Mathematics in Signal Processing V

J. G. McWhirter and I. K. Proudler Eds.

Oxford University Press

Contents

Mathematics in Signal Processing V, J. G. McWhirter and I. K. Proudler Eds., Oxford University Press, Oxford, UK, 2001

Tensor Decompositions

State of the Art and Applications

Pierre Comon

Lab. I3S, CNRS, BP121, F-06903 Sophia-Antipolis cedex, France

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. The concept of rank is itself difficult to define, and its calculation raises difficulties. 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 [59] (McCullagh 1987). When second order statistics suffice to establish identifiability, the corresponding algorithms are quite sensitive to model uncertainties [57] (Liavas Regalia and Delmas 1999), so that the complementary use of HOS statistics is often still recommended.

BSS finds applications in Sonar, Radar [13] (Chaumette Comon and Muller 1993), Electrocardiography [32] (DeLathauwer DeMoor at alterae 2000), Speech [61] [52] [26] (Nguyen-Thi and Jutten 1996; Lee and Lewicki 1999; DeLathauwer 1997), and Telecommunications [35] [36] [70] [12] [39] (Ferreol and Chevalier 2000; Gassiat and Gamboa 1997; Van der Veen 1996; Castedo and Macchi 1997; Grellier and Comon 2000), 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 [10] [11] (Caroll and Chang 1970; Caroll Pruzansky and Kruskal

1980), Chemometrics [37] [4] (Geladi et alterae 1989; Bro 1997), and Sensor Array Processing [18] [70] [65] (Comon 1989; Van der Veen and Paulraj 1996; Sidiropoulos Bro and Giannakis 2000; Comon 2000). Other fields where array decompositions can turn out to be useful include Exploratory Analysis [45] (Jones and Sibson 1987), Complexity Analysis [50] [43] (Kruskal 1977; Howell 1978), and Sparse Coding [44] (Hyvärinen Hoyer and Oja 1999).

Bibliographical survey. Bergman [1] (1969) and Harshman [41] (1970) were the first to notice that the concept of *rank* was difficult to extend from matrices to higher order arrays. Caroll [10] (1970) provided the first *canonical decomposition* algorithm of a three-way array, later referred to as *Candecomp* model. Several years later, Kruskal [50] (1977) conducted a detailed analysis of uniqueness, and related several definitions of rank. The algorithm *Candelinc* was devised by Caroll and others in the eighties [11] (Caroll et alterae 1980); it allowed to compute a canonical decomposition subject to a priori linear constraints.

Leurgans and others [54] (1993) 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 [15] (1991) to approximately decompose it into d terms. The problem was then reduced to finding an invertible linear transform (change of coordinates); see [16] [9] (Comon 1994; Cardoso 1993) 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 *over-complete* ICA [52] [18] (Lee et alterae 1999; Comon 1998).

The terminology of ICA is meaningful in the context of Signal Processing and BSS [46] [16] [8] (Jutten and Hérault 1991; Comon 1994; Cardoso 1999). Constructive algorithms for ICA either proceed by sweeping the pairs of indices [14] [16] [9] (Comon 1989; Comon 1991; Cardoso and Souloumiac 1993; Comon 1994), or are of iterative nature, like power methods [28] [49] (DeLathauwer Comon and others 1995; Kofidis and Regalia 2000), gradient descents [58] (Macchi and Moreau 1997), or Robbins-Monro algorithms [46] [61] (Jutten and Hérault 1991; Nguyen-Thi Jutten and others 1996). Some less efficient early methods were based on contracted versions of the array [6] (Cardoso 1989) or on noiseless observations[14] (Comon 1989).

A solid account on decompositions of 3-way arrays can also be found in DeLathauwer's PhD thesis [26] (1997); an interesting tool defined therein is the *HOSVD* [30] (DeLathauwer and others 1993), generalizing the concept of SVD to arrays of order 3, in a different manner compared to Caroll, but quite similar to the *Tuckals* decomposition [55] [68] [37] (Levin 1965; Tucker 1965; Geladi 1989). A good survey of rank issues can also be found in [49] (Kofidis et alterae 2000). An account on identifiability issues can be found in [5] (Cao and Liu 1996).

[25]

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, and unless otherwise specified, 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. For instance, 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:

$$T'_{ijk} = \sum_{abc} A_{ia} B_{jb} C_{kc} T_{abc} \tag{1}$$

In particular, moments and cumulants of random variables may be treated as tensors [59] (McCullagh 1987). This product is sometimes referred to as the Tucker product [49] (Kofidis et al. 2000) 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 $m+n$:

$$C_{ij..\ell\, ab..d} = A_{ij..\ell}\, B_{ab..d} \tag{2}$$

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

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

If unspecified, the contraction applies by default to the first index. Some authors denote this product as $\boldsymbol{A} \times_p \boldsymbol{B}$, but we find it less readable.

We define the Kronecker product $\boldsymbol{u} \otimes \boldsymbol{v}$ between two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ as the vector containing all the possible cross-products [3] (Brewer 1978). If $\boldsymbol{u}$ and $\boldsymbol{v}$ are of dimension J and K, then $\boldsymbol{u} \otimes \boldsymbol{v}$ is of dimension JK.

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 $d > 2$. For instance, $\boldsymbol{w}^{\oslash 3}$ is of size $K(K+1)(K+2)/6$.

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, $\mathbf{Unvecs}\{\boldsymbol{f}^{\oslash 2}\} = \boldsymbol{f}\,\boldsymbol{f}^{\mathrm{T}}$. Refer to [39] [70] (Van der Veen 1996; Grellier and Comon 2000) for a more detailed description.

2.3 Homogeneous polynomials

$d-$way arrays can be written in two different manners, as pointed out by several authors [59] (McCullagh 1987), related to each other by a bijective mapping, $\boldsymbol{f}$. Assume the notations $\boldsymbol{x}^{\boldsymbol{j}} \stackrel{\text{def}}{=} \prod_{k=1}^{K} x_k^{j_k}$ and $|\boldsymbol{j}| \stackrel{\text{def}}{=} \sum_k j_k$. Then for homogeneous monomials of degree d, $\boldsymbol{x}^{\boldsymbol{j}}$, we have $|\boldsymbol{j}| = d$.

To start with, take the example of $(K, d) = (4, 3)$: one can associate every entry T_{ijk} to a monomial $T_{ijk}\, x_i x_j x_k$. For instance, T_{114} is associated with $T_{114}\, x_1^2 x_4$, and thus to $T_{114}\, \boldsymbol{x}^{[2,0,0,1]}$; this means that $f([1,1,4]) = [2,0,0,1]$.

