IMA Conf. mathematics in Signal Processing, Warwick, UK, Dec. 18-20, 2000

TENSOR DECOMPOSITIONS
STATE OF THE ART AND APPLICATIONS

Pierre COMON

I3S Laboratory, Algorithmes-Euclide-B, 2000 route des Lucioles
BP 121, F-06903 Sophia-Antipolis cedex, France
comon@unice.fr www.i3s.unice.fr/~comon

ABSTRACT

In this paper, we present a partial survey of the tools borrowed from tensor algebra, which have been utilized recently in Statistics and Signal Processing. It is shown why the decompositions well known in linear algebra can hardly be extended to tensors. Numerical algorithms have nevertheless been developed, and some are reported here, but their limitations are emphasized. These reports hopefully open research perspectives for enterprising readers.

1. INTRODUCTION

Applications. The decomposition of arrays of order higher than 2 has proven to be useful in a number of applications. The most striking case is perhaps *Factor Analysis*, where statisticians early identified difficult problems, tackling the limits of linear algebra. The difficulty lies in the fact that such arrays may have more factors than their dimensions. Next, data are often arranged in many-way arrays, and the reduction to 2-way arrays sometimes results in a loss of information. Lastly, the solution of some problems, including the so-called *Blind Source Separation* (BSS) generally requires the use of High-Order Statistics (HOS), which are intrinsically tensor objects [39]. When second order statistics suffice to establish identifiability, the corresponding algorithms are quite sensitive to model uncertainties [37], so that the complementary use of HOS statistics is often still recommended.

BSS finds applications in Sonar, Radar [9], Electrocardiography [21], Speech [41] [33] [16], and Telecommunications [22] [23] [45] [8] [26], among others. In particular, the surveillance of radio-communications in the civil context, or interception and classification in military applications, resort to BSS. Moreover, in *Mobile Communications*, the mitigation of interfering users and the compensation for channel fading effects are now devised with the help of BSS; this is closely related to the general problem of *Blind Deconvolution*.

High-Order Factor Analysis is applied in many areas including Economy, Psychology [6] [7], Chemometrics [24], and Sensor Array Processing [13] [45] [43]. Other fields where array decompositions can turn out to be useful include Exploratory Analysis [29], Complexity Analysis [32], and Sparse Coding [28].

Bibliographical survey. Bergman [1] was probably the first to notice that the concept of *rank* was difficult to extend from matrices to higher order arrays. Caroll [6] provided the first *canonical decomposition* algorithm of a three-way array, later referred to as *Candecomp* model. Several years later, Kruskal [32] conducts a detailed analysis of uniqueness, and related several definitions of rank. The algorithm *Candelinc* [7] was devised by Caroll and others in the eighties; it allowed to compute a canonical decomposition subject to a priori linear constraints.

Leurgans [35] derived sufficient identifiability conditions for the 3-way array decomposition; as opposed to Kruskal, his proof was constructive and yielded a numerical algorithm running several matrix SVD's.

Instead of finding an exact decomposition of a $d-$way array, which requires more than d terms (as we shall subsequently see), Comon proposed in [10] to approximately decompose it into d terms. The problem was then reduced to finding an invertible linear transform (change of coordinates); see [11] [5] and references therein. This decomposition is now referred to as "*Independent Component Analysis*" (ICA), whereas the exact *Canonical Decomposition* is sometimes referred to as underdetermined or overcomplete ICA [33] [13]. The terminology of ICA is meaningful in the context of Signal Processing and BSS [30] [11] [4]. Constructive algorithms for ICA either proceed by sweeping the pairs of indices [11] [5], or are of iterative nature, like power methods [18] [31], gradient descents [38], or Robbins-

[1]This work has been supported in part by the Rnrt project "Paestum". The author wishes to thank P.Whitaker for proof-reading part of this paper.

Monro algorithms [30] [41]. Some less efficient early methods were based on contracted versions of the array [3] or on noiseless observations.

A solid account on decompositions of 3-way arrays can also be found in DeLathauwer's PhD thesis [16]; 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. A good survey of rank issues can also be found in [31]. An account on identifiability issues can be found in [2].

2. TENSORS

2.1. 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 $i_1..i_d$, with 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 thus has 2 dimensions. A vector is an array of order 1, and a scalar is of order 0.

