TENSORS VERSUS MATRICES USEFULNESS AND UNEXPECTED PROPERTIES

Pierre Comon

I3S Laboratory UMR6070, CNRS and University of Nice-Sophia Antipolis (UNS), France

ABSTRACT

Since the nineties, tensors are increasingly used in Signal Processing and Data Analysis. There exist striking differences between tensors and matrices, some being advantages, and others raising difficulties. These differences are pointed out in this paper while briefly surveying the state of the art. The conclusion is that tensors are omnipresent in real life, implicitly or explicitly, and must be used even if we still know quite little about their properties.

***Index Terms*—** Tensor rank, Canonical decomposition, Factor analysis, Parafac, High-order statistics, Separation of variables, Low-rank approximation, Deflation, Blind identification

1. INTRODUCTION

1.1. Overview

This paper summarizes several tutorials I gave on tensors since the IEEE SP *Workshop on High-Order Statistics* held at Banff in 1997, and in particular for the last five years. Some of the results I presented were quite standard, and others were the fruit of cooperations with colleagues, who are cited in references.

Various definitions of rank are given in Section 2 and properties are emphasized in Section 3. In particular: (i) the rank of a tensor generally exceeds its dimension, (ii) there exist tensors having a rank larger than generic, (iii) the maximal achievable rank of a tensor is not clearly known, (iv) the decomposition of a tensor into a sum of rank-1 terms can often be performed in an essentially unique manner.

In Section 4, it is shown that imposing a structure in the rank-1 decomposition of a tensor may change the rank, even if the structure is the same as the tensor to be decomposed. It is shown for instance that (v) the rank can differ in real and complex fields, and that (vi) a structure as simple as symmetry has not be proved to have no influence on rank. In Section 5, it is pointed out that (vii) searching for the best low-rank approximation of a tensor is often an ill-posed problem. An even more surprising fact is enlightened in Section 5.3: (viii) subtracting the best rank-1 approximate generally increases tensor rank. In Section 6, some algorithms are quoted, which allow to compute the canonical decomposition that reveals the tensor rank, but generally (ix) it is difficult to assess the rank of a tensor and to compute its decomposition into a sum of rank-1 terms.

Despite all these odd properties, which reveal our ignorance in the domain, tensors still remain attractive because they allow to restore identifiability in several problems, as explained in Section 7.

1.2. Notations

Vectors are denoted in bold lowercases, matrices in bold uppercases, and tensors in bold slanted uppercases. A matrix defines either a linear operator or a bilinear form, once the bases of the vector spaces are fixed. If bases are changed, so is the matrix representing the operator.

Similarly, a tensor $\boldsymbol{T}$ of order d, with coordinates $T_{i_1..i_d}$ represents a multi-linear operator, defined from a space outer product, $\otimes_i \mathcal{S}_i$, to another, $\otimes_j \mathcal{S}_j$. More precisely, a tensor of order d can be seen as an array with d indices, provided bases of d spaces $\mathcal{S}_k$ have been defined. Denote by N_k the dimension of these spaces $\mathcal{S}_k$, $1 \leq k \leq d$, and take a tensor of order $d = 3$ and dimensions $N_1 \times N_2 \times N_3$ to simplify the presentation. Such a tensor can define a trilinear form, or a bilinear operator. Thus, there are *several ways* to associate a tensor with an operator. If a change of bases is defined by 3 matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ in spaces $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ respectively, then the new tensor coordinates T'_{ijk} take the form

$$T'_{ijk} = \sum_{pqr} U_{ip}\, V_{jq}\, W_{kr}\, T_{pqr} \tag{1}$$

This *multi-linear property* is a specificity of tensors [1, 2]. It is often written in compact form as $\boldsymbol{T}' = (\mathbf{U}, \mathbf{V}, \mathbf{W}) \cdot \boldsymbol{T}$. Note that matrices also enjoy this property when they represent an operator, since they are then tensors of order 2.

Some tensors have the particular property of having "separated variables", which means that they take the form

$$T_{ijk} = u_i\, v_j\, w_k$$

or in compact form $\boldsymbol{T} = \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$. One can indeed view such a tensor expression as a discretized version of a functional equation $f(x, y, z) = u(x)\, v(y)\, w(z)$ representing a

Work funded by contract ANR-06-BLAN-0074 "DECOTES".

separation of variables. The latter tensors will be referred to as *rank-1 tensors*.

Tensors whose dimensions N_k are all equal may be said to be *cubic* [3]. A square matrix is a cubic tensor of order 2, for instance. Tensors whose entries do not change by permuting arbitrarily their indices are referred to as *symmetric*.

Symmetric tensors of order d are encountered every time dth order derivatives of a multivariate function are utilized. For instance, moments or cumulants of a Nth dimensional random variable are symmetric tensors of dimension N [2, 4]. The covariance matrix is a particular case of cumulant tensor, of order 2.