More generally, the $d-$dimensional vector index $\boldsymbol{i} = [i, j, k]$ can be associated with a $K-$dimensional vector index $\boldsymbol{f}(\boldsymbol{i})$ containing the number of times each variable x_k appears in the associated monomial. Whereas the d entries of $\boldsymbol{i}$ take their values in $\{1, \ldots, K\}$, the K entries of $\boldsymbol{f}(\boldsymbol{i})$ take their values in $\{1, \ldots, d\}$ with the constraint that $\sum_k f_k(\boldsymbol{i}) = d, \forall \boldsymbol{i}$.

As a consequence, the linear space of symmetric tensors can be bijectively associated with the linear space of homogeneous polynomials. To see this, it suffices to associate every polynomial $p(\boldsymbol{x})$ with the symmetric tensor $\boldsymbol{G}$ as:

$$p(\boldsymbol{x}) = \sum_{|\boldsymbol{f}(\boldsymbol{i})|=d} G_{\boldsymbol{i}}\, \boldsymbol{x}^{\boldsymbol{f}(\boldsymbol{i})} \tag{3}$$

where $G_{\boldsymbol{i}}$ are the entries of $\boldsymbol{G}$. The dimension of these spaces is $S = \binom{K+d-1}{d}$, and one can choose as a basis the set of monomials: $\mathcal{B}(K;d) = \{\boldsymbol{x}^{\boldsymbol{j}},\ |\boldsymbol{j}| = d\}$.

> *Example.* Let p and q be two homogeneous polynomials in K variables, associated with tensors $\boldsymbol{P}$ and $\boldsymbol{Q}$, possibly of different orders. Then, polynomial pq is associated with $\boldsymbol{P} \circ \boldsymbol{Q}$:
>
> $$p(\boldsymbol{x})q(\boldsymbol{x}) = \sum_{\boldsymbol{i}} \sum_{\boldsymbol{j}} P_{\boldsymbol{i}} Q_{\boldsymbol{j}} \boldsymbol{x}^{\boldsymbol{f}(\boldsymbol{i})+\boldsymbol{f}(\boldsymbol{j})} = \sum_{[\boldsymbol{i}\,\boldsymbol{j}]} [\boldsymbol{P} \circ \boldsymbol{Q}]_{[\boldsymbol{i}\,\boldsymbol{j}]}\, \boldsymbol{x}^{\boldsymbol{f}([\boldsymbol{i}\,\boldsymbol{j}])}$$

In practice, it is convenient to take into account the symmetry of the tensor, by defining $c(\boldsymbol{j})$ as the number of times the entry $T_{\boldsymbol{f}^{-1}(\boldsymbol{j})}$ appears in the array $\boldsymbol{T}$: $c(\boldsymbol{j}) = |\boldsymbol{j}|!/(\boldsymbol{j})!$, where $(\boldsymbol{j})! \stackrel{\text{def}}{=} \prod_k j_k!$. For binary quantics, $K = 2$ and $c(\boldsymbol{j}) = \binom{d}{j_1} = \binom{d}{j_2}$; for instance, for $d = 4$, $c([3,1]) = 4$ and $c([2,2]) = 6$.

Coefficients $\gamma(\boldsymbol{j}, p)$ of a polynomial p in basis $\mathcal{B}$ are chosen so as satisfy the relation:

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

Now, both spaces can be provided with a scalar product. For $d-$way arrays of dimension K, define the Froebenius scalar product:

$$\langle \boldsymbol{G}, \boldsymbol{H} \rangle = \sum_{\boldsymbol{i}} \boldsymbol{G}_{\boldsymbol{i}}\, \boldsymbol{H}_{\boldsymbol{i}}$$

and the induced Euclidian norm. For homogeneous polynomials of degree d in K variables, define the scalar product as

$$\langle p, q \rangle = \sum_{|\boldsymbol{j}|=d} c(\boldsymbol{j})\, \gamma(\boldsymbol{j}, p)\, \gamma(\boldsymbol{j}, q)$$

In particular, monomials in $\mathcal{B}$ satisfy $||\boldsymbol{x}^{\boldsymbol{j}}||^2 = (\boldsymbol{j})!/d! = 1/c(\boldsymbol{j})$. In the case of binary quantics ($K = 2$), this was called the *apolar* scalar product [51] (Kung and Rota 1984). The latter definition has several advantages [24] (Comon and Mourrain 1996). For instance, if $a(\boldsymbol{x})$ is a homogeneous linear form with coefficient vector $\boldsymbol{a}$, then the scalar product $\langle a^d, q \rangle$ with any homogeneous polynomial q of degree d turns out to be the value of q at $\boldsymbol{a}$:

$$\langle a(\boldsymbol{x})^d, q(\boldsymbol{x}) \rangle = \sum_{\boldsymbol{i}} c(\boldsymbol{i})\, \gamma(\boldsymbol{i}, q)\, \boldsymbol{a}^{\boldsymbol{i}} = q(\boldsymbol{a})$$

The interest in establishing a link between tensors and polynomials lies in the fact that polynomials have been studied rather deeply during the last century [63] (Salmon 1885). Some of the results obtained will be useful in this chapter, and in particular the classification of cubics.

2.4 Genericity

A property will be referred to as *generic* if it is true on a dense algebraic subset. The topology used is the standard one for homogeneous polynomials, namely that of Zariski [64] (Shafarevich 1977). Recall that in this topology, the closed subsets are defined by algebraic equations of the form $p(\boldsymbol{x}) = 0$, where p is a polynomial. Its particularity is that two open non empty subsets always intersect; in other words, the topology is not separated.

For instance, a symmetric matrix is generically of full rank. In fact, the set of singular matrices is defined by the polynomial relation $\det(\boldsymbol{A}) = 0$, which is associated with a closed subset, whose complementary is dense. We shall subsequently see that this does not hold true anymore for tensors of order higher than 2. For instance, the $2 \times 2 \times 2$ tensor $\boldsymbol{T}$ such that $T_{122} = T_{212} = T_{221} = 1$, and zero elsewhere (cf. figure 1), is known to be of rank 3. However, the generic rank is 2 in that case, as will be discussed in section 4.2.

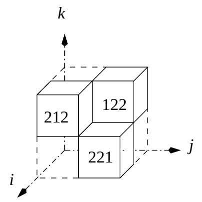

Figure 1: Non generic example of a binary cubic of maximal rank: position of non-zero entries in the $2 \times 2 \times 2$ associated symmetric tensor.