Throughout this paper, variables take their values in the real field, although all the statements hold true in the complex field with more complicated notations; boldface lowercase letters, like $\boldsymbol{u}$, will denote single-way arrays, *i.e.* vectors, whereas boldface uppercase letters, like $\boldsymbol{G}$, will denote arrays with more than one way, *i.e.* matrices or many-way arrays. The entries of arrays are scalar quantities and are denoted with plain letters, such as u_i or $G_{ijk\ell}$.

A tensor of order d is a d−way array that enjoys the *multilinearity property* after a change of coordinate system. In order to fix the ideas, 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:

$$T'_{ijk} = \sum_{abcd} A_{ia} B_{jb} C_{kc} T_{abc} \quad (1)$$

In particular, moments and cumulants of random variables may be treated as tensors [39]. This product is sometimes referred to as the Tucker product [31] between matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$, weighted by $\boldsymbol{T}$. Note that tensors enjoy property (1) even if the above matrices are not invertible; only linearity is required.

Tensor algebra is a well identified framework; in particular, two kinds of indices are distinguished, covariant or contravariant, depending on the role they play in the application under consideration: an array can indeed be seen as an operator from one space to another. For the sake of simplicity, we *shall not* pay too much attention to this distinction, although it turns out to be important in contexts other than the present one.

2.2. Notation

Given two arrays of order m and n, one defines their *outer product* $\boldsymbol{C} = \boldsymbol{A} \circ \boldsymbol{B}$ as the array of order $m+n$:

$$C_{ij..\ell\, ab..d} = A_{ij..\ell}\, B_{ab..d} \quad (2)$$

For instance, the outer product of two vectors, $\boldsymbol{u} \circ \boldsymbol{v}$, is a matrix.

Given two arrays, $\boldsymbol{A} = \{A_{ij..\ell}\}$ and $\boldsymbol{B} = \{B_{i'j'..\ell'}\}$ of orders d_A and d_B respectively, having the same first dimension, one can define the *mode*−1 contraction product:

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

For instance, the standard matrix-vector product is $\boldsymbol{A}\,\boldsymbol{u} = \boldsymbol{A}^{\mathrm{T}} \bullet \boldsymbol{u}$. Similarly, one defines the *mode*−p inner product when arrays $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same pth dimension, by summing over the pth index; the product is denoted as

$$\boldsymbol{A} \underset{p}{\bullet} \boldsymbol{B}$$

Some authors denote this product as $\boldsymbol{A} \times_p \boldsymbol{B}$, but we find it less readable.

Lastly, we define the symmetric Kronecker product of a $K-$ dimensional vector $\boldsymbol{w}$ by itself, denoted $\boldsymbol{w} \oslash \boldsymbol{w} = \boldsymbol{w}^{\oslash 2}$, as the $K((K+1)/2$−dimensional vector containing all the distinct products, with an appropriate weighting of the cross terms so that $||\boldsymbol{w} \oslash \boldsymbol{w}|| = ||\boldsymbol{w} \otimes \boldsymbol{w}||$. The product $\boldsymbol{w}^{\oslash d}$ is defined in a similar manner for $d > 2$. The $\mathbf{vecs}\{\cdot\}$ operator puts a $K \times K$ symmetric matrix in the form of a $K(K+1)/2$−dimensional vector; conversely, $\mathbf{Unvecs}\{\cdot\}$ puts it back in matrix form. For instance, $\mathbf{Unvecs}\{\boldsymbol{f}^{\oslash 2}\} = \boldsymbol{f}\,\boldsymbol{f}^{\mathrm{T}}$. Space does not permit a detailed description of the notation, and we refer to [26] [45].

3. CUMULANTS

3.1. Definitions

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

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

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 $\Phi^{\boldsymbol{z}}(\boldsymbol{u}) = \mathrm{E}\{exp(\jmath \boldsymbol{u}^{\mathrm{T}} \boldsymbol{z})\}$ about the origin, where

$\jmath = \sqrt{-1}$, cumulants are those of the second characteristic function, $\Psi^{\boldsymbol{z}}(\boldsymbol{u}) = \log(\Phi^{\boldsymbol{z}}(\boldsymbol{u}))$. Moments and cumulants enjoy the multilinearity property (1) and may be considered as tensors [39].

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 [39]. For instance, if all the z_i are mutually independent, then $\mathcal{C}^{\boldsymbol{z}}_{ij..\ell} = \delta(i, j, ..\ell)\, \mathcal{C}^{\boldsymbol{z}}_{ii..i}$; this property is not enjoyed by moments, hence the interest of cumulants.

