Blind Identification and Source Separation in 2×3 Under-determined mixtures

Pierre Comon, *Senior Member, IEEE*

Abstract— **Under-determined mixtures are characterized by the fact that they have more inputs than outputs, or, with the antenna array processing terminology, more sources than sensors. The problem addressed is that of identifying and inverting the mixture, which obviously does not admit a linear inverse. Identification is carried out with the help of tensor canonical decompositions. On the other hand, the discrete distribution of the sources is utilized for performing the source extraction, the under-determined mixture being either known or unknown. The results presented in this paper are limited to 2-dimensional mixtures of 3 sources.**

Keywords— **Blind Identification, Source Separation and Extraction, Under-Determined Mixtures, Tensor Decomposition, High-Order Statistics**

I. Introduction

It is assumed throughout the paper that N realizations $\boldsymbol{y}(n)$ of a $K-$dimensional random vector, $\boldsymbol{y}$, are observed, and that they follow the linear statistical static model below:

$$\boldsymbol{y}(n) = \boldsymbol{A}\,\boldsymbol{x}(n) + \boldsymbol{v}(n), \tag{1}$$

where $\boldsymbol{x}(n)$ are realizations of a so-called source vector, of dimension P, $\boldsymbol{A}$ is a $K \times P$ unknown matrix, and $\boldsymbol{v}(n)$ is a noise, assumed to be independent of $\boldsymbol{x}(n)$. Throughout the paper, bold lowercases will denote vectors ($1-$way arrays), whereas bold uppercases will represent matrices or tensors (entries A_{ij} of an array $\boldsymbol{A}$ are not boldfaced, since they are scalar quantities). Mixtures in which the diversity K is smaller than the number of sources, P, are referred to as *under-determined*. We are mainly interested in such mixtures in this paper.

In the Blind Source Separation (BSS) framework, the goal consists of identifying the mixture $\boldsymbol{A}$, or estimating the sources $x_i(n)$, or both, form the sole observation of realizations $\boldsymbol{y}(n)$. Note that under-determined mixtures cannot be linearly inverted, because the rank of the mixing matrix $\boldsymbol{A}$ is bounded by K, which makes it more difficult to extract the sources, even if the mixture were known, and hence the challenge. As pointed out in many of the works subsequently quoted, this problem is ill-posed, and needs further assumptions. We shall consider the most commonly encountered assumptions on sources, namely:

A1. Sources x_i are non Gaussian and statistically mutually independent

A2. Sources x_i are discrete, and take their values in a known finite alphabet.

With the help of one of these two assumptions (both are not necessary), the BSS problem can be solved only up to inherent indeterminations. It is now well known that these indeterminations can be expressed in terms of a diagonal invertible matrix, $\boldsymbol{\Lambda}$, and a permutation $\boldsymbol{P}$, so that solutions $\boldsymbol{\Lambda P x}$ are as acceptable as $\boldsymbol{x}$. In other words, one searches for a representative of this equivalence class of solutions [1] [2]. Because of its quite general formulation and its weak assumptions, the BSS problem finds various applications in numerous areas including Factor and Data Analysis, Digital Communications, Sonar, Radar, Compression, Speech, Image processing, Econometrics, Biomedical engineering... See [3] [4] [5] and references therein.

Submitted September 10, 2001; revised October 10, 2002; accepted March 28, 2003. The author is with the Laboratoire I3S, CNRS, university of Nice, Sophia-Antipolis, France (e-mail: comon@i3s.unice.fr)

If the diversity is sufficiently large, that is, if $K \geq P$ (over-determined mixtures), many general-purpose algorithms have been proposed, both for static and dynamic mixtures, see [2] [6] [3] [7] [8] [9] [10] [5] and references therein. These algorithms require that at most one source is Gaussian, and the possible discrete character of the sources is not explicitly exploited. Fewer algorithms dedicated to discrete sources are available in the literature [11] [12] [13] [14], some of them addressing the case of dynamic mixtures, namely [15] for SISO channels and [16] for MIMO, among others.

When the diversity is small, then $K < P$ (under-determined mixtures), the problem of extracting the sources is not trivial, even when the mixture is known. This problem has been little addressed despite its practical interest. The area where this problem is most known is that of factor analysis [17] [18] [19] [20] [21] [22]. In the terminology of factor analysis , the under-determined case $\omega > K$ is translated into "more factors than subjects" [18] [19]. Works done on Parallel Factor Analysis (Parafac) of unsymmetric tensors are valid under sufficient conditions on the dimensions [19] [22]. For instance, if we translate the conditions to the symmetric case, we must have $2P + 2 \leq 3K$, which is not satisfied if $(K, P) = (2, 3)$. These approaches will thus not be considered subsequently.

Identification of static under-determined mixtures also finds applications in downlink mobile communications (in the presence of flat fading or far apart multi-paths), and feature extraction of images, among others. The earliest work in the signal processing area is probably the nice identifiability paper by Cao and Liu [1], in which no constructive algorithm was proposed. Links between blind identification of the mixture and tensor diagonalization [23] [24] [21], or linear operators acting on matrix spaces [25] [21], have been established. In [25], a constructive algorithm is described, but it requires that

$2P(P-1) \leq K^2(K-1)^2$; again, this sufficient condition is not satisfied for $(K,P) = (2,3)$.

Surveys on tensor decompositions can be found in [26] [22] [4] [27]; as will be pointed out later on, decompositions of polynomials [28] [29] can be related to those of tensors. Sometimes, one finds this problem in the literature under the name of *over-complete* mixtures [30] [5], especially in the community of neuro-computing. Lastly, underdetermined dynamic mixtures have been little addressed [31], and no constructive algorithm has been demonstrated to be able to blindly identify the channel.

The links with tensors can be emphasized if cumulants of given order are utilized. But the problem can also be formulated with the help of characteristic functions, which carries all the information. This formulation is more powerful (it allows for instance to address non linear mixtures) but becomes more difficult to handle [32] [8] [2] [33] [34].

Lastly, it is well known that it is theoretically possible to identify a generic mixture of an arbitrarily large number of discrete sources in the absence of noise, regardless of the number of sensors, if sources have a known alphabet. This comes from the fact that the linear combination of 3 independent binary sources has generically a distribution concentrated on 6 masses. The construction is deterministic and geometrical. Asymptotic performances of BSS based on this geometrical idea have been analyzed in [35]. In practice, the technique is limited by the observation length, and requires clustering. Practical algorithms have been proposed to extract binary sources in [36] [37], but are very time consuming. In section V, the algorithm extracts binary sources with a very light numerical complexity.

In this paper, it is attempted to solve the identification and extraction problems by keeping the computational load as small as possible. Our contribution is three-fold. In section III, the problem of identifying the mixing matrix is addressed, in the case of 2 sensors and 3 sources, and the only assumption made on sources is that their marginal fourth-order cumulants are non-zero whereas the cross fourth-order cumulants are null. In section IV, an extraction algorithm is proposed for *discrete* sources (essentially BPSK or MSK), and when the 2×3 mixing matrix is known (this algorithm could then be applied after an identification stage, performed with the help of the identification procedure described in section III for instance). Third, another algorithm is described in section V, and performs directly a blind extraction of discrete sources without prior identification of the mixture. In each section, computer simulations are reported.

II. Mathematical tools

A. Tensors and Quantics

A.1 Array terminology

The *order* of an array refers 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 K_a; the integer K_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.

Given two arrays, $\boldsymbol{A} = \{A_{ij..\ell}\}$ and $\boldsymbol{B} = \{B_{i'j'..\ell'}\}$ of orders d_A and d_B respectively, the *outer* product $\boldsymbol{A} \circ \boldsymbol{B}$ yields an array of order $d_A + d_B$ and is simply defined as:

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

For example, rank-one matrices are of the form $\boldsymbol{u} \circ \boldsymbol{v}$. In the remaining, $(^{\mathsf{T}})$ will stand for transposition, $(^*)$ for complex conjugation, and $(^{\mathsf{H}})$ for transposition and conjugation (*i.e.* Hermitian transposition).

In the present framework, arrays of order higher than 2 will be called *tensors* if they enjoy the multilinearity property under changes of coordinate systems. More precisely, consider a 3rd order tensor $\boldsymbol{T}$ with entries T_{ijk}, and a change of coordinates defined by 3 square invertible matrices, $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$. Then, in the new coordinate system, the tensor $\boldsymbol{T}'$ can be written as a function of tensor $\boldsymbol{T}$ as:

$$T'_{ijk} = \sum_{abc} A_{ia} B_{jb} C_{kc} T_{abc} \tag{2}$$

In particular, moments and cumulants of random variables may be treated as tensors [38]. This product is sometimes referred to as the Tucker product [27] between matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$, weighted by $\boldsymbol{T}$. Note that tensors enjoy property (2) even if the above matrices are not invertible; only linearity is required. Tensors are referred to as *symmetric* if, for any permutation σ, $\boldsymbol{T}_{\sigma(ij..\ell)} = \boldsymbol{T}_{ij..\ell}$.

