third cumulant array is not of appropriate use. But the two others can be fully exploited. In fact, assume that model (2) is satisfied. Then we have:

$$G_{ijk\ell} = \sum_{m=1}^{r} \kappa_m\, A_{im}A_{jm}A_{km}A_{\ell m};\quad \tilde{G}_{ijk\ell} = \sum_{m=1}^{r} \tilde{\kappa}_m\, A_{im}A_{jm}A_{km}^*A_{\ell m}^*,$$

where $\kappa_m = \mathrm{Cum}\{x_m, x_m, x_m, x_m\}$ and $\tilde{\kappa}_m = \mathrm{Cum}\{x_m, x_m, x_m^*, x_m^*\}$ are unknown complex and real numbers, respectively. These relations clearly show that the two 4-way arrays G and $\tilde{G}$ admit canonical decompositions that are related to each other, because *the same* matrix A enters both of them. The idea is to compute the decomposition of G, up to some indeterminacies, and then to use the second 4-way array, $\tilde{G}$, to fix it.

3.2.3. Algorithm for $n = 2$

Given a finite set of samples $\{\mathbf{y}(t),\ 0 \le t \le T\}$,

1. Compute the two 4th order cumulant arrays as follows, $\{1 \le i, j, k, \ell \le 2\}$:

$$\mu_{ij} = \frac{1}{T}\sum_{t=1}^{T} y_i(t)y_j(t) \qquad \tilde{\mu}_{ij} = \frac{1}{T}\sum_{t=1}^{T} y_i(t)y_j^*(t)$$
$$\mu_{ijk\ell} = \frac{1}{T}\sum_{t=1}^{T} y_i(t)y_j(t)y_k(t)y_\ell(t) \qquad \tilde{\mu}_{ijk\ell} = \frac{1}{T}\sum_{t=1}^{T} y_i(t)y_j(t)y_k^*(t)y_\ell^*(t)$$
$$G_{ijk\ell} = \mu_{ijk\ell} - \mu_{ij}\mu_{k\ell} - \mu_{ik}\mu_{j\ell} - \mu_{i\ell}\mu_{jk} \qquad \tilde{G}_{ijk\ell} = \tilde{\mu}_{ijk\ell} - \tilde{\mu}_{ij}\tilde{\mu}_{k\ell} - \tilde{\mu}_{ik}\tilde{\mu}_{j\ell} - \tilde{\mu}_{i\ell}\tilde{\mu}_{jk}$$

   Note that in practice, because of symmetries, only a small part of these entries need to be computed ; these details are omitted here for the sake of clarity.

2. Construct the 2×4 Hankel matrix as in (13), with $\gamma_0 = G_{1111}$, $\gamma_1 = G_{1112}$, $\gamma_2 = G_{1122}$, $\gamma_3 = G_{1222}$, $\gamma_4 = G_{2222}$.
3. Compute two 4-dimensional vectors of its null space, $\mathbf{v}_1$ and $\mathbf{v}_2$.
4. Associate the 4-way array $\tilde{G}$ with a 4th degree real polynomial in 4 real variables. This polynomial lives in a 35-dimensional linear space, and can thus be expressed onto the basis of the 35 canonical homogeneous monomials of degree 4 in 4 variables. Denote $\tilde{g}$ the corresponding 35-dimensional vector of coordinates.
5. For $\theta \in [0,\ \pi)$ and $\varphi \in [0,\ 2\pi)$ compute $\mathbf{g}(\theta, \varphi) = \mathbf{v}_1 \cos\theta + \mathbf{v}_2 \sin\theta\, e^{\jmath\varphi}$
6. Compute the 3 linear forms $L_1(\mathbf{x}|\theta, \varphi)$, $L_2(\mathbf{x}|\theta, \varphi)$, $L_3(\mathbf{x}|\theta, \varphi)$, associated with $\mathbf{g}(\theta, \varphi)$.
7. Express $|L_r(\mathbf{x}|\theta, \varphi)|^4$, $r = \{1, 2, 3\}$ in the 35-dimensional linear space, by three vectors $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$.
8. Detect the values (θ_o, φ_o) for which the vector $\tilde{g}$ is closest to the linear space spanned by $[\mathbf{u}_1,\ \mathbf{u}_2,\ \mathbf{u}_3]$
9. Set $L_r = L_r(\theta_o, \varphi_o)$, and $A = [L_1,\ L_2,\ L_2]$, where the 3 forms L_r are expressed by their 2 coordinates.