The reverse is not true. In fact, unless the random variable $\boldsymbol{z}$ is Gaussian, an infinite number of cumulants must vanish in order to ensure their strict sense independence. Therefore, when a cumulant tensor $\mathcal{C}^{\boldsymbol{z}}$ of order d is diagonal, we shall say that the random variables z_i are *independent at order* d.

3.2. Blind Source Separation

Consider the linear statistical model

$$\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{v} \tag{3}$$

where $\boldsymbol{y}$ is an observed random variable of dimension K, $\boldsymbol{x}$ is a random vector of dimension P referred to as the *source vector*, $\boldsymbol{A}$ is a mixing matrix, and $\boldsymbol{v}$ stands for background noise and measurement errors, independent of $\boldsymbol{x}$.

The *Blind Source Separation* (BSS) problem consists of estimating the mixing matrix, $\boldsymbol{A}$, solely from measurements of $\boldsymbol{y}$, and possibly the corresponding estimates of $\boldsymbol{x}$. In the classical BSS framework, the components x_i of $\boldsymbol{x}$ are assumed to be statistically independent (generally not in the strict sense because a weaker independence is sufficient) [11]. In some cases however, sources x_i may be correlated [45] [26] [14], as we shall subsequently see.

To fix the ideas and simplify the notation, assume sources x_i are independent at order 4. Then, from the properties of cumulants we just described, we have:

$$\mathcal{C}^{\boldsymbol{y}}_{ijk\ell} = \sum_{p=1}^{P} A_{ip}A_{jp}A_{kp}A_{\ell p}\mathcal{C}^{\boldsymbol{x}}_{pppp} \tag{4}$$

up to an additive noise term, $\mathcal{C}^{\boldsymbol{v}}$. From measurements of $\boldsymbol{y}$, it is possible to estimate the cumulant tensor $\mathcal{C}^{\boldsymbol{y}}$. Estimating $\boldsymbol{A}$ then amounts to finding the decomposition (4). In practice, because of the noise $\boldsymbol{v}$, this decomposition is not exact. In addition, since the only property utilized is the source independence at a given order, matrix $\boldsymbol{A}$ can only be identified up to a multiplicative factor $\boldsymbol{P}\boldsymbol{\Lambda}$, where $\boldsymbol{P}$ is a permutation and $\boldsymbol{\Lambda}$ is diagonal invertible; see identifiability issues in [11] [2].

Blind Deconvolution is related to the above BSS modeling in two respects. First, a convolution with a finite impulse response can always be written as the product with a Töplitz matrix, which means that the modeling (3) still holds valid, provided matrix $\boldsymbol{A}$ is subject to the Töplitz structure [11] [27]. Second, if the source process is linear, then extracting the sources is equivalent to computing the linear prediction residue [12]. Then, the problem reduces to an unstructured static separation as in (3).

4. ARRAY DECOMPOSITIONS

4.1. Diagonalization by change of coordinates

If the mixing matrix $\boldsymbol{A}$ is square and invertible, then the BSS problem may be seen as a change of coordinates (ICA). In practice, one can always reduce the problem to the latter when the number of sources, P, is at most equal to the observation dimension, K [11].

Denote $\boldsymbol{A} = \boldsymbol{R}\boldsymbol{Q}^{\mathrm{T}}$ the QR factorization of $\boldsymbol{A}$. Because it can be easily estimated, the cumulant tensor of order 2 (the covariance matrix of $\boldsymbol{y}$) is always considered. Unfortunately, this only allows identification of the $\boldsymbol{R}$ factor of $\boldsymbol{A}$, so that the $\boldsymbol{Q}$ factor remains undetermined. It is thus necessary to resort to higher orders, namely 3 or 4, unless other hypotheses can be assumed. The choice between these two possibilities depends on the conditioning of the problem, directly linked to the value of the diagonal tensor $\mathcal{C}^{\boldsymbol{x}}$. At order 3, this tensor vanishes for all symmetrically distributed sources, which strongly limits its use.