2.5 Array ranks

Let $\boldsymbol{T}$ be a tensor, not necessarily symmetric, of dimensions $n_1 \times \ldots \times n_d$. One defines the *tensor rank* ω of $\boldsymbol{T}$ as the minimal number of rank one tensors whose linear combination yields $\boldsymbol{T}$. The properties of tensor rank will be extensively discussed in section 4.2. For completeness, let us also mention the definition of *mode−n ranks.*

The mode$-n$ vectors of $\boldsymbol{T}$ are obtained by varying index i_n and keeping the others fixed; there are thus as many mode$-n$ vectors as possibilities of fixing indices i_k, $k \neq n$. The mode$-n$ rank, R_n, is defined as the dimension of the linear space spanned by all mode$-n$ vectors of $\boldsymbol{T}$.

Bounds. Howell (1978) [43] showed that the tensor rank can be bounded as $\omega \leq \max_{i \neq j}\{n_i n_j\}$. On the other hand, mode$-n$ ranks and tensor rank are related by the inequality $R_n \leq \omega$, $\forall k$.

In the symmetric case, Reznick showed that the tensor rank can be bounded as a function of the dimension K and the order, d:

$$\omega \leq \begin{pmatrix} K+d-2 \\ d-1 \end{pmatrix} \tag{4}$$

but this bound is rather loose, except in some very particular cases, as will be commented in section 4.2.

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:

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

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 the dotless $\jmath$ denotes $\sqrt{-1}$, cumulants are those of the second characteristic function, $\Psi^{\boldsymbol{z}}(\boldsymbol{u}) = \log(\Phi^{\boldsymbol{z}}(\boldsymbol{u}))$; for complex random variables, it suffices to consider the joint distribution of their real and imaginary parts. Moments and cumulants enjoy the multilinearity property (1) and may be considered as tensors [59] (McCullagh 1987).

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 $\mathcal{C}^{\boldsymbol{z}}_{ij..\ell} = \delta(i, j, ..\ell)\, \mathcal{C}^{\boldsymbol{z}}_{ii..i}$ [48] (Kendall and Stuart 1977), where the Kronecker δ is null unless *all* its arguments are equal. This property is not enjoyed by moments, hence the interest in 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*.

Gaussian variables play a particular role, since they are the only random variables that have a finite set of non-zero cumulants [47] [34] (Kagan et al. 1973; Feller 1968). The cumulants of order 1 and 2 are better known under the names of statistical *mean* and *covariance*. To illustrate this property, I looked for a long time for a simple non Gaussian random variable having null cumulants of order 3 and 4. Take a random variable with values in the complex plane. If its distribution is 0 with probability one half, and uniformly distributed on the unit circle with probability one half, then its mean $\mathrm{E}\{z\}$ is zero, its variance $\mathrm{E}\{zz^*\} = 1$, its second-order non-circular cumulant $\mathrm{E}\{z^2\} = 0$, its 2 marginal cumulants of order 3, $\mathrm{E}\{z^3\}$ and $\mathrm{E}\{z^2 z^*\}$, are zero, as well as its 3 marginal cumulants of order 4, $\mathrm{Cum}\{z,z,z,z\}$, $\mathrm{Cum}\{z,z,z,z^*\}$, $\mathrm{Cum}\{z,z,z^*,z^*\}$. Yet, this variable is obviously not Gaussian. This is a very striking example, that one encounters in secondary radar applications.

3.2 Blind Source Separation

Consider the linear statistical model

$$\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{v} \tag{5}$$

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, possible interferers, and measurement errors, independent of $\boldsymbol{x}$.

The *Blind Source Separation* (BSS) problem consists of estimating the mixing matrix, $\boldsymbol{A}$, and possibly the corresponding estimates of $\boldsymbol{x}$, solely from measurements of $\boldsymbol{y}$. 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) [16] (Comon 1994). In some cases however, sources x_i may be correlated [70] [39] [21] (Van der Veen 1996; Grellier and Comon 2000; Comon and Grellier 1999), as we shall subsequently see, and this is not necessarily an obstacle to their separation.

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:

$$\mathcal{C}^{\boldsymbol{y}}_{ijk\ell} = \sum_{p=1}^{P} A_{ip} A_{jp} A_{kp} A_{\ell p} \mathcal{C}^{\boldsymbol{x}}_{pppp} \tag{6}$$

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 (6). 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 $\boldsymbol{P\Lambda}$, where $\boldsymbol{P}$ is a permutation and $\boldsymbol{\Lambda}$ is diagonal invertible; see identifiability issues in [16] [5] (Cao and Liu 1996; Comon 1994).

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 (5) still holds valid, provided matrix $\boldsymbol{A}$ is subject to the Töplitz structure [16] [40] (Comon 1994; Grigorascu and Regalia 1998). Second, if the source process is linear, then extracting the sources is equivalent to computing the linear prediction residue [17] (Comon 1994). Then, the problem reduces to an unstructured static separation as in (5).

4 Array decompositions

4.1 Diagonalization by change of coordinates

Preprocessing for square mixtures. If the mixing matrix $\boldsymbol{A}$ is square and invertible, which means that the number of sources, P, is equal to the observation dimension, K, then the BSS problem may be seen as a bijective congruent transformation (ICA).

Denote $\boldsymbol{R}_y$ the covariance matrix of the observation. The goal is to find an estimate $\boldsymbol{z}$ of $\boldsymbol{x}$ such that its components z_i are statistically independent. The first idea is thus to build a vector $\tilde{\boldsymbol{y}} = \boldsymbol{T}\,\boldsymbol{y}$ that has a diagonal covariance, yielding decorrelated components. This can be easily done by searching for a (non unique) square root factor of $\boldsymbol{R}_y$; it can be obtained by a Cholesky factorization or by an Eigen Value decomposition (EVD) of $\boldsymbol{R}_y$. We then define $\boldsymbol{T}$ as the inverse of this factor, so that $\boldsymbol{T}\boldsymbol{R}_y\boldsymbol{T}^{\mathrm{T}} = \boldsymbol{I}$.

With this preprocessing $\boldsymbol{T}$, the obtained random variable $\tilde{\boldsymbol{y}}$ has a covariance equal to identity. We say that this variable is *standardized.*

Now, it may be more appropriate, when the noise covariance, $\boldsymbol{R}_v$, (or conversely the signal covariance $\boldsymbol{R}_s = \boldsymbol{R}_y - \boldsymbol{R}_v$) is known, to build $\boldsymbol{T}$ as the inverse of a square root of the signal covariance: $\boldsymbol{T}\boldsymbol{R}_s\boldsymbol{T}^{\mathrm{T}} = \boldsymbol{I}$. In fact, this yields an unbiased solution in the presence of noise. Unfortunately, neither $\boldsymbol{R}_v$ nor $\boldsymbol{R}_s$ are known in general, hence the former procedure based on $\boldsymbol{R}_y$.

Preprocessing for rectangular mixtures. In practice, one can always reduce the problem to the latter when the number of sources, P, is smaller than the observation dimension, K, in the absence of noise, or when the noise covariance is known. This is now explained below.