Example 1 *For convenience, let's write the values of a $2 \times 2 \times 2$ tensor in two 2×2 matrices stacked one after the other. Here is an example of a rank-1 symmetric tensor:*

$$\boldsymbol{T}_1 = \left[\begin{array}{rr|rr} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{array} \right] \tag{2}$$

It is in fact the triple outer product of vector $\mathbf{v} = [1,\ -1]$ with itself.

2. RANKS

In this paper, we are mainly interested in decomposing tensors into a sum of rank-1 terms, as:

$$T_{ijk} = \sum_{r=1}^{R} u_i(r)\, v_j(r)\, w_k(r) \tag{3}$$

which is denoted as $\boldsymbol{T} = \sum_r \mathbf{u}(r) \otimes \mathbf{v}(r) \otimes \mathbf{w}(r)$ in compact form, where as before $\otimes$ stands for the tensor (outer) product.

Tensor rank. One defines the *rank* of a tensor as the minimal number of rank-1 terms that are necessary for equality (3) to hold. It is sometimes denoted as $R_\otimes(\boldsymbol{T})$, or simply $R(\boldsymbol{T})$. The *Canonical decomposition* (CAND) of a tensor $\boldsymbol{T}$ is the decomposition (3) obtained with the smallest number of terms, $R(\boldsymbol{T})$. The CAND can be traced back to [5] [1][6] [7], but it has been rediscovered in the seventies by psychometricians and chemometricians, who gave it other names (*e.g.* PARAFAC).

In the case of matrices, the CAND is not unique, since there are infinitely many ways of decomposing a matrix of rank R into a sum of R rank-1 terms. This is the reason why it is possible to impose orthogonality among the left (resp. right) singular vectors of a matrix. But for tensors of order higher than 2, the CAND can be unique, up to scale and permutation ambiguities. That's why the CAND is so attractive in applications.

Multilinear rank. There exist other definitions of rank for tensors. Some are pointed out in Section 4. But the simplest one consists of associating the tensor with a *linear operator*. Yet, there are $2^d - 2$ nontrivially different ways of making this association. If one restricts to operators mapping $\otimes_{i \neq k} \mathcal{S}_i$ to $\mathcal{S}_k$, $1 \leq k \leq d$, represented by a matrix of size $N_k \times \prod_{i \neq k} N_i$, there are d possible such ranks, which are usually called $k-$*mode ranks*. Hence, the $k-$mode rank of a tensor obviously cannot exceed its kth dimension, N_k. These matrices have found various names in the literature, *e.g.*, *unfolding* or *flattening* matrices.

The *multilinear rank* is defined as the $d-$uple of these mode ranks. For a matrix, the two *mode ranks* are equal, and coincide with the tensor rank defined by the CAND. For a tensor they can be different.

Example 2 *Take now the symmetric tensor*

$$\boldsymbol{T}_3 = \left[\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{array} \right] \tag{4}$$

One can check out that this tensor has at most rank 3 since:

$$2\, \boldsymbol{T}_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}^{\otimes 3} + \begin{pmatrix} -1 \\ 1 \end{pmatrix}^{\otimes 3} - 2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}^{\otimes 3}$$

It turns out that there exist no decomposition with only two terms, so that tensor $\boldsymbol{T}_3$ has rank 3; this can be proved with the help of Sylvester's theorem, see e.g. *[8, 9]. On the other hand, the multilinear rank of $\boldsymbol{T}_3$ is equal to $(2, 2, 2)$.*

This is the simplest example demonstrating that the tensor rank can exceed dimensions, a fact that does not exist for matrices.

Eigenvalues. Contrary to the Singular Value decomposition (SVD), which reveals the eigenvalues of some linear operator, the CAND is not directly linked to the eigenvalues of the linear operators that can be associated with a tensor, at least to our current knowledge. Some attempts have been made to define eigenvalues of tensors, but this question is still considered to be open [10, 11, 12].

The Waring problem. When dealing with symmetric tensors, it is often useful to make the connection with homogeneous polynomials. More precisely, if $\boldsymbol{T}$ is symmetric of order d and dimension N, it can be bijectively associated with the polynomial

$$p(\mathbf{x}) = \sum_{i,j,..k=1}^{N} c_{ij..k}\, T_{ij..k}\, x_i\, x_j\, \ldots x_k$$

of degree d in N variables, where $\mathbf{x} = (x_1, x_2, \ldots, x_N)$. The coefficient $c_{ij..k}$ could be arbitrarily set to 1, but it is more convenient to set it equal to the multinomial coefficient [13, 9]. Through this bijection, a vector $\mathbf{u}$ is mapped to a linear form $\ell(\mathbf{x}) = \mathbf{u}^\mathsf{T}\mathbf{x}$. Now because of the bijection, the CAND can be stated in terms of polynomials as:

$$p(\mathbf{x}) = \sum_{p=1}^{R(T)} \ell_p(\mathbf{x})^d$$

Given polynomial $p(\mathbf{x})$, finding the minimal number of linear forms $\ell_p(\mathbf{x})$ is known as the Waring problem [14].