Let $\boldsymbol{R}\boldsymbol{R}^{\mathrm{T}}$ denote the Cholesky factorization of the covariance matrix of $\boldsymbol{y}$, and $\tilde{\boldsymbol{y}} = \boldsymbol{R}^{-1}\boldsymbol{y}$ the so-called *standardized* observation vector. In order to find $\boldsymbol{Q}$, one can attempt to diagonalize (approximately) the cumulant tensor of $\boldsymbol{z} = \boldsymbol{Q}\,\tilde{\boldsymbol{y}}$, $\mathcal{C}^{\boldsymbol{z}}_{ijk\ell} = \sum_{pqrs} Q_{ip}Q_{jq}Q_{kr}Q_{\ell s}\mathcal{C}^{\tilde{\boldsymbol{y}}}_{pqrs}$. The random variable $\boldsymbol{z}$ is eventually an estimate of the source vector $\boldsymbol{x}$; in the absence of noise, we have $\boldsymbol{z} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{x}$. Because $\boldsymbol{Q}$ is orthogonal, minimizing the non diagonal entries is equivalent to maximizing the diagonal ones [10], so that $\boldsymbol{Q}$ can be determined by

$$\boldsymbol{Q} = Arg\,\operatorname*{Max}_{\boldsymbol{Q}} \Upsilon_{\alpha,4}\,;\; \Upsilon_{\alpha,4} = \sum_i |\mathcal{C}^{\boldsymbol{z}}_{iiii}|^{\alpha} \tag{5}$$

where $\alpha \geq 1$. Several optimization criteria of this type, called *contrasts*, have been proposed [11] [40] [16] [4] and are justified by Information Theory arguments. Contrary to the matrix case [25], it is generally impossible to exactly null the non diagonal entries of a symmetric tensor of order higher than 2, by just rotating the coordinate axes. More will be said in the next section. Numerical ICA algorithms are surveyed in section 4.5.

4.2. Decomposition into a sum of rank-1 arrays

When the number of sources, P, is strictly larger than the observation dimension K, the previous approach does not apply. In fact, the matrix $\boldsymbol{A}$ now has fewer rows than columns, and the relation $\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x}$ cannot be linearly inverted. In other words, $\boldsymbol{A}$ must be identified without attempting to extract the sources x_p. A symmetric tensor of order d can be expressed via a Canonical Decomposition (CAND) of the form:

$$\mathcal{C}^{\boldsymbol{y}} = \sum_{p=1}^{\omega} \gamma(p)\boldsymbol{a}(p) \circ \boldsymbol{a}(p) \circ \boldsymbol{a}(p) \circ \boldsymbol{a}(p) \qquad (6)$$

The number of terms, ω, reaches a minimum when it equals the *tensor rank*. This CAND decomposition allows identification of matrix $\boldsymbol{A}$ if: (i) it is unique up to $\boldsymbol{\Lambda P}-$indeterminations, and (ii) the tensor rank ω is larger than or equal to the number of sources, P.

Yet, it can be shown that, unfortunately, it is rarely the case [15]. We report below the generic value of the *tensor rank* as a function of the dimension K and the order d. We also report the dimension D of the manifold of solutions; when it is zero, it means that there are a finite number of CAND, and there is a chance of identifying matrix $\boldsymbol{A}$ this way.

ω	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

D	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

For instance, it can be seen that 3rd order tensors have a finite number of CAND for even dimensions. For 4th order tensors, this is satisfied for dimension 7, but not for lower ones. This is unfortunate, for 4th order cumulants are very often better conditioned than 3rd order ones. In addition, the proofs leading to these tables are not constructive. The only known constructive result is given by the:

Sylvester 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 ω distinct linear forms:

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

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

$$\begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_\omega \\ \vdots & & & \vdots \\ \gamma_{d-\omega} & \cdots & \gamma_{d-1} & \gamma_d \end{bmatrix} \boldsymbol{g} = \boldsymbol{0}. \qquad (7)$$

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

Sylvester's theorem not only proves the existence of the ω forms (second column in the tables), but also gives a means to compute them [13] [15]. For odd values of d, we have thus a generic rank of $\omega = \frac{d+1}{2}$, whereas for even values of d, $\omega = \frac{d}{2}+1$. So when d is odd, there is generically a unique vector $\boldsymbol{g}$ satisfying (7), but there are two of them when d is even.

In [17], several extensions to this theorem are proposed in the complex case. In [35], SVD-based algorithms are proposed to compute CAND of 3rd order tensors in larger dimensions.

4.3. Rank-1 approximation

Approximating a tensor by another of rank 1 has at least two applications in the present context. The first one is encountered when when $P \le K$ and when the source extraction is performed one source at a time in model (3), contrary to section 4.1; this is referred to as a *deflation* procedure.