A matlab code of the complete algorithm is downloadable from `www.i3s.unice.fr/~comon`.

3.3. Computer results

Source samples have been generated according to a discrete distribution with support $\{1, \jmath, -1, -\jmath\}$. Such sources are encountered in digital communications, when the (very common) QPSK modulation is used. They have as fourth order cumulants $\kappa = 1$ and $\tilde{\kappa} = -1$. The data length T was varied from 200 to 5000 samples, and the mixing matrix was taken to be $A = \begin{bmatrix} 0.81 + 0.39\jmath & 0 & 0.35 + 0.35\jmath \\ 0 & 0.5 - 0.86\jmath & 0.86 \end{bmatrix}$.

A key issue is the choice of the performance measure. In the present case, it is not trivial to measure the error between the matrix identified by the algorithm, say $\hat{A}$, and the true mixing matrix A, since each column is computed up to a multiplicative complex number, and up to a permutation among the columns. In other words, one should measure the norm of $||A - \hat{A} \cdot D||$, for the best matrix D, formed of the product of a diagonal matrix and a permutation (such matrices are sometimes called *generalized permutations*). In order to do this, the basic tool is the computation of a distance $\Upsilon(\mathbf{u}, \mathbf{v})$ between two vectors $\mathbf{u}$ and $\mathbf{v}$, invariant up to a multiplicative complex number. For this purpose, one defines

$$\Upsilon(\mathbf{u}, \mathbf{v}) = \operatorname*{Min}_z \frac{||\mathbf{u} - z\,\mathbf{v}||^2}{||\mathbf{u}|| \cdot ||\mathbf{v}||}$$

It can be seen that if $\mathbf{u}^\dagger \mathbf{v} = ||\mathbf{u}|| \cdot ||\mathbf{v}|| \cdot \cos\theta\, e^{\jmath\psi}$, then the minimal distance $\Upsilon(\mathbf{u}, \mathbf{v})$ is reached for $||\frac{\mathbf{u}}{||\mathbf{u}||} - \frac{\mathbf{v}}{||\mathbf{v}||}\, e^{-\jmath\psi}||$ and takes the value $2(1 - \cos\theta)$. The gap between two matrices is then computed as the minimum distance over the 6 possible permutations:

$$Gap(A, \hat{A}) = \operatorname*{Min}_P \sum_{i=1}^{r} \Upsilon(\mathrm{col}_i(A), \mathrm{col}_i(\hat{A}P))$$

The range of variation of this gap is thus $[0, 6]$ in the present problem where $r = 3$. Note that this is easy to compute because of the very small dimension. For larger dimensions, one can avoid the exhaustive search for the best permutation by assuming another gap measure, of more complicated (but compact) form.[9]

In figure 1 the average gap obtained over 15 independent experiments is plotted. The gap keeps small (compared to its maximal achievable value of 6), even for a data length as small as $T = 200$. This behavior holds excellent as long as the noise is negligible. If noise is present, the performance degrades rather fast, especially for short data length T and non Gaussian noise. On the other hand, cumulant arrays are asymptotically insensitive to Gaussian noise (for large T), providing some robustness to the method.