If noise is present, denote $\boldsymbol{T}_v$ the inverse of a square root of $\boldsymbol{R}_v$, such that we have $\boldsymbol{T}_v\boldsymbol{R}_v\boldsymbol{T}_v{}^{\mathrm{T}} = \sigma_v\boldsymbol{I}$; if noise is absent, set $\boldsymbol{T}_v = \boldsymbol{I}$ and $\sigma_v = 0$. Now consider the matrix $\boldsymbol{T}_v\boldsymbol{R}_y\boldsymbol{T}_v = \boldsymbol{T}_v\boldsymbol{R}_s\boldsymbol{T}_v + \sigma\boldsymbol{I}$. Its EVD allows to detect the number of non-zero eigenvalues in $\boldsymbol{R}_s$ [2] (Bienvenu and Kopp 1983), equal to P by definition, as well as to estimate the source space spanned by the associated eigenvectors: $\boldsymbol{T}_v\boldsymbol{R}_y\boldsymbol{T}_v = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{T}} + \sigma\boldsymbol{I}$. The matrix $\boldsymbol{U}$ is here of dimension $K \times P$ and of full rank. The preprocessing defined as $\boldsymbol{T} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{T}_v$ eventually yields a $P-$dimensional standardized vector $\tilde{\boldsymbol{y}} = \boldsymbol{T}\boldsymbol{y}$ whose noiseless part has a unit covariance, as in the previous paragraph.

Lastly, if the mixture is rectangular, but with more sources than sensors, *i.e.*, $P > K$, the mixture cannot be linearly inverted. Such mixtures are referred to as *underdetermined* or *over-complete*, as already pointed out in the bibliographical survey, and their identification will be addressed separately in section 4.2. In such a case, the preprocessing is unuseful, and not recommended.

Orthogonal change of coordinates. In the preprocessing, we have done only part of the job. In fact, we have constructed a matrix $\boldsymbol{T}$ such that ideally

$\boldsymbol{TAA}^{\mathrm{T}}\boldsymbol{T}^{\mathrm{T}} = \boldsymbol{I}$, but this only implies that $\boldsymbol{TA} = \boldsymbol{Q}^{\mathrm{T}}$, for some $P \times P$ orthogonal matrix $\boldsymbol{Q}$. This $\boldsymbol{Q}$ factor still remains undetermined. It is thus necessary to resort statistics of order higher than 2, 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. At order 4, this tensor is generally non zero, except in some exceptional pathological cases, as that mentioned in section 3.1.

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 $\boldsymbol{z} = \boldsymbol{P\Lambda x}$. Because $\boldsymbol{Q}$ is orthogonal, minimizing the non diagonal entries is equivalent to maximizing the diagonal ones [15] (Comon 1991), so that $\boldsymbol{Q}$ can be determined by

$$\boldsymbol{Q} = Arg\,\underset{\boldsymbol{Q}}{\mathrm{Max}}\,\Upsilon_{\alpha,4}\,;\ \Upsilon_{\alpha,4} = \sum_i |\mathcal{C}^{\boldsymbol{z}}_{iiii}|^{\alpha} \tag{7}$$

where $\alpha \geq 1$. Several optimization criteria of this type, called *contrasts*, have been proposed [16] [60] [26] [8] [20] (Comon 1994; Moreau and Pesquet 1997; DeLathauwer 1997; Cardoso 1999; Comon 2001) and are justified by Information Theory arguments. Contrary to the matrix case [38] (Golub and Van Loan 1989), 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. In other words, the class of decompositions presented in this section lead to rank$-K$ *approximations* of $K-$dimensional symmetric tensors. More will be said in the next section. Numerical ICA algorithms are surveyed in section 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 noiseless 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:

$$\mathcal{C}^{\boldsymbol{y}} = \sum_{p=1}^{\omega} \gamma(p)\,\boldsymbol{a}(p) \circ \boldsymbol{a}(p) \circ \boldsymbol{a}(p) \circ \boldsymbol{a}(p) \tag{8}$$

The number of terms, ω, reaches a minimum when it equals the *tensor rank*. This CAND decomposition allows the 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.

Generic rank. We report in the tables below the generic value of the *tensor rank* as a function of the dimension K and the order d [24] (Comon and Mourrain 1996). 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 (at most d^S), and there is a chance of identifying matrix $\boldsymbol{A}$ this way.

> *Example.* Fore matrices ($d = 2$), it is known that a quadratic form cannot be uniquely decomposed into a sum of squares. The manifold of solutions is of dimension $D = K(K-1)/2$.

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

Table 1: Generic rank ω of symmetric tensors as a function of the dimension K and the order d

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

Table 2: Generic dimension D of the manifold of solutions

The first striking fact that appears in table 1 is that *the rank can exceed the dimension*, which is not true for matrices. For instance, it can be seen that $P = 5$ sources can be identified in dimension $K = 4$ with a 3rd order cumulant tensor, whereas this number increases to $P = 10$ with a 4th order tensor.

One can also deduce from table 2 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. Furthermore, most of the proofs leading to these tables are not constructive. The only known constructive result is given by the Sylvester theorem (section 5.3).

$\mathcal{GI}-$orbit	$\omega(p)$
x^3	1
$x^2y + xy^2$	2 (**generic**)
x^2y	3

Table 3: Equivalence classes of binary cubics: orbits under the action of $\mathcal{GI}$, the group of invertible 2−dimensional changes of coordinates.

Non generic rank. In addition, these results are only valid in generic cases. And it turns out that, contrary to matrices (*i.e.*, 2nd order tensors), *the generic rank is not always maximal*. In other words, the rank can exceed its generic value. Unfortunately, the maximal achievable rank is not known for all pairs (P, K). But we can still illustrate this odd fact with particular values.

> *Example.* For instance, for $K = 2$ and $d = 3$, the maximal rank is 3. The symmetric tensors having rank 3 are associated with polynomials in the orbit of x^2y. The tensor associated with the latter homogeneous polynomial is represented in figure 1, where only 3 entries are equal to 1, the others being null. As reported in table 3 there is a single class associated with every value of the tensor rank.
>
> Now to make it more explicit, the polynomial x^2y can be written as:
>
> $$6\,x^2y = (x+y)^3 + (-x+y)^3 - 2y^3$$
>
> This relation can be rewritten in tensor form as:
>
> $$\boldsymbol{T} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}^{\circ 3} + \begin{pmatrix} -1 \\ 1 \end{pmatrix}^{\circ 3} - 2\begin{pmatrix} 0 \\ 1 \end{pmatrix}^{\circ 3}$$
>
> which is an explicit irreducible CAND. This decomposition is depicted in figure 2. Also note that in this case, the Reznick bound (4) is reached: $\omega = \binom{3}{2} = 3$.