The maximization of the contrast (5) then reduces to that of a single output standardized cumulant (here the *kurtosis*), because a single unit-norm vector is sought, instead of a whole orthogonal matrix:

$$\boldsymbol{w} = Arg \underset{||\boldsymbol{w}||=1}{\mathrm{Max}} \sum_{ijk\ell} w_i\, w_j\, w_k\, w_\ell\, \mathcal{C}^{\boldsymbol{y}}_{ijk\ell} \qquad (8)$$

Yet, it has been shown [18] [13] [31] that this maximization problem is equivalent to minimizing $||\mathcal{C}^{\boldsymbol{y}} - \sigma\, \boldsymbol{w} \circ \boldsymbol{w} \circ \boldsymbol{w} \circ \boldsymbol{w}||$, which is simply finding the best rank-1 approximate of tensor $\mathcal{C}^{\boldsymbol{y}}$.

The second application is found in analytical BSS when sources are discrete [26] or of constant modulus [45]. In this problem, we have to solve a system of N equations of the form $(\boldsymbol{f}^{\mathsf{T}}\boldsymbol{y}_n)^d = 1$, $1 \le n \le N$. This is equivalent to solving a larger linear system $\boldsymbol{Y}\, \boldsymbol{f}^{\oslash d} = \boldsymbol{1}$, under the constraint of $\mathbf{Unvecs}\{\boldsymbol{f}^{\oslash d}\}$ being a rank-1 tensor. Denote $\{\boldsymbol{u}_q^{\oslash d}, 1 \le q \le \bar{P}\}$ a basis of $\mathrm{Ker}\{\boldsymbol{Y}\}$. The solution to this system takes the form $\boldsymbol{f}^{\oslash d} = \boldsymbol{f}^{\oslash d}_{min} + \sum_{p=1}^{\bar{P}} \lambda_p\, \boldsymbol{u}_p^{\oslash d}$ where $\boldsymbol{f}^{\oslash d}_{min}$ is the minimum norm solution. Unfolding these vectors in tensor form leads to the relation

$$\boldsymbol{F} = \boldsymbol{F}_{min} + \sum_{p=1}^{\bar{P}} \lambda_p\, \boldsymbol{U}_p \qquad (9)$$

4.4. Rank-1 combination

The rank-1 combination problem consists of finding the numbers λ_p so that, given matrices $\boldsymbol{U}_p$, matrix $\sum_p \lambda_p \boldsymbol{U}_p$ has a rank of 1. Up to now, this problem has spawned solutions that are not entirely satisfactory. As a consequence, so are the solutions to (9).

Incidentally, we can restate the Joint Approximate Diagonalization (JAD) problem addressed in [5] for the BSS into rank-1 combinations.

The Joint Approximate Diagonalization of $\bar{P}$ matrices $\boldsymbol{N}_p$ consists of finding a square matrix $\boldsymbol{T}$ such that $\boldsymbol{N}_p \approx \boldsymbol{T}\boldsymbol{\Lambda}_p\boldsymbol{T}^{\mathrm{T}}$, for all p, where $\boldsymbol{\Lambda}_p$ are diagonal matrices. This relation can be rewritten in vector form as $\sum_i \lambda_{pi}\boldsymbol{t}_i^{\oslash 2} \approx \boldsymbol{n}_p^{\oslash 2}$. If the matrix $[\lambda_{pi}]$ is full rank and has more columns than rows, then there exists a matrix $\boldsymbol{B}$ such that $\boldsymbol{t}_j^{\oslash 2} \approx \sum_p B_{pj}\boldsymbol{n}_p^{\oslash 2}$. Thus, given matrices $\boldsymbol{N}_p$, the problem is to find for every j, scalar coefficients β_p such that $\sum_p \beta_p \boldsymbol{N}_p$ is a rank-1 matrix.

However, the two problems are not equivalent, for matrix $\boldsymbol{B}$ is not necessarily square.