A.2 Canonical Decomposition

Assume we are given a $d-$way array, $\{G_{i\,j..\ell}\}$. This array is *decomposable* [19] if it is equal to the outer product of d vectors: $\boldsymbol{G} = \boldsymbol{u} \circ \boldsymbol{v} \circ \cdots \circ \boldsymbol{w}$. A general array is a superposition of decomposable arrays, and a common problem in data analysis is precisely to determine those constituting factors [19] [20]. In order to focus our attention, let us concentrate on the fourth order case, and consider the array $\{G_{ijk\ell}\}$. The problem consists of finding a family of vector quadruples, $(\boldsymbol{t}(p), \boldsymbol{u}(p), \boldsymbol{v}(p), \boldsymbol{w}(p))$, such that $\boldsymbol{G} = \sum_{p=1}^{\omega} \boldsymbol{t}(p) \circ \boldsymbol{u}(p) \circ \boldsymbol{v}(p) \circ \boldsymbol{w}(p)$. Clearly, three of the four factors can be determined only up to a constant multiplicative scalar number. It is thus legitimate to assume that these vectors have unit norm, without restricting the generality, so that the model to identify is eventually:

$$\boldsymbol{G} = \sum_{p=1}^{\omega} \gamma(p)\, \boldsymbol{t}(p) \circ \boldsymbol{u}(p) \circ \boldsymbol{v}(p) \circ \boldsymbol{w}(p) \tag{3}$$

where $\gamma(p)$ are unknown scalars, $1 \leq p \leq \omega$. We shall refer to this equation as the Canonical Decomposition (CAND) of $\boldsymbol{G}$ when it is obtained for the smallest possible ω. Note that this problem deflates to the standard factor analysis in the case of 2-way arrays [39]. 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 [24] [4] [21], and this is precisely what makes orders higher than 2 attractive. In fact, orthogonality between the CAND vectors may not be in accordance with the actual structure of the data. A striking example is that where the number of factors exceeds the smallest dimension: it is then impossible to orthogonalize more vectors than their dimension.

A.3 Ranks

The *array rank* [19] is defined as the minimal value of ω allowing to obtain the decomposition given in (3). Note that other terminologies such as *tensor rank* [21] [4] or *polynomial width* [28], or just *rank*, are also encountered. This definition will be used in sections III-A and III-C. For consistency, a decomposable array enjoying the multilinearity property (2) will be referred to as a *rank−1 tensor*.

Other definitions have been proposed [19] [17] [26], and are related to matrix ranks; let us mention them for completeness. 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 *mode−k vectors*, $1 \leq k \leq d$. These vectors are obtained by letting index i_k vary, $1 \leq i_k \leq K$, 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 K^{d-1} of them).

For symmetric tensors, 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 *tensor rank* are not related to each other in a simple way, and we just know that the former is bounded above by the latter, by construction [26]. On the other hand, there exist close links between the High-Order SVD introduced by De-Lathauwer [21] [26] and mode−k ranks. These ranks will not be used in this paper.

A.4 Link between tensors and quantics

As will be now explained, symmetric tensors of order d and dimension K can be associated *bijectively* to homogeneous polynomials of degree d in K variables [24] [4], called *quantics* in the early works on invariant theory [40] [41]. Based on this remark, decomposing a symmetric tensor is equivalent to decomposing a homogeneous polynomial into a sum of linear forms raised to the dth power.

As a consequence, the problem can then be connected to early works [40] [42]. The first results go back to the beginning of the century with the works of Sylvester (see section II-B) and Wakeford [43]. One can also mention, among others, the works of Rota [29] on binary quantics, and those of Reznick on quantics of even degree, especially in the complex case [44]. Reznick introduced the concept of *width*, which corresponds to the rank in the case of matrices, and to our *tensor rank* introduced in section II-A.3.

Denote $\boldsymbol{i}$ a vector of K integer indices, sometimes called *multi-index*. If $\boldsymbol{a}$ is a vector of size K, then the following conventions are assumed:

$$\boldsymbol{a}^{\boldsymbol{i}} = \prod_{k=1}^{K} a_k^{i_k} \quad ; \quad (\boldsymbol{i})! = \prod_{k=1}^{K} i_k! \quad ; \quad |\boldsymbol{i}| = \sum_{k=1}^{K} i_k \tag{4}$$

As pointed out by several authors [38] [45], symmetric tensors can be indexed in two different manners, related to each other by a bijective mapping, $\boldsymbol{h}$. Let $\boldsymbol{i}$ be a multi-index of dimension d whose entries vary in $\{1, 2, \dots K\}$ and are sorted in ascending order. The mapping $\boldsymbol{j} = \boldsymbol{h}(\boldsymbol{i})$ is defined as a $K-$dimensional multi-index containing the number of times every value of $\{1, 2, \dots K\}$ appears in $\boldsymbol{i}$. For instance, take $d = 3$ and $K = 4$; then $\boldsymbol{h}([1, 1, 4]) = [2, 0, 0, 1]$. For the sake of simplicity, denote $\boldsymbol{\mu} = \boldsymbol{h}^{-1}$; then conversely, if $\boldsymbol{j}$ is a $K-$dimensional multi-index, whose entries satisfy $|\boldsymbol{j}| = d$, then one can associate a unique $d-$dimensional multi-index $\boldsymbol{i} = \boldsymbol{\mu}(\boldsymbol{j})$, with entries sorted in ascending order.

Now, every symmetric tensor $\boldsymbol{G}$ of order d and dimension K can be associated with a homogeneous polynomial p of degree d in K variables as follows:

$$p(x_1, \cdots x_K) = \sum_{\boldsymbol{i}} G_{\boldsymbol{i}}\, \boldsymbol{x}^{\boldsymbol{h}(\boldsymbol{i})} \tag{5}$$

In the above expressions, some terms appear several times. For instance at order $d = 4$ and in dimension $K = 2$:

$$\begin{aligned} p(x_1, x_2) = \quad & G_{1111} x_1^4 + 4\, G_{1112} x_1^3 x_2 + 6\, G_{1122} x_1^2 x_2^2 \\ & + 4\, G_{1222} x_1 x_2^3 + G_{2222} x_2^4 \end{aligned}$$

whereas the total number of terms is $K^d = 16$. For this reason, a compact notation needs to be introduced, as shown in the above writing in the particular case $(K, d) = (2, 4)$:

$$p(x_1, \cdots x_K) = \sum_{|\boldsymbol{j}|=d} c(\boldsymbol{j})\, G_{\boldsymbol{\mu}(\boldsymbol{j})}\, \boldsymbol{x}^{\boldsymbol{j}} \tag{6}$$

where $c(\boldsymbol{j})$ denotes the multinomial coefficient:

$$c(\boldsymbol{j}) = \frac{|\boldsymbol{j}|!}{(\boldsymbol{j})!} \tag{7}$$

For instance, in the binary case, $c(4, 0) = 1$, $c(3, 1) = 4$, and $c(2, 2) = 6$; notice that in the binary case, the multi-index $\boldsymbol{j}$ is necessarily of the form $(j, d-j)$ so that one can denote $c(\boldsymbol{j}) \equiv c(j)$, with some abuse of notation. In addition, $c(j) = \binom{d}{j} = \frac{d!}{j!\,(d-j)!}$.

With this notation, any homogeneous polynomial of the form (5) can be written compactly without redundancy:

$$p(\boldsymbol{x}) = \sum_{|\boldsymbol{j}|=d} c(\boldsymbol{j}) \gamma(\boldsymbol{j}; p)\, \boldsymbol{x}^{\boldsymbol{j}} \tag{8}$$

where by definition $\gamma(\boldsymbol{j}; p) = G_{\boldsymbol{\mu}(\boldsymbol{j})}$. The above notation is widely utilized for quantics, and enjoys some useful properties [24], one of them being the *apolar* scalar product, as explained in the next section.

A.5 Linear spaces and scalar products

The set of symmetric tensors of order d and dimension K forms a linear space of dimension [24]:

$$D(K;d) = \left(\begin{array}{c} K+d-1 \\ d \end{array} \right) \tag{9}$$

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

Now the last ingredient we need is a scalar product. Let $\boldsymbol{P}$ and $\boldsymbol{Q}$ be two symmetric tensors of order d and dimension K. Then define the Froebenius scalar product:

$$\langle \boldsymbol{P}, \boldsymbol{Q} \rangle = \sum_{\boldsymbol{i}} {P_{\boldsymbol{i}}}^* \, Q_{\boldsymbol{i}}$$

and the induced Euclidian norm.