Example 3 *Take the tensor of example 2. The associated polynomial is* $p(x_1, x_2) = 3\,x_1^2\,x_2$ *and the corresponding* CAND *is:*

$$2\,p(x_1, x_2) = (x_1 + x_2)^3 + (-x_1 + x_2)^3 - 2\,x_2^3$$

Orbits. The orbit of a tensor $\boldsymbol{T}_0$ is the set of tensors that are obtained by applying an invertible multilinear transform on $\boldsymbol{T}_0$. Obviously, all tensors in the same orbit have the same rank, and the same multilinear rank.

Example 4 *By permuting the two rows in the 1st mode, one easily checks out that the non symmetric tensor below is in the orbit of* $\boldsymbol{T}_3$*, studied in examples 2 and 3:*

$$\boldsymbol{T}'_3 = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right] \quad (5)$$

and is hence of rank 3.

3. GENERICITY

Informally, one says that a property is *typical* if it holds true on a nonzero-volume set. A property is *generic* if is true almost everywhere. In other words, there can be several typical ranks, but only a single generic rank [15, 16]. In algebraic geometry, it has been proved that a generic homogeneous polynomial always assumes the same rank value in the complex field. Hence there exist a single generic rank for symmetric tensors in $\mathbb{C}$. Thus the distinction between typical and generic ranks is irrelevant in the complex field [3, 17]. Nevertheless, it is convenient to make the distinction in the present framework by reserving the use of the terminology of *generic rank* to decompositions in $\mathbb{C}$. We should note that the notion of zero-volume (or zero-measure) is related to an underlying topology. In algebraic geometry, this is the Zariski topology. But these results hold true for other topologies, including Euclidean [4].

Importance. The consequence is that real or complex symmetric tensors with entries randomly drawn according to a continuous probability distribution are generic, and always have the same rank in the complex field. Considering that in Engineering, measurements are always corrupted by noise, and that this noise is generally assumed additive and has a continuous probability density, we see that estimated tensors will be generic most of the time. That's why the study of generic tensors is of prime importance.

Uniqueness. One could believe that uniqueness is ensured, at least up to permutation and scale, if the number of free parameters are the same on both sides of (3). It turns out that this is not true. Clebsch was the first to notice in 1850 that ternary quartics cannot generally be written as a sum of 5 fourth powers of linear forms. Table 1 reports the values of the rank of generic symmetric tensors, and one can indeed notice the value of 6 for $(d, N) = (4, 3)$. More precisely, for symmetric tensors, left and right hand sides of (3) can have the same number of free parameters only if the quantity F_s (resp. F for cubic tensors with unconstrained entries) below is an integer

$$F_s = \frac{1}{N}\binom{N+d-1}{d}; \quad F = \frac{N^d}{Nd - d + 1}$$

	N	2	3	4	5	6	7	8
d	2	2	3	4	5	6	7	8
	3	2	4	5	8	10	12	15
	4	3	6	10	15	21	30	42

Table 1. Generic rank $\bar{R}_s$ of symmetric tensors of order d and dimension N. In the real field, it corresponds to the smallest typical rank.

In Tables 1 (resp. Table 2), the exceptions to the ceil rule $\bar{R}_s = \lceil F_s \rceil$ (resp. $\bar{R} = \lceil F \rceil$) are distinguished by an underlining. When the ratios F_s or F are integers, and when we do not face an exception, then there is a finite number of CAND's, and the corresponding value of the generic rank is represented in a frame in Tables 1 and 2. Otherwise, the generic rank $\bar{R}_s > F_s$ (resp. $\bar{R} > F$) and the CAND is not unique. See [18] for further practical details.

For symmetric tensors in the complex field, all the values of the generic rank are now known and proved theoretically: this is the Waring problem [14], for which it has been proved in 1996 by Alexander and Hirschowitz that the number of exceptions is finite. On the other hand, for tensors with free entries, most values have been found by computer simulations (in bold in Table 2). Finiteness of exceptions for non symmetric tensors is still a conjecture.

As shown in Table 2, generic rank values are larger when tensors have free entries, which makes sense. Generic ranks of non cubic tensors can be found in [16], as well as tensors of other particular forms.