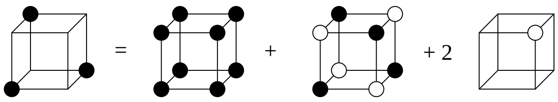

Figure 2: Explicit decomposition of the non generic example of binary cubic of maximal rank. Black bullets represent $+1$'s and white bullets -1's.

> *Example.* Now take $K = 3$ and $d = 3$. We are thus handling $3 \times 3 \times 3$ symmetric tensors, or equivalently, ternary cubics. The generic rank is 4, but the maximal rank is 5, according to table 4. The class of maximal rank is unique, and a representative is depicted in figure 3; the 6 non-zero entries are all equal. Note that other non generic classes occur with also a rank of 4, as pointed out in table 4.

> *Example.* Finally, consider ternary quartics, *i.e.*, $(K, d) = (3, 4)$. In this case, the number of free parameters in the tensor is $S = \binom{6}{4} = 15$. The number of free parameters in CAND exceeds 15 as soon as $\omega \geq 5$. So we could hope that we are lucky, because the number of free parameters is the same on both sides of CAND. Unfortunately, this is not the case, and Clebsh showed that the maximal achievable rank was 6 [33] (Ehrenborg and Rota 1993). Of course, this maximal rank is not generic, as reported in table 1.

$\mathcal{GI}-$orbit	$\omega(p)$
x^3	1
$x^2y + xy^2$	2
x^2y	3
$x^3 + 3\,y^2z$	4
$x^3 + y^3 + 6\,xyz$	4
$x^3 + 6\,xyz$	4
$a\,(x^3 + y^3 + z^3) + 6b\,xyz$	4 **(generic)**
$x^2y + xz^2$	5

Table 4: Equivalence classes for ternary cubics: orbits under the action of $\mathcal{GI}$, the group of invertible 3−dimensional changes of coordinates.

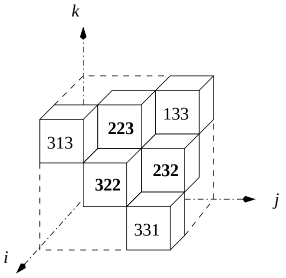

Figure 3: Non generic example of ternary cubic, proved to be of maximal rank: position of non-zero entries

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 $P \leq K$ and when the source extraction is performed one source at a time in model (5), contrary to section 4.1; this is referred to as a *deflation* procedure.

The maximization of the contrast (7) 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:

$$\boldsymbol{w} = Arg \operatorname*{Max}_{||\boldsymbol{w}||=1} \sum_{ijk\ell} w_i w_j w_k w_\ell \, \mathcal{C}^{\boldsymbol{y}}_{ijk\ell} \tag{9}$$

Yet, it has been shown [28] [18] [49] (DeLathauwer Comon and others 1995; Comon 1998; Kofidis and regalia 2000) 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 [39] (Grellier and Comon 2000) or of constant modulus [70] (Van der Veen 1996).

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 $\boldsymbol{Y}\,\boldsymbol{f}^{\oslash d} = \mathbf{1}$, 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

$$\boldsymbol{f}^{\oslash d} = \boldsymbol{f}_{min}^{\oslash d} + \sum_{p=1}^{\bar{P}} \lambda_p\, \boldsymbol{u}_p^{\oslash d}$$

where $\boldsymbol{f}_{min}^{\oslash d}$ is the minimum norm solution. Unfolding these vectors in tensor form leads to the relation

$$\boldsymbol{F} = \boldsymbol{F}_{min} + \sum_{p=1}^{\bar{P}} \lambda_p\, \boldsymbol{U}_p \tag{10}$$

This problem can be shown to be related to the rank-1 combination problem that we describe below.

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 (10).

Incidentally, we can restate the Joint Approximate Diagonalization (JAD) problem addressed in [9] (Cardoso and Souloumiac 1993) 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 $\boldsymbol{N}_p \approx \boldsymbol{T}\boldsymbol{\Lambda}_p\boldsymbol{T}^{\mathsf{T}}$, for all p, where $\boldsymbol{\Lambda}_p$ are diagonal matrices. From a property recalled in section 2.2, this relation can be rewritten in vector form as $\mathbf{vecs}\{\boldsymbol{N}_p\} \stackrel{\text{def}}{=} \boldsymbol{n}_p \approx \sum_i \lambda_{pi}\boldsymbol{t}_i^{\oslash 2}$, $\boldsymbol{t}_i$ denoting the ith column of $\boldsymbol{T}$. If the matrix $[\lambda_{pi}]$ is full rank and has more columns than rows, then there exists a matrix $\boldsymbol{B}$ such that $\boldsymbol{t}_j^{\oslash 2} \approx \sum_p B_{pj}\boldsymbol{n}_p$. 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, and hence the link with the rank-1 combination problem.

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

5 Numerical algorithms

5.1 Contrast maximization

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 [15] [16] (1991) proposed a sweeping of the pairs of indices, in a similar manner as in the Jacobi diagonalization algorithm for Hermitian matrices [38] (Golub and Van Loan 1989). This idea has been later applied to JAD by Cardoso [9] (Cardoso and Souloumiac 1993). 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 (7), 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 $\Upsilon_{1,3}$ and $\Upsilon_{1,4}$ are real quadratic forms in $\boldsymbol{u}$, 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 (7)). 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 [16] [20] (Comon 1994; Comon 2001). Nevertheless, polynomials of degree 4 can still be rooted analytically.

The procedure originally proposed by Comon (1989) consisting of sweeping all the pairs, like in some numerical algorithms dedicated to matrices, has never been proved to always lead to one of the $\boldsymbol{\Lambda P}-$equivalent absolute maxima, even if this is always observed in practice. Counter-examples have never been found either. So we consider this convergence issue as an open problem, belonging to the general class of optimization problems over multiplicative groups. However, some elements of convergence are now reported below.