Next, let p and q be two homogeneous polynomials of degree d in K variables, $p(\boldsymbol{x}) = \sum_{|\boldsymbol{i}|=d} c(\boldsymbol{i})\, \gamma(\boldsymbol{i};p)\, \boldsymbol{x}^{\boldsymbol{i}}$, and $q(\boldsymbol{x}) = \sum_{|\boldsymbol{i}|=d} c(\boldsymbol{i}) \gamma(\boldsymbol{i};q)\, \boldsymbol{x}^{\boldsymbol{i}}$. Then define their scalar product as:

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

This definition corresponds to the so-called *apolar* scalar product [29], divided by $d!$. This choice may seem strange and arbitrary, but it is justified by the property below. Let $L(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x}$ be a linear form. Then, $L^d(\boldsymbol{x}) = \sum_{|\boldsymbol{j}|=d} c(\boldsymbol{j})\, \boldsymbol{a}^{\boldsymbol{j}}\, \boldsymbol{x}^{\boldsymbol{j}}$, and:

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

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

Lastly, note that with these definitions, the scalar product between two quantics coincides with that of the associated tensors.

B. Sylvester's theorem

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

$$p(x,y) = \sum_{j=1}^{r} \lambda_j\, (\alpha_j\, x + \beta_j\, y)^d, \tag{12}$$

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

$$\left[\begin{array}{cccc} \gamma_0 & \gamma_1 & \cdots & \gamma_r \\ \vdots & & & \vdots \\ \gamma_{d-r} & \cdots & \gamma_{d-1} & \gamma_d \end{array} \right] \boldsymbol{g}^* = \boldsymbol{0}. \tag{13}$$

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 [24] [46] [47].

Proof: Let's prove the forward assertion, and assume (12) is true. Then, define vector $\boldsymbol{g}$ via the coefficients of the polynomial $q(x,y)$:

$$q(x,y) \stackrel{\text{def}}{=} \prod_{j=1}^{r} ({\beta_j}^*\, x - {\alpha_j}^*\, y). \tag{14}$$

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 (12) 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, by construction (14) of q, 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 section III-A. ■

C. Cumulants

C.1 Definition

Let $\boldsymbol{z}$ be a random variable of dimension K, with components z_i. Then its moment and cumulant tensors of order d are defined as [38]:

$$\begin{array}{rcl} \mathcal{M}^{\boldsymbol{z}}_{i_1 i_2 .. i_d} & = & \mathrm{E}\{z_{i_1} z_{i_2} \dots z_{i_d}\} \\ \mathcal{C}^{\boldsymbol{z}}_{i_1 i_2 .. i_d} & = & \mathrm{Cum}\{z_{i_1}, z_{i_2}, \dots z_{i_d}\} \end{array}$$

When the moment tensors of order less than or equal to d exist and are finite, the cumulant tensor of order d exists and is finite. Whereas moments are the coefficients of the expansion of the first characteristic function, $\Phi^{\boldsymbol{z}}(\boldsymbol{u}) = \mathrm{E}\{exp(\jmath \Re\{\boldsymbol{u}^{\mathrm{H}} \boldsymbol{z}\})\}$, about the origin (the dotless $\jmath$ denotes the square root of -1, and $\Re$ the real part), cumulants are those of the second characteristic function, $\Psi^{\boldsymbol{z}}(\boldsymbol{u}) = \log(\Phi^{\boldsymbol{z}}(\boldsymbol{u}))$. $\Phi^{\boldsymbol{z}}(\boldsymbol{u})$ is everywhere continuous, and equals 1 at the origin; consequently, $\Psi^{\boldsymbol{z}}(\boldsymbol{u})$ always exists in a neighborhood of the origin [45] [7]. Note that for complex random variables, it suffices to consider the joint distribution of their real and imaginary parts [7]. It turns out that moments and cumulants enjoy the multilinearity property (2) and may be considered as tensors [38].

One important property of cumulant tensors is the following: if at least two variables, or groups of variables, among $\{z_1, ..z_K\}$ are statistically independent, then all cumulants involving these variables are null. For instance, if all the z_i are mutually independent, then:

$$\mathcal{C}^{\boldsymbol{z}}_{ij..\ell} = \delta(i,j,..\ell)\, \mathcal{C}^{\boldsymbol{z}}_{ii..i} \tag{15}$$

where the Kronecker δ equals 1 when *all* its arguments are equal, and is null otherwise. This property is not enjoyed by moments, hence the interest in cumulants. The converse is not true, as we shall subsequently see on a nice

simple example in section V: dependent random variables can have (a finite number of) null cross-cumulants. See [45] [48] [38] [7] for further properties of cumulants.

C.2 Input-Output relations

Random variables x_i, with finite cumulants, are statistically independent if all their cross-cumulants vanish [38]. Denote by $\boldsymbol{G}$ the fourth order cumulant tensor of $\boldsymbol{y}$, and by $\kappa(p)$ that of x_p. Then, we should have

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

Yet, cumulants satisfy the multilinearity property [38], so that (1) implies, if noise is Gaussian:

$$G_{ijk\ell} = \sum_{p=1}^{P} \kappa(p)\, A_{ip}\, A_{jp}\, A_{kp}\, A_{\ell p} \tag{16}$$

Now denoting by $\boldsymbol{a}(p)$ the pth column of $\boldsymbol{A}$, it is easily seen that (16) can be written as a *symmetric* CAND:

$$\boldsymbol{G} = \sum_{p=1}^{\omega} \kappa(p)\, \boldsymbol{a}(p) \circ \boldsymbol{a}(p) \circ \boldsymbol{a}(p) \circ \boldsymbol{a}(p) \tag{17}$$

The minimal number ω of factors coincides with the number of sources, P, if all of them have a non-zero marginal cumulant [25] [21] [4]. This shows more explicitly that (i) Independent Component Analysis (ICA) can be seen as a *symmetric version* of CAND [46], and (ii) decomposition (17) completely accounts for the underlying linear model and independence at order 4 between sources, in the presence of Gaussian noise.

In the remainder, we shall be interested mainly in this symmetric decomposition, with $d = 4$, and with moderate values of the dimension, K. In particular, in downlink mobile communications, it is realistic to assume that the receiver diversity will range between $K = 2$ and $K = 4$. In fact current equipment offers only $K = 1$ antenna, but it is reasonable to assume that $K = 2$ is available, by exploiting either spatial, bandwidth (oversampling), or polarization diversities. In (17), $K-$dimensional vectors $\boldsymbol{a}(p)$ account for this diversity by the fact that they are not mutually collinear; this condition is required to meet identifiability [1].

III. Blind identification of the mixture

A. Algorithm with 2 sensors and 3 sources

A.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 $\boldsymbol{g}$ 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 $\boldsymbol{g}$ satisfying (13), but there are two of them when d is even. In the latter case, any linear combination of the two vectors yields a Canonical Decomposition; in other words, the dimensionality of the variety of solutions is 1 (as shown in entry $(d, K) = (4, 2)$ of table II). In order to fix this indeterminacy, the idea proposed is to use another tensor, which should admit a related decomposition, as explained in the next section.

A.2 Choice of two cumulant tensors

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 (this has been already seen in Sylvester's theorem, and will be emphasized in section III-C). 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 (1), denote $\boldsymbol{y}$ the random variable representing the $K-$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:

$$\begin{array}{l} G_{ijk\ell} = \mathrm{Cum}\{y_i, y_j, y_k, y_\ell\};\ \tilde{G}_{ijk\ell} = \mathrm{Cum}\{y_i, y_j, y_k^*, y_\ell^*\}; \\ \tilde{\tilde{G}}_{ijk\ell} = \mathrm{Cum}\{y_i, y_j, y_k, y_\ell^*\} \end{array} \tag{18}$$

Again because of conditioning, the third cumulant tensor is not of appropriate use. But the two others can be fully exploited. In fact, assume that model (1) is satisfied. Then we have, as in (16):

$$\begin{array}{l} G_{ijk\ell} = \sum_{m=1}^{r} \kappa_m\, A_{im} A_{jm} A_{km} A_{\ell m}; \\ \tilde{G}_{ijk\ell} = \sum_{m=1}^{r} \tilde{\kappa}_m\, A_{im} A_{jm} A_{km}^* A_{\ell m}^*, \end{array} \tag{19}$$

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 tensors 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 $\boldsymbol{G}$, up to some indeterminacies, and then to use the second tensor, $\tilde{\boldsymbol{G}}$, to fix it. Of course, for real mixtures and sources, the two tensors coincide, and the algorithm

does not work. It does not work either if $\boldsymbol{G}$ is null, which occurs for sources that are circular at order 4, like PSK-8 for instance.