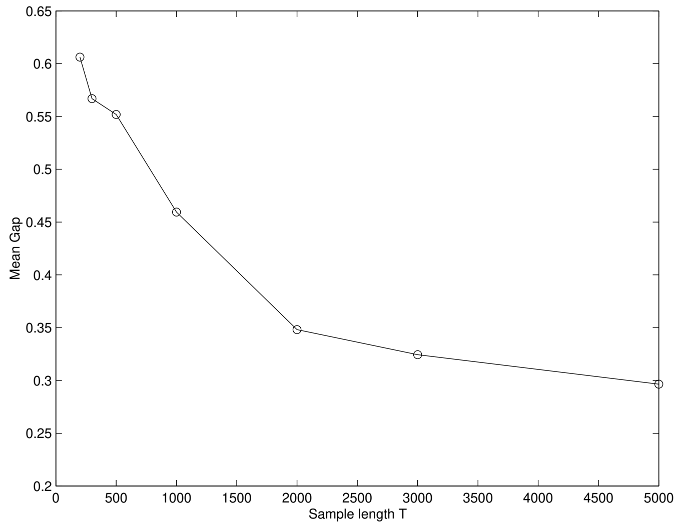

Figure 1. Mean Gap obtained over 15 snapshots and data lengths ranging from 200 to 5000 samples.

3.4. Extension to larger dimensions

For dimensions $n > 2$, Sylvester's theorem cannot apply. Instead, one should resort to Lasker-Wakeford theorem,[4] whose proof is not constructive. Consequently, efficient blind identification algorithms still remain to be devised. But there are a number of results that are already known, especially concerning the *rank* (the *width* according to Reznick[25]). More precisely, for $2 \leq n \leq 8$, it has been shown that the generic value of r is given by the following table:[4]

r	n	**2**	**3**	**4**	**5**	**6**	**7**	**8**
d	**3**	2	4	5	8	10	12	15
	4	3	6	10	15	22	30	42

The number of free parameters that remain to be fixed in order to ensure uniqueness (in the sense that a finite set of equivalent solutions can be obtained) is given by the dimension of the manifold of solutions:[4]

	n	**2**	**3**	**4**	**5**	**6**	**7**	**8**
d	**3**	0	2	0	5	4	0	0
	4	1	3	5	5	6	0	6

One can check out, for instance, that for $n = 2$, we have indeed 1 free parameter to fix when decomposing 4-way arrays, whereas there are a *finite number* of solutions in the case of 3-way arrays. This has already been pointed out in section 3.2.1: a second 4-way cumulant array had been necessary in order to fix the parameter. It is interesting to notice that the latter property often holds true for 3-way arrays; on the contrary for 4-way arrays, it occurs only for $n = 7$ in the above table. Before to close this section, it is worth insisting that there is no simple rule or formula that would yield the values of these tables: it is the result of several theorems of various origins.[4]

4. NON LINEAR INVERSION OF THE LINEAR CHANNEL

In the previous section, the problem of *blind* identification of the mixture A has been addressed, but the recovery of the sources themselves was left open. In this section, it is assumed that the mixing matrix A is given, and the goal is to estimate the input (source) vector, $\mathbf{x}$, from the observation vector, $\mathbf{y}$. Because A has more columns than rows (under-determination), it cannot be linearly inverted.

4.1. Principle

In digital communications, the discrete-time processes always take values in a finite set. Therefore, their values can be completely characterized by a set of polynomial equations. Take the example of $d-$PSK modulated sources, which satisfy $x^d = 1$. There are d solutions in the complex field, and hence a set of d allowed values. By adding to the n observation equations (2) all the $\binom{n+d}{d+1}$ homogeneous monomials of degree $d+1$, one gives oneself the opportunity to use the r additional equations $x_i^d = 1$, $1 \leq i \leq r$. The system can be solved thanks to these additional equations. In order to illustrate the idea, we shall now consider the easiest case.

This system becomes especially simple when enough linear equations can be extracted. This is in particular the case if $d = 2$ and $n = 2$. In fact we have then $r + 6$ equations in r unknowns. The first 2 equations are given by (2). The next 4 equations are given by y_1^3, $y_1^2\, y_2$, $y_1\, y_2^2$, y_2^3. These equations involve products of the form x_i^3, $x_i^2 x_j$, or $x_i x_j x_k$. By using the r remaining ones, namely $x_j^2 = 1$, the latter products reduce simply to x_i, x_j, and $x_i x_j x_k$, respectively. In other words, the system becomes almost linear, beside the $\binom{r}{3}$ terms of the form $x_i x_j x_k$.