Convergence. For compactness, denote $\boldsymbol{G}$ the cumulant tensor of the standardized observation, $\bar{\boldsymbol{y}}$, which has been denoted $\mathcal{C}^{\bar{\boldsymbol{y}}}$ up to now. Also denote $\boldsymbol{Z} = \mathcal{C}^{\boldsymbol{z}}$ the cumulant tensor obtained after an orthogonal transformation $\boldsymbol{Q}$. According to the multi-linearity property, we have that:

$$Z_{pq..r} = \sum_{ij..\ell} Q_{pi} Q_{qj} \dots Q_{r\ell}\, G_{ij..\ell} \tag{11}$$

Consider first the matrix case (order 2) in order to fix the ideas. The contrast (7) can then be written as:

$$\Upsilon_{2,2} = \sum_p |Z_{pp}|^2 \tag{12}$$

Because $\boldsymbol{Q}$ is orthogonal, its differential can be written as

$$d\boldsymbol{Q} = d\boldsymbol{S}\,\boldsymbol{Q} \tag{13}$$

where matrix $\boldsymbol{S}$ is skew-symmetric. This yields the relation characterizing stationary points, $\boldsymbol{Z}$: $\frac{1}{2}\, d\Upsilon_{2,2} = 2\sum_{p,t} Z_{pp} S_{pt} Z_{tp} = 0$. Yet, this is true for any skew-symmetric matrix $\boldsymbol{S}$, and hence for every skew-symmetric matrix having only two non zero entries (one $+1$ and one -1); based on this argument, one concludes that:

$$(Z_{qq} - Z_{rr})Z_{qr} = 0, \text{ for } q \neq r \tag{14}$$

Next, the local convexity can be examined with the help of the same tools, observing that:

$$\frac{1}{4} d^2\Upsilon_{2,2} = 4Z_{qr}^2 - (Z_{qq} - Z_{rr})^2 \tag{15}$$

Thus, there are three kinds of stationary points: (i) those for which all diagonal entries are equal, which correspond to minima of $\Upsilon_{2,2}$, (ii) those for which all non-diagonal entries are null, which correspond to maxima, and (iii) saddle points, for which some diagonal entries are equal and some non-diagonal entries vanish. This result is well known, and proves that the only maxima are diagonal matrices, which can be deduced from each other by mere permutation within the diagonal.

Now let us develop the same calculations for tensors of order 3 and 4. Stationary values are given by the relations:

$$\begin{aligned} \frac{1}{2}\, d\Upsilon_{2,3} &= 3\sum_{p,t} Z_{ppp}\, dS_{pt}\, Z_{tpp} = 0, \\ \frac{1}{2}\, d\Upsilon_{2,4} &= 4\sum_{p,t} Z_{pppp}\, dS_{pt}\, Z_{tppp} = 0 \end{aligned}$$

or, on the basis of skew-symmetric matrices, for $q \neq r$:

$$Z_{qqq}Z_{qqr} - Z_{rrr}Z_{qrr} = 0, \tag{16}$$

$$Z_{qqqq}Z_{qqqr} - Z_{rrrr}Z_{qrrr} = 0, \tag{17}$$

whereas local convexity conditions are governed by (Comon 1994):

$$\frac{1}{6} d^2\Upsilon_3 = 4Z_{qqr}^2 + 4Z_{qrr}^2 - (Z_{qqq} - Z_{qrr})^2 - (Z_{rrr} - Z_{qqr})^2 \tag{18}$$

$$\frac{1}{8} d^2\Upsilon_4 = \frac{9}{2} Z_{qqrr}^2 + 4Z_{qqqr}^2 + 4Z_{qrrr}^2 - (Z_{qqqq} - \frac{3}{2} Z_{qqrr})^2 - (Z_{rrrr} - \frac{3}{2} Z_{qqrr})^2 \tag{19}$$

The comparison of these results with (14) and (15) lead to two conclusions: (a) non-diagonal terms do not factorize anymore in (16) and (17), so that stationary values are more difficult to characterize, and (b) diagonal tensors are still local maxima, but there are *a priori* others. This is another problem, linked to optimization in groups, that this author considers as open.

Sweeping strategies. We have presented several numerical algorithms aiming at separating $P = 2$ sources from $K = 2$ sensors in the presence of noise of unknown statistics. Inspired from the Jacobi cyclic-by rows sweeping strategy proposed for matrices, we can process all the $K(K-1)/2$ pairs one by one sequentially (Comon 1989; Comon 1994). However, as in the matrix case, the noise part (constituted by the actual background noise and all the other $K-2$ sources) changes at every step, so that a single sweeping is not sufficient. In practice, an order of $\sqrt{K}$ sweeps have been shown to be sufficient.

Other strategies have been analyzed, and consist of processing first the pair of sensors that yields the maximal increase in the contrast criterion. This strategy has also been implemented successfully, but is not always numerically efficient.

When processing one pair (i, j), one can either recompute all the entries of the cumulant tensor that have been affected (*i.e.*, those whose indices contain i or j), or compute the rotated data instead. The two possibilities do not have the same numerical complexity, and the best choice depends on the number of sensors, K, and on the number of samples, N.

5.2 Parafac algorithm

In [54] (Leurgans et alterae 1993), SVD-based algorithms are proposed to compute CAND of 3rd order tensors in larger dimensions. However, these algorithms, called PARAFAC, need the number of sources, P, to be smaller than or equal to $\frac{3}{2}K-1$, in the symmetric case we are interested in. See also [50] [4] (Kruskal 1977; Bro 1997) for more details. In view of table 1 reported above, this value of P is strictly smaller than the generic rank, ω, except for $(d, K) = (3, 2)$ or $(d, K) = (3, 4)$. As a consequence, PARAFAC algorithms can only *approximate* $d-$way arrays, in general.

In the unsymmetric problem, the goal is to find three matrices, $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$, such that $G_{ijk} = \sum_p A_{ip}B_{jp}C_{kp}$. One possible numerical algorithm is based on alternating least squares, as explained below for $3-$way arrays [10] (Caroll and Chang 1970):

- Start with $(\boldsymbol{A}(0), \boldsymbol{B}(0), \boldsymbol{C}(0))$
- Define matrices $\boldsymbol{G}^{(1)}$, $\boldsymbol{G}^{(2)}$, $\boldsymbol{G}^{(3)}$:
  $\boldsymbol{G}_{ijk} = \boldsymbol{G}^{(1)}_{ip} = \boldsymbol{G}^{(2)}_{jq} = \boldsymbol{G}^{(3)}_{kr}$; $p = (jk)$, $q = (ik)$, $r = (ij)$
- Estimate stage $t+1$ from stage t by pseudo-inversion:
  - Update mode 1:
    $\boldsymbol{A}(t+1) = \boldsymbol{G}^{(1)}\,[\boldsymbol{B}(t)^{\mathrm{T}}\,\boldsymbol{C}(t)^{\mathrm{T}}]^{-}$
  - Update mode 2:
    $\boldsymbol{B}(t+1) = \boldsymbol{G}^{(2)}\,[\boldsymbol{A}(t+1)^{\mathrm{T}}\,\boldsymbol{C}(t)^{\mathrm{T}}]^{-}$
  - Update mode 3:
    $\boldsymbol{C}(t+1) = \boldsymbol{G}^{(3)}\,[\boldsymbol{A}(t+1)^{\mathrm{T}}\,\boldsymbol{B}(t+1)^{\mathrm{T}}]^{-}$

where $\boldsymbol{M}^{-}$ denotes the Moore-Penrose pseudo inverse of $\boldsymbol{M}$. See also [4] [26] [50] (Bro 1997; DeLathauwer 1997; Kruskal 1977) for more details on PARAFAC algorithms.