A.3 Numerical algorithm

Given a finite set of samples $\{\boldsymbol{y}(t),\ 0 \leq t \leq T\}$,
1. Compute the two 4th order sample cumulant tensors as follows, $\{1 \leq i,j,k,\ell \leq 2\}$:

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

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, $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$.
4. Associate the 4-way array $\tilde{\boldsymbol{G}}$ 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 $\tilde{\boldsymbol{g}}$ the corresponding 35-dimensional vector of coordinates.
5. For $\theta \in [0,\ \pi)$ and $\varphi \in [0,\ 2\pi)$, do:
 (a) Compute $\boldsymbol{g}(\theta,\varphi) = \boldsymbol{v}_1 \cos\theta + \boldsymbol{v}_2 \sin\theta\, e^{\jmath\varphi}$
 (b) Compute the 3 linear forms $L_1(\boldsymbol{x}|\theta,\varphi)$, $L_2(\boldsymbol{x}|\theta,\varphi)$, $L_3(\boldsymbol{x}|\theta,\varphi)$, associated with $\boldsymbol{g}(\theta,\varphi)$.
 (c) Express $|L_r(\boldsymbol{x}|\theta,\varphi)|^4$, $r = \{1,2,3\}$ in the 35-dimensional linear space, by three vectors $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{u}_3$.
Enddo
6. Detect the values (θ_o,φ_o) for which the vector $\tilde{g}$ is closest to the linear space spanned by $[\boldsymbol{u}_1,\ \boldsymbol{u}_2,\ \boldsymbol{u}_3]$
7. 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.
At this point, some comments are useful. Steps 1, 2, 3, 5 generate admissible vectors, $\boldsymbol{g}$. The 3 linear forms in Step 6 are computed by rooting the polynomial defined by coefficients g_ℓ. In fact, from (14), one obtains α_j and β_j by factorizing the polynomial $\sum_\ell g_\ell x^\ell$; and from (12), these α_j and β_j are the coefficients of the linear forms. Steps 4, 7, 8 measure the distance between $\tilde{g}$ and the linear space spanned by the 3 linear forms raised to the fourth power, and allow to compute coefficients λ_i. The minimal distance is found by exhaustive search of the (θ,φ) set, thanks to a loop. A matlab code of the complete algorithm can be down-loaded from the URL `www.i3s.unice.fr/~comon/spie98.htm`.

B. 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

$$\boldsymbol{A} = \begin{bmatrix} 0.81 + 0.39\jmath & 0 & 0.35 + 0.35\jmath \\ 0 & 0.5 - 0.86\jmath & 0.86 \end{bmatrix}$$

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 $\hat{\boldsymbol{A}}$, and the actual mixing matrix $\boldsymbol{A}$, 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 $||\boldsymbol{A} - \hat{\boldsymbol{A}} \cdot \boldsymbol{D}||$, for the best matrix $\boldsymbol{D}$, 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 $\Upsilon(\boldsymbol{u},\boldsymbol{v})$ between two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, invariant up to a multiplicative complex number. For this purpose, define

$$\Upsilon(\boldsymbol{u},\boldsymbol{v}) = \underset{z}{\mathrm{Min}}\ \frac{||\boldsymbol{u} - z\,\boldsymbol{v}||^2}{||\boldsymbol{u}|| \cdot ||\boldsymbol{v}||}$$

It can be seen that if $\boldsymbol{u}^{\mathsf{H}}\boldsymbol{v} = ||\boldsymbol{u}|| \cdot ||\boldsymbol{v}|| \cdot \cos\theta\, e^{\jmath\psi}$, then the minimal distance $\Upsilon(\boldsymbol{u},\boldsymbol{v})$ is reached for $||\frac{\boldsymbol{u}}{||\boldsymbol{u}||} - \frac{\boldsymbol{v}}{||\boldsymbol{v}||}\, e^{-\jmath\psi}||$ 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:

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

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 [2].

In figure 1 the average gap obtained over 15 independent noiseless experiments is plotted, for *finite* data lengths. 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 tensors are asymptotically insensitive to Gaussian noise (for large T), providing some robustness to the method.

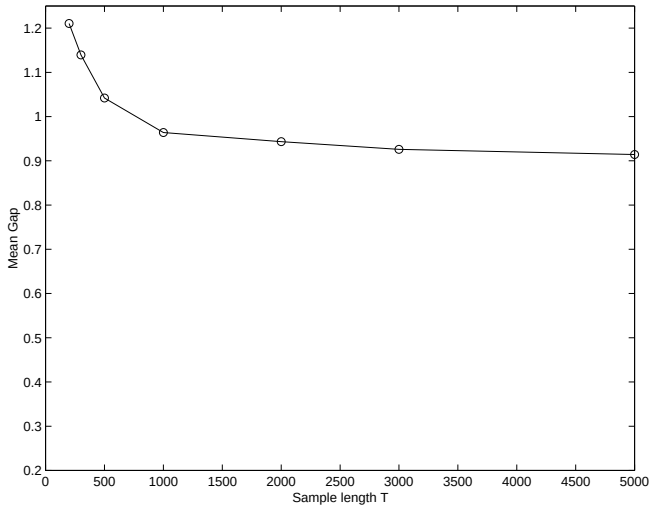


Figure 1: Blind identification. Mean Gap obtained over 15 snapshots and data lengths ranging from 200 to 5000 samples.

C. Extension to larger dimensions

For dimensions $K > 2$, Sylvester's theorem cannot apply. For instance, in its constructive proof, the fact that $d \geq r$ was used; but it turns out that this property is true only in the particular case of binary quantics, as we shall point out in this section. Instead, one should resort to Lasker-Wakeford theorem [43] [24], 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* (or *width* according to Reznick [28]). More precisely, for $2 \leq n \leq 8$, it has been shown [24] that the generic value r of the tensor rank ω is given by table I.

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

Table I: Generic rank r of symmetric tensors as a function of dimension K and order d.

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 [24], as reported in table II.

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

Table II: Generic dimensionality of the variety of solutions for the CAND of symmetric tensors, as a function of dimension K and order d.

One can check out, for instance, that for $K = 2$, 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 III-A.1: a second 4-way cumulant array had been necessary in order to fix the extraneous parameter. It is interesting to notice that the number of solutions is often finite for 3-way arrays; on the contrary for 4-way arrays, it occurs only for $K = 7$ in the above table! Before to close this section, it is worth insisting that there is no simple rule or formula that would yield all the values of table I: the result of several theorems of various origins allow to compute half of the values, but a numerical algorithm is still required to fill table I completely [24]. Eventually, there exist other formulations of the Sylvester theorem in the complex case, and one could take advantage of them for deriving new algorithms [47].

IV. Source extraction from a known mixture

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

The Maximization of the A Posteriori (MAP) distribution is quite natural for recovering discrete inputs, but is generally iterative and requires an (almost) exhaustive search [36] [49] [30] [37]. Actually, this idea of extracting sources in the presence of lack of diversity is certainly not new, and can be traced back to the Viterbi algorithm [50], in which the discrete character is fully exploited, and the most probable sequence is searched for among a set of admissible candidates, which is adaptively reduced. As a consequence, it is not hard to see that for sufficiently low noise and sufficiently long observations, one can extract an arbitrarily large number of sources from a single observation [35].

Our contribution is here different: the goal is to devise an entirely analytical algorithm that does not resort to a computationally heavy search. It can be used either as a means to extract the sources, or as an initial guess for an ascent algorithm maximizing the *a posteriori* distribution.

A. Algorithms

As pointed out in [15], among others, cumulant-based criteria [2] [51] [52] obviously do not carry all the information contained in the discrete character. On the other hand, **all discrete distributions** in the complex plane can be entirely characterized by a polynomial equation in two variables (the real and imaginary parts). For some discrete distributions, this equation even turns out to be a polynomial in the complex variable [53]. For example, for $M-$PSK modulated sources [54], we have that $x^M = 1$. There are M solutions in the complex field, and hence a set of M allowed values. By adding to the K observation equations (1) all the $\binom{K+M}{M+1}$ homogeneous monomials of degree $M+1$, one gives oneself the opportunity to use the P additional equations $x_i^M = 1$, $1 \leq i \leq P$. The system can

be solved thanks to these additional equations. In order to illustrate the idea, we shall now consider the easiest case.

A.1 BPSK sources

This augmented linear system becomes especially simple when enough linear equations can be extracted. In particular, this is the case if $M = 2$ and $K = 2$. In fact we have then $P+6$ equations in P unknowns. The first 2 equations are given by (1). 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 P 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 $\binom{P}{3}$ terms of the form $x_i x_j x_k$.