Now take our example of section 3, where $r = 3$. Then there is a single non linear term, $x_1 x_2 x_3$. If this term is considered as a plain unknown, independently of the 3 others, we end up with a *linear system* of 6 equations in 4 unknowns:

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} A & 0 \\ & B \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_1\, x_2\, x_3 \end{pmatrix}, \quad \text{where} \quad \mathbf{z} = \begin{pmatrix} y_1^3 \\ y_1^2\, y_2 \\ y_1\, y_2^2 \\ y_2^3 \end{pmatrix}.$$

Since the 4×4 matrix B is a known function of A, it is given as soon as A is known. The above system can thus be solved for the 4 unknowns in the Least Squares (LS) sense. In the approach proposed here, the LS solution obtained for $x_1 x_2 x_3$ is simply discarded.

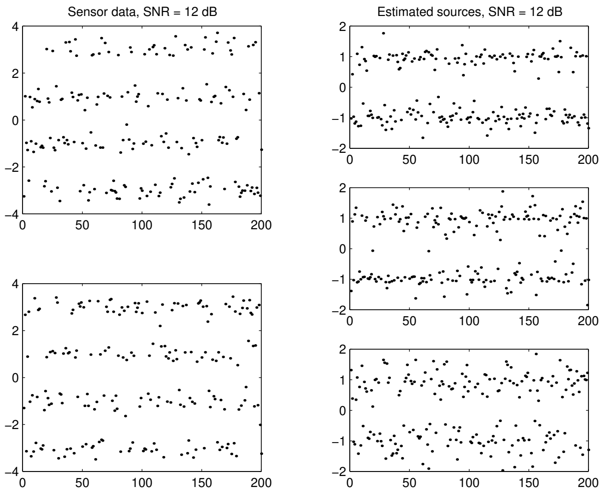

Figure 2. Typical inversion results obtained with a SNR of 12dB.

4.2. Computer results

Data have been generated according to the following model:

$$\mathbf{y} = A\,\mathbf{x} + \frac{1}{\rho}\,\mathbf{w}, \quad \text{where} \quad A = \begin{bmatrix} 2 & 0.1 & -1 \\ 0.1 & 2 & 1 \end{bmatrix}$$

Parameter ρ allows to control the noise power. The signal to Noise Ratio (SNR) is defined here as $20\log_{10}\rho$. It is thus a global value for the 3 sources, that implicitly depends on the matrix A, of course. Figure 2 shows a typical example, obtained for SNR=12dB. Note that the Signal to Interference Ratio (SIR) is of about 6dB for the first two sources, but of -6dB for the third one, measured on the space spanned by the exact directional vector. This yields an average SIR of about 3dB, coming *in addition to* the noise corruption.

As the SNR varies from 10 to 50dB, two performance measures have been calculated. First, one has computed a mean standard deviation of the normalized source estimation error:

$$\sigma_{\text{error}} = \left(\sum_{i=1}^{3} \text{variance}\left[\frac{x_i}{\sigma(x_i)} - \frac{\hat{x}_i}{\sigma(\hat{x}_i)} \right] \right)^{1/2}$$

where x_i denotes the true source value of source i and $\hat{x}_i$ its estimate. Figure 3 (left) reports the performances obtained. Second, the Bit Error Rate (BER) has been estimated over $10,000$ samples for each value of the SNR; the errors made for each of the 3 sources have been cumulated, so that the accuracy is $1/30,000 \approx 3.10^{-5}$. The BER stays below 2 percent until 10dB (see figure 3, right).

5. CONCLUSION

One of the objectives of this paper was to demonstrate that it was possible to blindly identify a static system with more inputs than outputs, and to estimate its inputs from the outputs. The principles were illustrated in the case of $n = 2$ outputs and $r = 3$ inputs. The extension to larger numbers of inputs and outputs turns out to be much more complicated, and is still being studied. With this goal, other tracks can also be followed, as briefly explained in section 2.3. One of them, in the case of even orders (*e.g.* $d = 4$), is to consider the $d-$way array as an operator mapping the space of $d/2-$way arrays onto itself; another one is to attempt to compute the stationary values of Ω_m. Despite their potential usefulness, numerical algorithms have been to date only little studied.