5.3 Sylvester theorem

As already pointed out earlier, a rank-one tensor is associated with a linear form raised to the dth power. In terms of polynomials, the CAND decomposition can thus be rephrased: how can one decompose a quantic into a sum of dth powers of linear forms [24] (Comon and Mourrain 1996) ? This is this topic that addresses this theorem, restricted to the binary case however (*i.e.*, two variables).

Theorem 5.1 *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

$$\begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_\omega \\ \vdots & & & \vdots \\ \gamma_{d-\omega} & \cdots & \gamma_{d-1} & \gamma_d \end{bmatrix} \boldsymbol{g}=\mathbf{0}. \tag{20}$$

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 [18] [24] (Comon 1998; Comon and Mourrain 1996). 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 (20), but there are two of them when d is even. This theorem shows that in column $K=2$ of table 2, we have $D=0$ when d is odd, and $D=1$ when d is even.

In [27] (DeLathauwer Comon and DeMoor 1999), several extensions to this theorem are proposed in the complex case. The basic idea remains the same, but the result becomes more complicated.

The disappointing fact is that Sylvester's theorem cannot be extended to dimensions higher than 2. In fact, a key step in the proof [24] [18] (Comon and Mourrain 1996; Comon 1998) is that for any polynomial p of degree d, and any monomial m of degree $d-\omega$, there exists a polynomial q of degree ω such that qm is orthogonal to p. Equation (20) expresses that orthogonality in terms of polynomial coefficients. It is clear that this holds true only when $d\geq\omega$, which is unfortunately satisfied only in the binary case, according to table 1. Possibilities of extension to more than 2 variables is discussed in [24] (Comon and Mourrain 1996).

Simultaneous CAND. Let us go back to table 2. Among others, this table reports that there are infinitely many CAND for even orders, d. In order to fix this indeterminacy in the case $(d,K,P)=(4,2,3)$ (the manifold of solutions is of dimension 1 in that situation), it is proposed in [18] (Comon 1998) to simultaneously diagonalize a second cumulant tensor of order 4.

The help of virtual sources. In [18] [21] (Comon 1998; Comon and Grellier 1999) 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 $(d, K, P) = (2, 2, 3)$.

In a few words, assume three sources x_i are mixed and received on two sensors, and assume these sources are all distributed in $\{-1, +1\}$ (they are called BPSK in digital communications). One can prove, if sources x_i are statistically independent, that the "virtual" source $x_1x_2x_3$ is also BPSK-distributed, but obviously statistically dependent of the three former ones. However, one can still prove that all its fourth-order pairwise cross-cumulants vanish. Yet, only *pairwise* cumulants are utilized in the sweeping strategies maximizing contrasts such as $\Upsilon_{2,4}$ in (7). As a consequence, viewed by the algorithm, sources are independent; one can thus build from $\boldsymbol{y}^{\mathrm{T}} = [y_1, y_2]$ virtual measurements y_1^3, $y_1^2y_2$, $y_1y_2^2$, and y_2^3, that can be modeled as linear mixtures of 4th order pairwise independent unknown sources. This allows the separation of the four sources (three actual and one virtual) from six sensors (two actual and four virtual).

5.4 Rank-one approximation

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 [28] [26] [49] (DeLathauwer Comon and others 1995; DeLathauwer 1997; Kofidis and Regalia 2000).

Criteria. Given tensor $\mathcal{C}^{\boldsymbol{y}}$, the goal is to find a vector $\boldsymbol{w}$ minimizing:

$$\Omega_o = ||\mathcal{C}^{\boldsymbol{y}} - \sigma\, \boldsymbol{w} \circ \boldsymbol{w} \circ \boldsymbol{w} \circ \boldsymbol{w}|| \tag{21}$$

for some scalar number σ. One can prove that minimizing (21) is equivalent to maximizing [18] (Comon 1998):

$$\Omega_d = ||\mathcal{C}^{\boldsymbol{y}} \bullet \boldsymbol{w} \bullet \boldsymbol{w} \bullet \ldots \bullet \boldsymbol{w}|| \tag{22}$$

or to minimizing:

$$\Omega_{d-1} = ||\mathcal{C}^{\boldsymbol{y}} \bullet \boldsymbol{w} \bullet \ldots \bullet \boldsymbol{w} - \lambda\, \boldsymbol{w}|| \tag{23}$$

However, the other criteria Ω_r, $0 < r < d-1$, are generally not equivalent.

Stationary uplets $(\boldsymbol{v}, \lambda)$ of Ω_o, Ω_{d-1} or Ω_d are the same and satisfy:

$$\mathcal{C}^{\boldsymbol{y}} \bullet \underbrace{\boldsymbol{v} \bullet \ldots \bullet \boldsymbol{v}}_{d-1 \text{ times}} = \lambda\, \boldsymbol{v}$$

this suggests a Rayleigh-like iteration, tat we can call the *Tensor Rayleigh symmetric iteration*:

$$\begin{aligned} \boldsymbol{w} &\leftarrow \mathcal{C}^{\boldsymbol{y}} \bullet \underbrace{\boldsymbol{w} \bullet \ldots \bullet \boldsymbol{w}}_{d-1 \text{ times}} \\ \boldsymbol{w} &\leftarrow \boldsymbol{w}/||\boldsymbol{w}|| \end{aligned}$$

In [28] (Delathauwer Comon et al. 1995), it is suggested to run a non symmetric iteration, and to initialize the algorithm with the HOSVD.

The rank-1 combination problem (section 4.4) has been solved in a sub-optimal way up to now in [70] [39] (Van der Veen and Paulraj 1996; Grellier and Comon 2000) by solving a large unconstrained linear system, and trying to restore the structure afterwards. The optimal one-stage solving still remains to be devised.

6 Concluding remarks

In this chapter, we have partly surveyed the tools dedicated to tensor decompositions, mainly through the problem of source separation. Thus, this presentation has been restrictive, but hopefully still informative.

Many other source separation algorithms do not resort to tensor tools, and have not been reported here. It is worth noting that some of them do not need the sources to be statistically independent, so that the output cumulant tensor is not even aimed at being diagonal. Instead, other properties of the sources can be exploited, such as their discrete character, or their constant modulus [70] [66] [39] (Van der Veen and Paulraj 1996; Talwar Viberg and Paulraj 1996; Grellier and Comon 2000). When more sources than sensors are present, general results state that it is sometimes possible to identify the mixture, but source extraction requires more knowledge about the sources (*e.g.*, their distribution). These issues have been tackled herein. Let us now turn to research perspectives.