Now take our example of section III, where $P = 3$. 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:

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix} = \boldsymbol{C}\,\bar{\boldsymbol{x}},\ \bar{\boldsymbol{x}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_1\, x_2\, x_3 \end{pmatrix},$$

$$\text{where} \quad \boldsymbol{z} = \begin{pmatrix} y_1^3 \\ y_1^2\, y_2 \\ y_1\, y_2^2 \\ y_2^3 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{C} = \begin{pmatrix} \boldsymbol{A} & & 0 \\ & \boldsymbol{B} & \end{pmatrix}$$

Since the 4×4 matrix $\boldsymbol{B}$ is a known function of $\boldsymbol{A}$, it is given as soon as $\boldsymbol{A}$ is known (see appendix VII-A). 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, so that the sources are given by:

$$\hat{\boldsymbol{x}} = \boldsymbol{C}^{-}(1:3, 1:6) \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix}$$

where $\boldsymbol{C}^{-}(1:3, 1:6)$ denotes the 3×6 matrix formed of the 3 first lines of the Moore-Penrose pseudo-inverse of $\boldsymbol{C}$.

A.2 MSK sources

It is well known that the Gaussian Minimum Shift Keying (GMSK) modulation, utilized in the GSM standard, can be approximated by a Minimum Shift Keying (MSK) modulation [55], which in turn can be viewed as a QPSK modulation with transition constraints [54]. This motivates the study of the MSK source separation problem.

Similarly to the previous subsection, one can show (cf. appendix VII-B) that any odd-degree monomial function of the 2 observations is a linear combination of 4 sources:

$$\begin{aligned} \epsilon^n\, y_1(n)^k y_2(n)^\ell = {} & b_1(k,\ell)\, x_1(n) + b_2(k,\ell)\, x_2(n) \\ & + b_3(k,\ell)\, x_3(n) + b_4(k,\ell)\, x_4(n) \end{aligned} \qquad (20)$$

where the coefficients b_i are known functions of k, ℓ, and $\boldsymbol{A}$, $\epsilon = -1$ if $k+\ell = 3$ modulo 4, and $\epsilon = 1$ if $k+\ell = 1$ modulo 4. The sources can then be recovered along the same lines as in section IV-A.1. It is also shown in appendix VII-B that the extraneous source $x_4(n) = (-1)^n\, x_1(n) x_2(n) x_3(n)$ is an MSK source, uncorrelated with the actual ones, but this is not used in the present section.

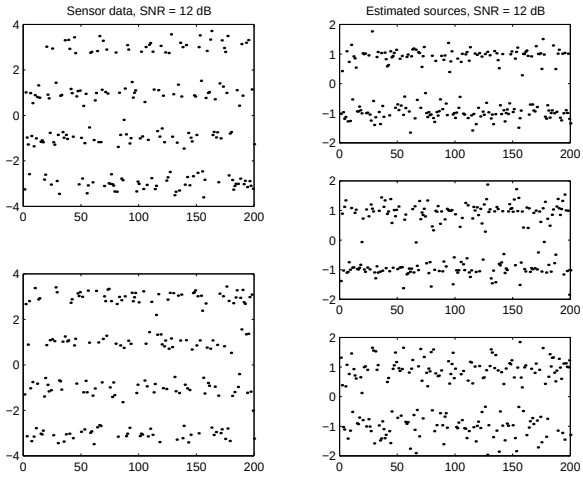


Figure 2: Inversion of a known under-determined mixture: typical results obtained with BPSK sources and a SNR of 12dB.

B. Computer results

Data have been generated according to the following model:

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

where x_i are BPSK distributed. 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 $\boldsymbol{A}$, 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 actual source value of source i, and $\hat{x}_i$ its estimate (before hard decision). Figure 3 (top) reports the performances obtained. Second, after hard decision, 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, bottom), which is quite satisfactory.

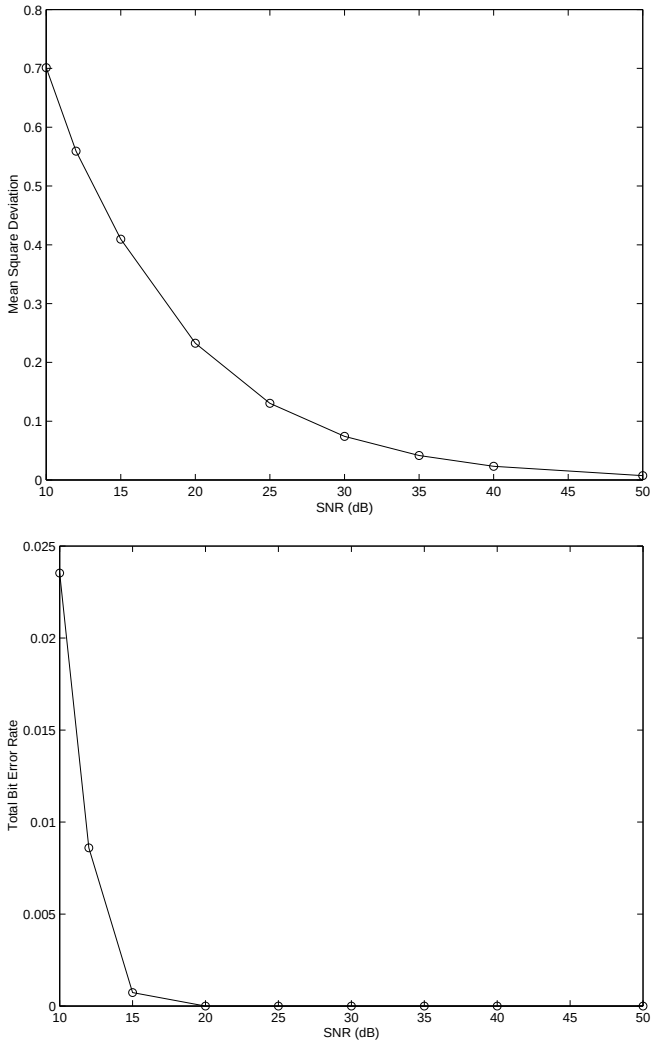


Figure 3: Extraction performances for a known mixture: mean Standard Deviation (top) and Total Bit Error Rate (bottom) for the 3 estimates.

V. Source extraction from an unknown mixture

As already pointed out, when the mixture $\boldsymbol{A}$ is unknown and sources are discrete with known alphabet, there are several algorithms available in the literature [14] [56] [11] [36] [49] [37]. Contrary to the latter algorithms, the goal is to devise a low-complexity numerical algorithm (*e.g.* without exhaustive search or clustering). Therefore, the idea is quite different: it is shown that the under-determined BSS problem is equivalent to an *invertible* BSS in *larger* dimension [57], namely:

$$\bar{\boldsymbol{y}} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix} = \boldsymbol{C}\,\bar{\boldsymbol{x}} \tag{21}$$

with the notation of the previous section. However, any standard BSS algorithm may not be used, since some virtual sources are introduced and are correlated to each other.

A. Algorithms

If the mixture is unknown, the principle of building virtual measurements is the same as in section IV, except that the source extraction is not as easy because, in addition to $\boldsymbol{A}$, matrix $\boldsymbol{B}$ (and hence $\boldsymbol{C}$) is unknown. The case of 2×3 mixtures of binary or MSK sources is now analyzed in detail.

A.1 BPSK sources

As previously, we have $P = 4$ BPSK sources (including one virtual) in the mixture. Let $x_4 = x_1 x_2 x_3$; then, since x_1, x_2, and x_3 are real i.i.d. BPSK sources, so is x_4. But x_4 is obviously not independent of the former sources. One can even stress that $\text{Cum}\{x_1, x_2, x_3, x_4\} = 1$. Nevertheless, it can be shown that $\text{E}\{x_i x_j\} = 0$, $\text{Cum}\{x_i, x_i, x_j, x_j\} = 0$, $\text{Cum}\{x_i, x_i, x_i, x_j\} = 0$, $\text{Cum}\{x_i, x_i, x_i, x_i\} = -2$, $\forall i, j \in \{1, 2, 3, 4\}$, $i \neq j$. This shows that all *pairwise* source cross-cumulants of order 2 and 4 vanish, which is sufficient for applying the ICA algorithm proposed in [2] [9]. In fact, only pairwise cumulants are utilized in order to estimate a sequence of plane rotations. As a result, in the absence of noise, the $P = 4$ sources may be estimated. Denote $\hat{\bar{\boldsymbol{x}}}$ the estimate of this augmented source vector, obtained by the BSS algorithm. The extraction of the actual source vector $\hat{\boldsymbol{x}}$ from the augmented one, $\hat{\bar{\boldsymbol{x}}}$, is addressed in section V-A.3.