4.5. Numerical algorithms

The ICA diagonalization of section 4.1 (as well as the JAD briefly mentioned in section 4.4) can be solved entirely analytically in dimension $K=2$, in a number of instances. In order to exploit this property, Comon [10] [11] proposed a sweeping of the pairs of indices, in a similar manner as in the Jacobi diagonalization algorithm for Hermitian matrices [25]. This idea has been later applied to JAD by Cardoso [5]. To see this more in detail, consider the Givens rotation

$$\boldsymbol{Q} = \begin{pmatrix} \cos\phi & \sin\phi\,\exp(\jmath\theta) \\ -\sin\phi\,\exp(-\jmath\theta) & \cos\phi \end{pmatrix}$$

where the angle ϕ is imposed to lie in the interval $(-\pi/2, \pi/2]$, because of inherent $\boldsymbol{\Lambda P}-$indeterminacies. Thus this matrix is entirely defined by the vector $\boldsymbol{u} = [\cos 2\phi,\ \sin 2\phi\cos\theta,\ \sin 2\phi\sin\theta]$. Now, as in (5), define the contrast $\Upsilon_{\alpha,d}$ as the sum of the $d-$th order tensor diagonal entries raised to the power α. Then it can be shown that $\Upsilon_{1,d}$ is a real quadratic form $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}\,\boldsymbol{u}$, and can thus be easily maximized with respect to $\boldsymbol{u}$, and hence to (θ,ϕ) (by convention, if $\alpha=1$, the absolute value is dropped in (5)). On the other hand, this holds true for $\Upsilon_{2,3}$ but not for $\Upsilon_{2,4}$, which can be shown to be a quartic [11]. In addition, this maximization procedure over the whole group of unitary matrices has unfortunately never been proved to always converge to one of the $\boldsymbol{\Lambda P}-$equivalent absolute maxima.

The computation of CAND (section 4.2) has been carried out only in the restrictive case of dimension $K=2$. In the case $(d,K,P)=(4,2,3)$, it is proposed in [13] to simultaneously diagonalize a second cumulant tensor of order 4 in order to fix the indeterminacy exhibited by the Sylvester theorem (the manifold of solutions is of dimension 1). In [14] an algorithm dedicated to discrete sources is proposed, and performs both the identification of $\boldsymbol{A}$ and the extraction of sources x_i, in the case $(d,K,P)=(2,2,3)$.

The rank-1 approximation problem (section 4.3) has been partly solved by algorithms inspired from the matrix *power method* and devised for arrays of higher orders [18] [16] [31].

The rank-1 combination problem (section 4.4) has been solved in a sub-optimal way up to now in [45] [26] by solving a large unconstrained linear system, and trying to restore the structure afterwards.

5. REFERENCES

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

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

[3] J. F. CARDOSO, "Source separation using higher order moments", in *Proc. ICASSP* Glasgow, 1989, pp. 2109–2112.

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

[5] J. F. CARDOSO, A. SOULOUMIAC, "Blind beamforming for non-Gaussian signals", *IEE Proceedings - Part F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[6] J. D. CAROLL, 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.

[7] J. D. CAROLL, S. PRUZANSKY, J. B. KRUSKAL, "CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters", *Psychom.*, vol. 45, no. 1, pp. 3–24, 1980.

[8] L. CASTEDO, O. MACCHI, "Maximizing the information transfer for adaptive unsupervised source separation", in *Proc. SPAWC 97 Conf.*, Paris, France, Apr. 1997, pp. 65–68.

[9] E. CHAUMETTE, P. COMON, D. MULLER, "An ICA-based technique for radiating sources estimation; application to airport surveillance", *IEE Proceedings - Part F*, vol. 140, no. 6, pp. 395–401, Dec. 1993.

[10] P. COMON, "Independent Component Analysis", in *Proc. Int. Sig. Proc. Work. HOS*, Chamrousse, France, Jul. 1991, pp. 111–120, Republished in *Higher-Order Statistics*, J.L.Lacoume ed., Elsevier, 1992, pp 29–38.

[11] P. COMON, "Independent Component Analysis, a new concept ?", *Signal Processing, Elsevier*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[12] P. COMON, "Tensor diagonalization, a useful tool in signal processing", in *IFAC-SYSID, 10th IFAC Symp. Syst. Ident.*, M. Blanke, T. Soderstrom, Eds., Copenhagen, Denmark, July, 1994, vol. 1, pp. 77–82.

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

[14] P. COMON, O. GRELLIER, "Non linear inversion of underdetermined mixtures", in *Proc. IEEE Work. ICA99*, Aussois, France, Jan 11–15 1999, pp. 461–465.

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

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

[17] L. DeLATHAUWER, P. COMON, 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.