ACKNOWLEDGMENTS

This work has been supported in part by the CNRS telecommunications program No.TL-97104, and in part by Eurecom's industrial partners: Ascom, Cegetel, Hitachi, IBM France, Motorola, Swisscom, and Thomson CSF.

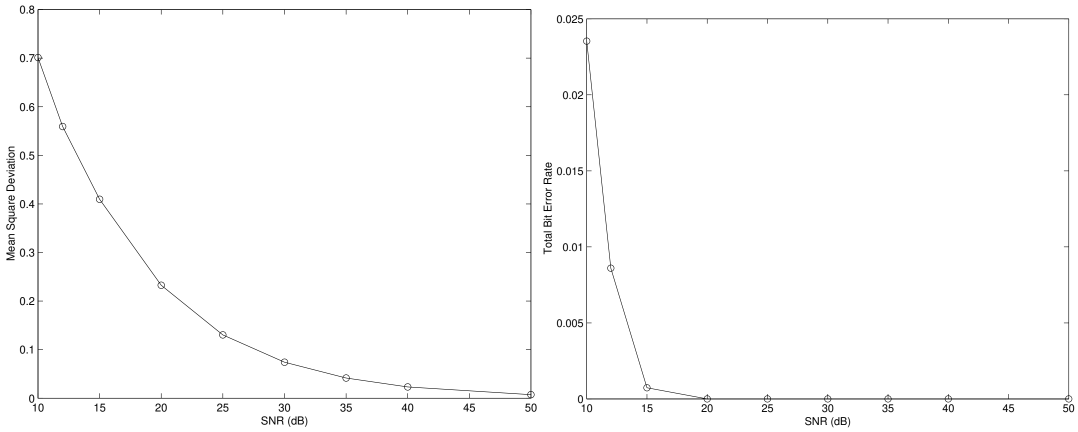

Figure 3. Mean Standard Deviation (left) and Total Bit Error Rate (right) for the 3 estimates.

REFERENCES

1. J. B. KRUSKAL, "Three-way arrays: Rank and uniqueness of trilinear decompositions," *Linear Algebra and Applications* **18**, pp. 95–138, 1977.
2. J. D. CAROLL, S. PRUZANSKY, and J. B. KRUSKAL, "CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters," *Psychometrika* **45**, pp. 3–24, Mar. 1980.
3. K. V. MARDIA, J. T. KENT, and J. M. BIBBY, *Multivariate Analysis*, Academic Press, 1979.
4. P. COMON and B. MOURRAIN, "Decomposition of quantics in sums of powers of linear forms," *Signal Processing, Elsevier* **53**, pp. 93–107, Sept. 1996. special issue on High-Order Statistics.
5. B. W. SILVERMAN, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 1986.
6. P. COMON, "Supervised classification, a probabilistic approach," in *ESANN-European Symposium on Artificial Neural Networks*, Verleysen, ed., pp. 111–128, D facto Publ., (Brussels), Apr 19-21 1995. invited paper.
7. E. CHAUMETTE, P. COMON, and D. MULLER, "An ICA-based technique for radiating sources estimation; application to airport surveillance," *IEE Proceedings - Part F* **140**, pp. 395–401, Dec. 1993. Special issue on Applications of High-Order Statistics.
8. E. DeCARVALHO and D. T. M. SLOCK, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *Proc. SPAWC 97 Conf.*, pp. 129–132, (Paris, France), Apr. 1997.
9. P. COMON, "Independent Component Analysis, a new concept ?," *Signal Processing, Elsevier* **36**, pp. 287–314, Apr. 1994. Special issue on Higher-Order Statistics.
10. J. F. CARDOSO, "High-order contrasts for independent component analysis," *Neural Computation* , 1998. To appear.
11. P. McCULLAGH, *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability, Chapman and Hall, 1987.
12. L. TONG, "Identification of multichannel MA parameters using higher-order statistics," *Signal Processing, Elsevier* **53**, pp. 195–209, Sept. 1996. special issue on High-Order Statistics.
13. V. S. GRIGORASCU and P. A. REGALIA, "Tensor displacement structures and polyspectral matching," in *Fast Reliable Algorithms for Matrices with Structure*, T. Kailath and A. H. Sayed, eds., ch. 9, SIAM Publ., Philadelphia, PA, 1998.
14. A. SWAMI, G. GIANNAKIS, and S. SHAMSUNDER, "Multichannel ARMA processes," *IEEE Trans. on Signal Processing* **42**, pp. 898–913, Apr. 1994.
15. L. DeLATHAUWER, *Signal Processing based on Multilinear Algebra*. Doctorate, Katholieke Universiteit Leuven, Sept. 1997.
16. X. R. CAO and R. W. LIU, "General approach to blind source separation," *IEEE Trans. Sig. Proc.* **44**, pp. 562–570, Mar. 1996.
17. J. D. CAROLL and J. J. CHANG, "Analysis of individual differences in multidimensional scaling via n-way generalization of eckart-young decomposition," *Psychometrika* **35**, pp. 283–319, Sept. 1970.