A.2 MSK sources

As in section IV-A.2, let us turn to MSK-modulated sources. Such sources can be split into two *independent* BPSK sources: MSK signals are alternatively real and imaginary (up to some fixed complex phase, which is part of the inherent indetermination). In other words, when n is odd, $\boldsymbol{y}(n) = \boldsymbol{A}\,\boldsymbol{b}(n)$, and when n is even, $\boldsymbol{y}(n) = \jmath\,\boldsymbol{A}\,\boldsymbol{b}(n)$, for some BPSK i.i.d. process, $\boldsymbol{b}(n)$. To simplify the notation, define $\boldsymbol{y}'(n) = \boldsymbol{y}(n)$ for n odd, and $\boldsymbol{y}'(n) = -\jmath \boldsymbol{y}(n)$ for n even. Then one can rewrite the observation model as:

$$\boldsymbol{y}'(n) = \boldsymbol{A}\,\boldsymbol{b}(n) \tag{22}$$

Now with this writing, 3 BPSK sources remain to be found. The same ICA algorithm as in section V-A.1, working with pairwise cumulants [2], will consequently successfully separate those sources. In order to obtain an estimate of the MSK source vector, $\hat{\boldsymbol{x}}$, after BSS, it will suffice to recombine samples of the separator output: $\hat{\boldsymbol{x}}_{msk}(n) = \hat{\boldsymbol{b}}(n)$ for n odd, and $\hat{\boldsymbol{x}}(n) = \jmath\hat{\boldsymbol{b}}(n)$ for n even. The way $\hat{\boldsymbol{b}}(n)$ is obtained from BSS outputs $\hat{\bar{\boldsymbol{b}}}(n)$ is explained in section V-A.3 below.

A.3 Extraction of actual sources

In this section, we assume we have separated $\bar{P}$ BPSK sources, $\bar{P} = P + \binom{P}{3}$, and we denote by $\hat{\bar{\boldsymbol{x}}}$ the $\bar{P}-$dimensional estimated source vector. Among these $\bar{P}$ BPSK sources, $\binom{P}{3}$ are virtual, for they are cubics in the P others, by definition of $\bar{\boldsymbol{x}}$. In order to discriminate between actual and extraneous sources, one computes the multiple

coherence vector between the estimated augmented source vector, $\hat{\hat{\boldsymbol{x}}}$, and actual observation, $\boldsymbol{y}$:

$$c_p \stackrel{\text{def}}{=} \frac{1}{\sigma_p^2}\, \boldsymbol{r}_p{}^{\text{H}} \boldsymbol{R}_y^{-1} \boldsymbol{r}_p,\ 1 \leq p \leq \bar{P}, \tag{23}$$

where $\sigma_p^2 = \text{var}\{\hat{\hat{x}}_p\}$, $\boldsymbol{r}_p = \mathbf{covar}\{\boldsymbol{y}, \hat{\hat{x}}_p\}$, and $\boldsymbol{R}_y = \mathbf{covar}\{\boldsymbol{y}\}$. Recall that the multiple coherence satisfies $0 \leq c_p \leq 1$, and that it equals 1 if and only if $\hat{\hat{\boldsymbol{x}}}$ is a linear function of $\boldsymbol{y}$. The consequence is that c_p is close to zero for extraneous sources, and the P actual sources have the largest coherence (even if it does not generally reach 1, because $\boldsymbol{A}$ is not invertible). Thus, as an estimate $\hat{\boldsymbol{x}}$ of the actual source vector, it suffices to retain among the entries of $\hat{\hat{\boldsymbol{x}}}$ those P that have the largest multiple coherence, c_p. The whole processing line is sketched in figure 4.

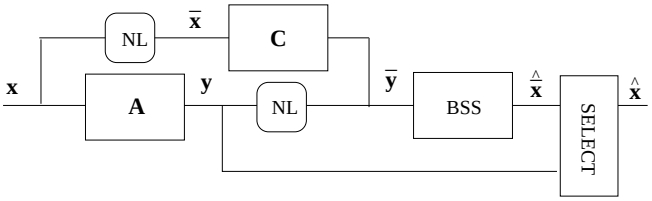


Figure 4: Processing line for the algorithm of section V.

B. Computer results

B.1 BPSK sources

The most favorable mixtures have been chosen to run the computer experiments; in fact, this is the only means to obtain mixture independent performances, comparable to the ultimate ones [35]. The mixture has thus the following structure :

$$\boldsymbol{A} = \begin{pmatrix} a_1 & ia_2 & 0 \\ 0 & ia_3 & a_4 \end{pmatrix}$$

where the values of a_i are given by: $(a_1,\ a_2,\ a_3,\ a_4) = (0.9,\ 0.42,\ 0.42,\ 0.9)$. Computer experiments have been run for data length $N = 1000$, and error rates have been estimated over 5000 snapshots. The results reported in figure 5 show good performances, compared to the ultimate BER performances plotted in dashed line.

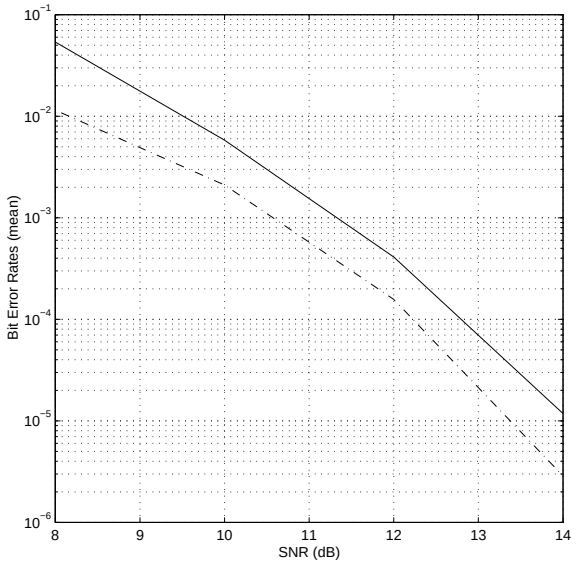


Figure 5: Separation performances for an unknown mixture: Bit Error rates (BER) for the separation of 3 BPSK sources from 2 sensors.

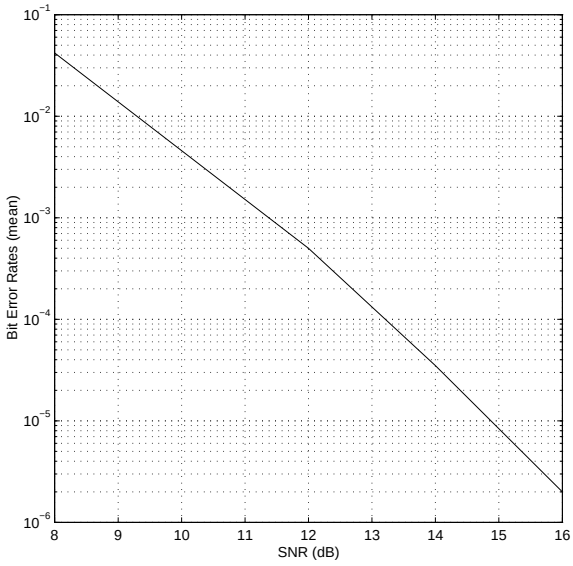


Figure 6: Separation performances for an unknown mixture: Bit Error Rates (BER) for the separation of 3 MSK sources from 2 sensors.

B.2 MSK sources

Since MSK sources can be assumed to be alternatively imaginary and real, and can be split into two independent BPSK sources as in (22), it is quite clear that the most favorable mixtures are the same as in the case of BPSK sources [35] :

$$A = \begin{pmatrix} a_1 & ia_2 & 0 \\ 0 & ia_3 & a_4 \end{pmatrix}$$

Again, the data length is $N = 1000$, and BER's are averaged over 5000 trials. The results presented in figure 6 demonstrate a good behavior of our algorithm, only 2dB above the ultimate bound given in [35].

C. Extensions

The case of QPSK sources, satisfying $x_i^4 = 1$, is more complicated, but we sketch the procedure in this section. In fact, a QAM4 source is a sum of two independent BPSK sources in quadrature:

$$x_p = \varepsilon_p + \jmath\, \varepsilon'_p$$

Hence, equation (1) can be rewritten as:

$$\begin{pmatrix} \Re[\boldsymbol{y}] \\ \Im[\boldsymbol{y}] \end{pmatrix} = \begin{pmatrix} \Re[A] & -\Im[A] \\ \Im[A] & \Im[A] \end{pmatrix} \begin{pmatrix} \epsilon \\ \epsilon' \end{pmatrix}$$

As a consequence, 6 independent BPSK inputs need now to be recovered from 4-sensor measurements.

Let us state a general result now. If we have K sensor measurements, and P sources, satisfying $x_i^M = x_i$. Then raising the measurements to the Mth power yields

$\binom{K+M-1}{M}$ additional equations, but introduces $\binom{P+M-1}{M}$ distinct source monomials, among which P are of the form $x_i^M = x_i$ and $P(P-1)$ of the form $x_i^{M-1}x_j = x_j$. Thus, the augmented linear system can be solved only if the necessary condition below is satisfied:

$$K + \binom{K+M-1}{M} \geq \binom{P+M-1}{M} - P(P-1)$$

It can be shown that if $P > K$, this inequality is satisfied only for $M \leq 3$. One can also prove that for $M = 3$, the inequality holds true only for $P \leq K+1$.