[18] L. DeLATHAUWER, P. COMON et al., "Higher-order power method, application in Independent Component Analysis", in *NOLTA Conference*, Las Vegas, 10–14 Dec 1995, vol. 1, pp. 91–96.

[19] L. DeLATHAUWER, 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.

[20] L. DeLATHAUWER, B. DeMOOR, J. VANDEWALLE, "A singular value decomposition for higher-order tensors", in *Second ATHOS workshop*, Sophia-Antipolis, France, Sept 20-21, 1993.

[21] L. DeLATHAUWER, B. DeMOOR, J. VANDEWALLE, "Fetal electrocardiogram extraction by blind source subspace separation", *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 567–572, May 2000.

[22] A. FERREOL, P. CHEVALIER, "On the behavior of current second and higher order blind source separation methods for cyclostationary sources", *IEEE Trans. Sig. Proc.*, vol. 48, pp. 1712–1725, June 2000.

[23] E. GASSIAT, 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.

[24] P. GELADI, "Analysis of multi-way data", *Chemom. Intel. Lab. Syst.*, vol. 7, pp. 11–, 1889.

[25] G. H. GOLUB, C. F. VAN LOAN, *Matrix computations*, The John Hopkins University Press, 1989.

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

[27] V. S. GRIGORASCU, P. A. REGALIA, "Tensor displacement structures and polyspectral matching", in *Fast Reliable Algorithms for Matrices with Structure*, T. Kailath, A. H. Sayed, Eds., chap. 9. SIAM, 1998.

[28] A. HYVÄRINEN, P. HOYER, E. OJA, "Denoising of non Gaussian data by Independent Component Analysis and sparse coding", in *Proc. IEEE Work. ICA99*, Aussois, France, Jan 11–15 1999, pp. 485–489.

[29] M. C. JONES, R. SIBSON, "What is projection pursuit", *Jour. Royal Statist. Society*, vol. 150, no. 1, pp. 1–36, 1987.

[30] C. JUTTEN, J. HERAULT, "Blind separation of sources, Part I", *Signal Processing, Elsevier*, vol. 24, no. 1, pp. 1–20, 1991.

[31] E. KOFIDIS, P. A. REGALIA, "Tensor approximation and signal processing applications", in *AMS Conf. Struct. Matrices Oper. Theory, Num. Ana., Control, Sig. and Im. Proc.* 2000, AMS Publ.

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

[33] 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.

[34] A. LESHEM, A. J. van der VEEN, "Direction of arrival estimation for constant modulus signals", *IEEE Trans. Sig. Proc.*, vol. 47, no. 11, pp. 3125–3129, 1999.

[35] S. LEURGANS, R. T. ROSS, R. B. ABEL, "A decomposition for three-way arrays", *SIAM J. Matrix Anal. Appl.*, vol. 14, no. 4, pp. 1064–1083, Oct. 1993.

[36] M. LEWICKI, T. J. SEJNOWSKI, "Learning nonlinear overcomplete representations for efficient coding", in *Advances in Neural Information Processing Systems*, 1998, pp. 815–821.

[37] A. P. LIAVAS, P. A. REGALIA, J. P. DELMAS, "Robustness of least-squares and subspace methods with respect to effective channel undermodeling/overmodeling", *Trans. on Sig. Proc*, vol. 47, no. 6, pp. 1636–1645, June 1999.

[38] O. MACCHI, E. MOREAU, "Self-adaptive source separation. Part I: Convergence analysis", *IEEE Trans. Sig. Proc.*, vol. 45, no. 4, pp. 918–926, Apr. 1997.

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

[40] E. MOREAU, J. C. PESQUET, "Generalized contrasts for multichannel blind deconvolution of linear systems", *IEEE Signal Processing Letters*, vol. 4, no. 6, pp. 182–183, June 1997.

[41] H. L. NGUYEN-THI, C. JUTTEN, et al., "Separation of sources: A method for speech enhancement", *Applied Sig. Process.*, vol. 3, pp. 177–190, 1996.

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

[43] N. D. SIDIROPOULOS, R. BRO, G. B. GIANNAKIS, "Parallel factor analysis in sensor array processing", *IEEE Trans. Sig. Proc.*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.

[44] 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.

[45] A. J. van der VEEN, A. PAULRAJ, "An analytical constant modulus algorithm", *IEEE Trans. Sig. Proc.*, vol. 44, no. 5, pp. 1136–1155, May 1996.