18. G. M. BERGMAN, "Ranks of tensors and change of base field," *Journal of Algebra* **11**, pp. 613–621, 1969.
19. S. LEURGANS, R. T. ROSS, and R. B. ABEL, "A decomposition for three-way arrays," *SIAM Jour. Matrix Anal. Appl.* **14**, pp. 1064–1083, Oct. 1993.
20. L. DeLATHAUWER, B. DeMOOR, and J. VANDEWALLE, *A Singular Value Decomposition for Higher-Order Tensors.* Second ATHOS workshop, Sophia-Antipolis, France, Sept 20-21 1993.
21. K. V. PARSHALL, "The one-hundred anniversary of the death of invariant theory," *The Mathematical Intelligencer* **12**(4), pp. 10–16, 1990.
22. J. A. DIEUDONNÉ and J. B. CARREL, "Invariant theory, old and new," in *Computing Methods in Optimization Problems-2*, pp. 1–80, Academic Press, (San Remo, Italy, Sept 1968), 1970.
23. E. K. WAKEFORD, "Apolarity and canonical forms," in *Proceedings of the London Mathematical Society*, pp. 403–410, 1918.
24. R. EHRENBORG and G. C. ROTA, "Apolarity and canonical forms for homogeneous polynomials." European Journal of Combinatorics, 1993.
25. B. REZNICK, "Sums of even powers of real linear forms," *Memoirs of the AMS* **96**, Mar. 1992.
26. B. REZNICK, "Sums of powers of complex linear forms." Preprint, private correspondence, Aug. 1992.
27. L. DeLATHAUWER and B. DeMOOR, "From matrix to tensor: Multilinear algebra and signal processing," in *Mathematics in Sig. Proc., IMA Conf. Series*, Oxford Univ Press, (Warwick), Dec 17-19 1996.
28. J. F. CARDOSO and A. SOULOUMIAC, "Blind beamforming for non-Gaussian signals," *IEE Proceedings - Part F* **140**, pp. 362–370, Dec. 1993. Special issue on Applications of High-Order Statistics.
29. J. F. CARDOSO, "Fourth-order cumulant structure forcing, Application to blind array processing," in *Proc. IEEE SP Workshop on SSAP*, pp. 136–139, (Victoria), Oct. 1992.
30. P. COMON and J. F. CARDOSO, "Eigenvalue decomposition of a cumulant tensor with applications," in *SPIE Conf. Adv. Sig. Proc. Alg.*, vol. 1348, Arch. and Implem., pp. 361–372, (San Diego, CA), July 10-12 1990.
31. L. DeLATHAUWER, P. COMON, *et al.*, "Higher-order power method, application in Independent Component Analysis," in *NOLTA Conference*, vol. 1, pp. 91–96, (Las Vegas), 10–14 Dec 1995.
32. P. COMON, "Cumulant tensors". *IEEE SP Workshop on HOS*, Banff, Canada, July 21-23 1997. Keynote address, `www.i3s.unice.fr/~comon/plenaryBanff97.html`.