This result shows that with 6 BPSK sources, the approach followed in the previous sections cannot apply with 4-sensor measurements: by raising to the third power, one introduces 20 equations, but also 20 new unknowns. On the other hand, it works if we raise the sensor measurements to the fifth power. In fact, in that case, one gets $\binom{8}{5} = 56$ equations and introduces only 26 new unknowns (20 of the form $\varepsilon_1\varepsilon_2\varepsilon_3$ and 6 of the form $\varepsilon_1\varepsilon_2\varepsilon_3\varepsilon_4\varepsilon_5$), among the $\binom{10}{5} = 252$ source monomials, because $\varepsilon_p^4 = \varepsilon_p^2 = 1$.

The problem is generically solvable in the LS sense (under the rank conditions guaranteeing identifiability). The process of choosing the interesting 6 BPSK sources among the 32 extracted ones remains the same as in the previous cases. However, the association of two BPSK sources to form an actual QPSK source is somewhat more complicated, but can be performed by taking advantage of the mixture structure, as will be explained in a companion paper.

VI. Concluding remarks

The BSS problem in the under-determined context has been addressed, and three numerical algorithms have been described. The first performs blind identification of 2×3 under-determined mixtures, under the assumption that sources have non-zero kurtoses. The second extracts sources when the mixture is known and the sources are discrete (BPSK or MSK). The third algorithm is of very low complexity, and directly separates independent discrete (BPSK or MSK) sources. Spurious sources are introduced in the procedure, but are detected and eliminated afterwards with the help of the multiple coherence. The algorithms have been illustrated by computer simulations, mainly for 2×3 mixtures of binary sources. Extensions have been discussed for four-state sources (QPSK), but the presentation in depth is postponed to another paper.

VII. Appendix

A. Expression of $\boldsymbol{B}$ as a function of $\boldsymbol{A}$ for BPSK sources

We give below the expression of $y_i^2 y_j$ as a function of x_p, $x_1x_2x_3$, and $A_{k\ell}$. It is a straightforward application of the multilinearity property (2):

$$y_i^2 y_j = \begin{pmatrix} A_{i2}^2A_{j1} + A_{i1}^2A_{j1} + 2A_{i1}A_{i2}A_{j2} + A_{i3}^2A_{j1} + 2A_{i1}A_{i3}A_{j3} \\ A_{i2}^2A_{j2} + A_{i1}^2A_{j2} + 2A_{i1}A_{i2}A_{j1} + A_{i3}^2A_{j2} + 2A_{i2}A_{i3}A_{j3} \\ 2A_{i1}A_{i3}A_{j1} + 2A_{i2}A_{i3}A_{j2} + A_{i2}^2A_{j3} + A_{i1}^2A_{j3} + A_{i3}^2A_{j3} \\ 2A_{i2}A_{i3}A_{j1} + 2A_{i1}A_{i2}A_{j3} + 2A_{i1}A_{i3}A_{j2} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_1\,x_2\,x_3 \end{pmatrix}$$

By varying i and j in the set $\{1, 2\}$, each of the four columns of $\boldsymbol{B}$ is readily obtained.

B. Expression of the extraneous source in the MSK case

A MSK process, $x(n)$, can be defined as $x(n) = \jmath^n\, b(n)$, up to an initial phase irrelevant in our problem, where $b(n)$ is BPSK, i.e. $b(n)^2 = 1$. As a consequence, $x(n)^2 = (-1)^n$ and $x(n)^3 = (-1)^n\, x(n)$. Consider first the case where $k+\ell = 3$. Then the product $y_1(n)^k\, y_2(n)^\ell$ is a linear combination of monomials of the form $x_i(n)^3$, $x_i(n)x_j(n)^2$, $i \neq j$, and $x_1(n)x_2(n)x_3(n)$. The two first terms are of the form $(-1)^n x_i(n)$. Let us look at the third. Taking the definition, $x_1(n)x_2(n)x_3(n) = (-\jmath)^n b_1(n)b_2(n)b_3(n)$, where $b_i(n)$ are BPSK. Yet, the product of BPSK's is again a BPSK. Consequently, $x_1(n)x_2(n)x_3(n) = (-\jmath)^n b_4(n)$ for some BPSK process, $b_4(n)$, showing that $x_1(n)x_2(n)x_3(n)$ is also of the form $(-1)^n x_4(n)$, where $x_4(n)$ is MSK. As a conclusion, $(-1)^n y_1(n)^k\, y_2(n)^\ell$ is a linear combination of 4 MSK processes.

Take now a more general case where $k + \ell > 3$. Clearly, because $x_i(n)^4 = 1$, there are only two distinct cases to consider, because $k + \ell$ is odd by hypothesis: $k + \ell = 3$ and $k + \ell = 5$ modulo 4. So let us look at $k + \ell = 5$. Skipping the details, we could state by a similar reasoning that $y_1(n)^k\, y_2(n)^\ell$ is a linear combination of $x_i(n)$ and $x_1(n)x_2(n)x_3(n) = x_4(n)$, all being MSK processes.

Lastly, one can notice that the extraneous MSK process is uncorrelated (at order 2) with the 3 others. In fact, without restricting the generality, take $\mathrm{E}\{x_1(n)x_4(n)\}$. It is clear from above that this cross covariance is equal to $\mathrm{E}\{x_1(n)^2x_2(n)x_3(n)\} = (-1)^n\mathrm{E}\{x_2(n)x_3(n)\}$, which is null because $x_2(n)$ and $x_3(n)$ are independent.

Acknowledgment. I wish to thank the anonymous reviewers for their careful reading and helpful remarks, which have contributed in improving the clarity of the paper.

References

[1] X. R. CAO and R. W. LIU, "General approach to blind source separation", *IEEE Trans. Sig. Proc.*, vol. 44, no. 3, pp. 562–570, Mar. 1996.

[2] P. COMON, "Independent Component Analysis, a new concept ?", *Signal Processing, Elsevier*, vol. 36, no. 3, pp. 287–314, Apr. 1994, Special issue on Higher-Order Statistics.

[3] T. W. LEE, *Independent Component Analysis*, Kluwer, 1998.

[4] P. COMON, "Tensor decompositions", in *Mathematics in Signal Processing V*, J. G. McWhirter and I. K. Proudler, Eds., pp. 1–24. Clarendon Press, Oxford, UK, 2002.

[5] A. HYVÄRINEN, J. KARHUNEN, and E. OJA, *Independent Component Analysis*, Wiley, 2001.

[6] A. J. BELL and T. J. SEJNOWSKI, "An information-maximization approach to blind separation and blind deconvolution", *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.

[7] J. L. LACOUME, P. O. AMBLARD, and P. COMON, *Statistiques d'ordre supérieur pour le traitement du signal*, Collection Sciences de l'Ingénieur. Masson, 1997.

[8] J. F. CARDOSO, "High-order contrasts for independent component analysis", *Neural Computation*, vol. 11, no. 1, pp. 157–192, Jan. 1999.

[9] P. COMON, "From source separation to blind equalization, contrast-based approaches", in *Int. Conf. on Image and Signal Processing (ICISP'01)*, Agadir, Morocco, May 3-5, 2001, invited plenary.

[10] P. LOUBATON, E. MOULINES, and P. A. REGALIA, "Subspace methods for blind identification and deconvolution", in *Signal Processing Advances in Wireless and Mobile Communications*, Giannakis, Hua, Stoica, and Tong, Eds., chapter 3. Prentice-Hall, 2001.

[11] A. J. van der VEEN, "Analytical method for blind binary signal separation", *IEEE Trans. Sig. Proc.*, vol. 45, no. 4, pp. 1078–1082, Apr. 1997.

[12] O. GRELLIER and P. COMON, "Analytical blind discrete source separation", in *Eusipco*, Tampere, Finland, 5-8 sept. 2000.

[13] E. GASSIAT and F. GAMBOA, "Source separation when the input sources are discrete or have constant modulus", *IEEE Trans. Sig. Proc.*, vol. 45, no. 12, pp. 3062–3072, Dec. 1997.

[14] A. KANNAN and V. U. REDDY, "Maximum likelihood estimation of constellation vectors for blind separation of co-channel BPSK signals and its performance analysis", *IEEE Trans. Sig. Proc.*, pp. 1736–1741, 1997.

[15] D. YELLIN and B. PORAT, "Blind identification of FIR systems excited by discrete-alphabet inputs", *IEEE Trans. Sig. Proc.*, vol. 41, no. 3, pp. 1331–1339, 1993.

[16] S. TALWAR, M. VIBERG, and A. PAULRAJ, "Blind estimation of multiple co-channel digital signals arriving at an antenna array: Part I, algorithms", *IEEE Trans. Sig. Proc.*, pp. 1184–1197, May 1996.