In the area of source separation, current hot research topics include (i) blind identification of under-determined mixtures, (ii) blind equalization of convolutive mixtures, (iii) the theoretical proof of convergence of pair-sweeping algorithms, and, in the context of telecommunications, (iv) handling properly carrier residuals when present in the measurements. In all cases, analytical block-algorithms are suitable when computer power is available and when the stationarity duration is short.

As far as tensors are concerned, open research directions include: (i) the determination of the maximal achievable rank for arbitrary order and dimensions, (ii) the actual calculation of general Canonical Decompositions for $K > 2$, (iii) efficient numerical algorithms for computing an approximate of given rank.

[31]

References

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

[2] G. BIENVENU, L. KOPP, "Optimality of high-resolution array processing using the eigensystem approach", *IEEE Trans. ASSP*, vol. 31, no. 5, pp. 1235–1248, Oct. 1983.

[3] J. W. BREWER, "Kronecker products and matrix calculus in system theory", *IEEE Trans. on Circuits and Systems*, vol. 25, no. 9, pp. 114–122, Sept. 1978.

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

[5] 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.

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

[7] J. F. CARDOSO, "A tetradic decomposition of fourth order tensors: application to the source separation problem", in *SVD and Signal Processing III*, Elsevier, Ed., Leuven, Belgium, Aug. 22-25 1995, pp. 375–382.

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

[9] J. F. CARDOSO, A. SOULOUMIAC, "Blind beamforming for non-Gaussian signals", *IEE Proceedings - Part F*, vol. 140, no. 6, pp. 362–370, Dec. 1993, Special issue on Applications of High-Order Statistics.

[10] 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.

[11] J. D. CAROLL, S. PRUZANSKY, 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.

[12] 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.

[13] 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, Special issue on Applications of High-Order Statistics.

[14] P. COMON, "Separation of stochastic processes", in *Proc. Workshop on Higher-Order Spectral Analysis*, Vail, Colorado, June 28-30 1989, IEEE-ONR-NSF, pp. 174–179.

[15] P. COMON, "Independent Component Analysis", in *Proc. Int. Sig. Proc. Workshop on Higher-Order Statistics*, Chamrousse, France, July 10-12 1991, pp. 111–120, Republished in *Higher-Order Statistics*, J.L.Lacoume ed., Elsevier, 1992, pp 29–38.

[16] 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.

[17] P. COMON, "Tensor diagonalization, a useful tool in signal processing", in *IFAC-SYSID, 10th IFAC Symposium on System Identification*, M. Blanke, T. Soderstrom, Eds., Copenhagen, Denmark, July 4-6 1994, vol. 1, pp. 77–82, invited session.

[18] 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.

[19] P. COMON, "Block methods for channel identification and source separation", in *IEEE Symposium on Adaptive Systems for Sig. Proc. Comm. Contr.*, Lake Louise, Alberta, Canada, Oct. 1-4 2000, pp. 87–92, invited plenary.

[20] 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.

[21] P. COMON, 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.

[22] P. COMON, O. GRELLIER, "Analytical blind identification of a SISO communication channel", in *IEEE Workshop on Statistical Signal and Array Processing*, Pocono Manor, Pennsylvania, Usa, Aug. 14-16, 2000, pp. 206–210.

[23] P. COMON, O. GRELLIER, B. MOURRAIN, "Closed-form blind channel identification with MSK inputs", in *Asilomar Conference*, Pacific Grove, California, November 1-4 1998, pp. 1569–1573, invited session.

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

[25] R. COPPI, S. BOLASCO, Eds., *Multi-Way Data Analysis*, Elsevier Publ., 1889.

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

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] L. DeLATHAUWER, B. DeMOOR, J. VANDEWALLE, "Fetal electrocardiogram extraction by source subspace separation", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp. 134–138.

[32] L. DeLATHAUWER, B. DeMOOR, J. VANDEWALLE, "Fetal electrocardiogram extraction by blind source subspace separation", *IEEE Trans. Biomedical Engineering*, vol. 47, no. 5, pp. 567–572, May 2000, Special Topic Section on Advances in Statistical Signal Processing for Biomedicine.

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

[34] W. FELLER, *An Introduction to Probability Theory and its Applications*, Wiley, 1968.

[35] 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.

[36] 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.

[37] P. GELADI, "Analysis of multi-way data", *Chemom. Intel. Lab. Syst.*, vol. 7, pp. 11–30, 1989.

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

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

[40] 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., chapter 9. SIAM Publ., Philadelphia, PA, 1998.

[41] R. A. HARSHMAN, "Determination and proof of minimum uniqueness conditions for parafac1", *UCLA Working Papers in Phonetics*, vol. 22, pp. 111–117, 1972.

[42] S. HAYKIN, *Unsupervised Adaptive Filtering*, vol. 1, Wiley, 2000, series in Adaptive and Learning Systems for Communications, Signal Processing, and Control.

[43] T. D. HOWELL, "Global properties of tensor rank", *Linear Algebra and Applications*, vol. 22, pp. 9–23, 1978.

[44] A. HYVÄRINEN, P. HOYER, E. OJA, "Denoising of non gaussian data by Independent Component Analysis and sparse coding", in *ICA99, IEEE Workshop on Indep. Comp. Anal. and Blind Source Separation*, Aussois, France, Jan 11–15 1999, pp. 485–489.

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

[46] C. JUTTEN, J. HÉRAULT, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture", *Signal Processing, Elsevier*, vol. 24, no. 1, pp. 1–20, 1991.

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

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

[49] E. KOFIDIS, 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.

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

[51] J. P. KUNG, G. C. ROTA, "The invariant theory of binary forms", *Bulletin of the Am. Math. Soc.*, vol. 10, no. 1, pp. 27–85, Jan. 1984.

[52] 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.

[53] 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, Nov. 1999.

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

[55] J. LEVIN, "Three-mode factor analysis", *Psychological Bulletin*, vol. 64, pp. 442–452, 1965.

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

[57] 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.

[58] O. MACCHI, E. MOREAU, "Self-adaptive source separation. part I: Convergence analysis of a direct linear network controlled by the Herault-Jutten algorithm", *IEEE Trans. Sig. Proc.*, vol. 45, no. 4, pp. 918–926, Apr. 1997.

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

[60] 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.

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

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

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

[64] I.R. SHAFAREVITCH, *Basic Algebraic Geometry*, Springer Verlag, 1977.

[65] 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.

[66] S. TALWAR, M. VIBERG, 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.

[67] 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.

[68] L. R. TUCKER, "Some mathematical notes for three-mode factor analysis", *Psychometrika*, vol. 31, pp. 279–311, 1966.

[69] 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.

[70] 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.