[17] G. M. BERGMAN, "Ranks of tensors and change of base field", *Journal of Algebra*, vol. 11, pp. 613–621, 1969.

[18] J. D. CAROLL and J. J. CHANG, "Analysis of individual differences in multidimensional scaling via n-way generalization of Eckart-Young decomposition", *Psychometrika*, vol. 35, no. 3, pp. 283–319, Sept. 1970.

[19] J. B. KRUSKAL, "Three-way arrays: Rank and uniqueness of trilinear decompositions", *Linear Algebra and Applications*, vol. 18, pp. 95–138, 1977.

[20] 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*, vol. 45, no. 1, pp. 3–24, Mar. 1980.

[21] L. DeLATHAUWER, *Signal Processing based on Multilinear Algebra*, Doctorate, Katholieke Universiteit Leuven, Sept. 1997.

[22] R. BRO, "Parafac, tutorial and applications", *Chemom. Intel. Lab. Syst.*, vol. 38, pp. 149–171, 1997.

[23] P. COMON, "Independent component analysis, and the diagonalization of symmetric tensors", in *European Conference on Circuit Theory and Design – ECCTD*, H. Dedieu, Ed., Davos, Aug 30-Sept 3, 1993, pp. 185–190, Elsevier.

[24] P. COMON and B. MOURRAIN, "Decomposition of quantics in sums of powers of linear forms", *Signal Processing, Elsevier*, vol. 53, no. 2, pp. 93–107, Sept. 1996.

[25] J. F. CARDOSO, "Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors", in *Proc. ICASSP*, Toronto, 1991, pp. 3109–3112.

[26] L. DeLATHAUWER and B. DeMOOR, "From matrix to tensor: Multilinear algebra and signal processing", in *Mathematics in Sig. Proc., IMA Conf. Series*, Warwick, Dec 17-19 1996, Oxford Univ Press.

[27] E. KOFIDIS and P. A. REGALIA, "Tensor approximation and signal processing applications", in *AMS Conf. on Structured Matrices in Operator theory, Numerical Analysis, Control, Signal and Image Processing*. 2000, AMS Publ.

[28] B. REZNICK, "Sums of even powers of real linear forms", *Memoirs of the AMS*, vol. 96, no. 463, Mar. 1992.

[29] R. EHRENBORG and G. C. ROTA, "Apolarity and canonical forms for homogeneous polynomials", European Journal of Combinatorics, 1993.

[30] T. W. LEE, M. S. LEWICKI, et al., "Blind source separation of more sources than mixtures using overcomplete representations", *IEEE Sig. Proc. Letters*, vol. 6, no. 4, pp. 87–90, Apr. 1999.

[31] L. TONG, "Identification of multichannel MA parameters using higher-order statistics", *Signal Processing, Elsevier*, vol. 53, no. 2, pp. 195–209, Sept. 1996, special issue on High-Order Statistics.

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

[33] A. TALEB, "An algorithm for the blind identification of n independent signals with 2 sensors", in *Sixth International Symposium on Signal Processing and its Applications (ISSPA'01)*. 2001, IEEE.

[34] S-I. AMARI and H. NAGAOKA, *Methods of Information Geometry*, Am. Math. Soc., 2000.

[35] O. GRELLIER and P. COMON, "Performance of blind discrete source separation", in *Eusipco*, Rhodes, Greece, Sept 8–11 1998, vol. IV, pp. 2061–2064, invited session.

[36] P. PAJUNEN, "Blind separation of binary sources witrh less sensors than sources", in *Int. Conf. Neural Networks*, Houston, Texas, June 1997, pp. 1994–1997.

[37] K. I. DIAMANTARAS and E. CHASSIOTI, "Blind separation of N binary sources from one observation: a deterministic approach", in *2nd Int. Work. Indep. Comp. Ana. Blind Source Sep. (ICA'2000)*, Helsinki, Finland, June 19-22 2000, pp. 93–98.

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

[39] K. V. MARDIA, J. T. KENT, and J. M. BIBBY, *Multivariate Analysis*, Academic Press, 1979.

[40] K. V. PARSHALL, "The one-hundred anniversary of the death of invariant theory", *The Mathematical Intelligencer*, vol. 12, no. 4, pp. 10–16, 1990.

[41] G. SALMON, *Lessons introductory to the Modern Higher Algebra*, Chesla publ., New York, 1885.

[42] J. A. DIEUDONNÉ and J. B. CARREL, "Invariant theory, old and new", in *Computing Methods in Optimization Problems-2*, San Remo, Italy, Sept 1968, 1970, pp. 1–80, Academic Press.

[43] E. K. WAKEFORD, "Apolarity and canonical forms", in *Proceedings of the London Mathematical Society*, 1918, pp. 403–410.

[44] B. REZNICK, "Sums of powers of complex linear forms", Preprint, private correspondence, Aug. 1992.

[45] M. KENDALL and A. STUART, *The Advanced Theory of Statistics, Distribution Theory*, vol. 1, C. Griffin, London, 1977.

[46] P. COMON, "Blind channel identification and extraction of more sources than sensors", in *SPIE Conference*, San Diego, July 19-24 1998, pp. 2–13, keynote address.

[47] L. DeLATHAUWER, P. COMON, and B. DeMOOR, "ICA algorithms for 3 sources and 2 sensors", in *Sixth Sig. Proc. Workshop on Higher Order Statistics*, Caesarea, Israel, June 14–16 1999, pp. 116–120.

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

[49] O. GRELLIER and P. COMON, "Blind separation of discrete sources", *IEEE Signal Processing Letters*, vol. 5, no. 8, pp. 212–214, Aug. 1998.

[50] G. D. FORNEY, "The Viterbi algorithm", *Proceedings of the IEEE*, vol. 61, no. 3, Mar. 1973, pp. 268–278.

[51] A. MANSOUR and C. JUTTEN, "Fourth-order criteria for blind source separation", *IEEE Trans. Sig. Proc.*, vol. 43, no. 8, pp. 2022–2025, Aug. 1995.

[52] E. MOREAU and N. THIRION, "Non symmetrical contrasts for sources separation", *IEEE Trans. Signal Processing*, vol. 47, no. 8, pp. 2241–2252, August 1999.

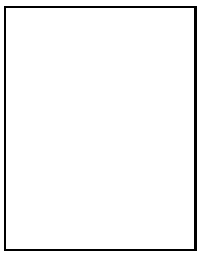
[53] O. GRELLIER, P. COMON, B. MOURRAIN, and P. TREBUCHET, "Analytical blind channel identification", *IEEE Trans. Signal Processing*, vol. 50, no. 9, pp. 2196–2207, Sept. 2002.

[54] J. G. PROAKIS, *Digital Communications*, McGraw-Hill, 1995, 3rd edition.

[55] P. A. LAURENT, "Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses", *IEEE Trans. Com.*, vol. 34, pp. 150–160, Feb. 1986.

[56] T. H. LI and K. MBAREK, "A blind equalizer for nonstationary discrete-valued signals", *IEEE Trans. Sig. Proc.*, vol. 45, no. 1, pp. 247–254, Jan. 1997, Special issue on communications.

[57] P. COMON and O. GRELLIER, "Non linear inversion of underdetermined mixtures", in *ICA99, IEEE Workshop on Indep. Comp. Anal. and Blind Source Separation*, Aussois, France, Jan 11–15 1999, pp. 461–465.



Pierre Comon (M'87 - SM'95) graduated in 1982, and received the Doctorate degree in 1985, both from the University of Grenoble, France. He later received the Habilitation to lead Researches in 1995, from the University of Nice. He has been for nearly 13 years in industry, first with Crouzet-Sextant, Valence, France, between 1982 and 1985, and then with Thomson Marconi, Sophia-Antipolis, France, between 1988 and 1997. He spent year 1987 with the ISL laboratory, Stanford University, CA. He joined Eurecom Institute, Sophia-Antipolis, France, in 1997 and left fall 1998. He was an Associate Research Director with CNRS from 1994 to 1998. He is now Research Director at CNRS since 1998 at laboratory I3S, Sophia-Antipolis, France. His research interests include High-Order statistics, Blind deconvolution and Equalization, Digital communications, and Statistical Signal and Array Processing.

Dr. Comon was Associate Editor of the IEEE Transactions on Signal Processing from 1995 to 1998, and a member of the French National Committee of scientific research from 1995 to 2000. He was the coordinator of the European Basic Research Working Group ATHOS, from 1992 to 1995. Between 1992 and 1998, he was a member of the Technical and Scientific Council of the Thomson Group. Since July 2001, he acts as launching Associate Editor with the IEEE Transactions on Circuits and Systems I, in the area of Blind Techniques; he is currently IEEE Distinguished Lecturer for the period 2002-2003.