	N	2	3	4	5	6	7	8
d	2	2	3	4	5	6	7	8
	3	2	5	**7**	**10**	**14**	**19**	**24**
	4	**4**	**9**	**20**	**37**	**62**	**97**	

Table 2. Generic rank $\bar{R}$ of free tensors of order d and equal dimensions, N. In the real field, it corresponds to the smallest typical rank.

Beside the generic cases appearing in framed boxes in Tables 1 or 2, can we have a finite number of possible CAND for a given tensor? There are two results that can help us.

Necessary condition: if the rank of a tensor $\boldsymbol{T}$ is strictly smaller than the expected rank:

$$R_s(\boldsymbol{T}) < F_s \quad \text{or} \quad R(\boldsymbol{T}) < F$$

then there is almost surely a finite number of CAND. In fact, there is a set of zero-volume in which the number of CAND can be infinite (cf. Example 5 below; see [16] for a summary of complementary results).

Example 5 *Take symmetric tensors of order $d = 4$ and dimension $N = 3$. From Table 1 we see that generic tensors admit a finite number of CanD in $\mathbb{C}$ in that case, and that the generic rank is 5. But by drawing randomly such tensors, there is a zero probability to draw a tensor in the orbit of $x_1^2 x_2$, which has infinitely many CanD's as in Example 3.*

Sufficient condition: if the rank of a tensor $\boldsymbol{T}$ satisfies the condition below:

$$2\,R(\boldsymbol{T}) + d - 1 \leq \sum_{i=1}^{d} N_i$$

then the CAND is unique. The above condition is a simplification of the exact statement, and is true with probability one; see [19] for the exact statement, involving the concept of k-rank of a matrix, introduced by Kruskal [20].

Another useful sufficient condition leading to almost sure uniqueness can be found in [21] for tensors having one dimension much larger than the others.

Maximal rank. The maximal achievable rank of a tensor is not clearly known, only bounds exist. But we know that there always exist tensors having a rank larger than generic, as soon as the order exceeds 2 (which is the matrix case). Remember that the rank of a matrix cannot exceed its dimensions.

Example 6 *In Tables 1 and 2, we see that 3rd order cubic tensors of dimension 2 are generically of rank 2, should they be symmetric or not. But tensors studied in Examples 2 and 4 had a rank equal to 3.*

Example 7 *There is a one to one correspondence between the space of $3 \times 3 \times 3$ symmetric tensors and ternary cubics (homogeneous polynomials of degree 3 in 3 variables). In $\mathbb{C}$ the generic rank is equal to 4 [9]. But polynomials lying in the orbit of polynomial $x^2y + xz^2$ are of rank 5, which is the maximal achievable rank of ternary cubics in $\mathbb{C}$ [22, 23, 9].*

Hence there exist tensors having a rank larger than generic, but they are difficult to find because they form a set of zero volume.

4. RANK UNDER CONSTRAINT

Rank depends on the field. The results obtained so far do not apply directly to real tensors, because it was assumed that the underlying field was algebraically closed. In fact, there may exist several typical ranks in the real field. The only thing we know is that the smallest typical rank coincides with the generic rank computed in $\mathbb{C}$.

Example 8 *Consider again symmetric tensors with $(d, N) = (3, 2)$, but this time in $\mathbb{R}$. Beside the zero tensor, there are 4 orbits Ω_i [24]. Ω_1: rank-1 tensors generated by polynomial x^3, Ω_2: rank-2 generated by $x^3 + y^3$, Ω_3: rank-3 generated by x^2y, and Ω_4: rank-3 generated by $xy^2 - x^3$. It turns out that there are 2 typical orbits in $\mathbb{R}$, namely Ω_2 and Ω_4 and 2 negligible orbits, Ω_1 and Ω_3. In $\mathbb{C}$, Ω_4 does not exist, because it is included in Ω_2; this can be seen by observing that Ω_2 is mapped to Ω_4 by the transform $(x, y) \to (x + \imath y, x - \imath y)$.*

The conclusions are similar as Example 8 for tensors with free entries [25]: there can be several typical ranks. ten Berge conjectured that there can only be at most two typical ranks for $p \times q \times 2$ tensors. But tensors with dimensions larger than 2 are likely to admit more than 2 typical ranks in $\mathbb{R}$.

Now what happens to the rank of a given real tensor? We have for any real tensor $\boldsymbol{T}$:

$$\text{rank}\{\boldsymbol{T}\}_{\mathbb{C}} \leq \text{rank}\{\boldsymbol{T}\}_{\mathbb{R}} \tag{6}$$

Example 9 *The $2 \times 2 \times 2$ tensor :*

$$\boldsymbol{T}_2 = \left[\begin{array}{cc|cc} -1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right]$$

is of rank 3 in $\mathbb{R}$, but 2 in $\mathbb{C}$ [4]. In fact we have in $\mathbb{R}$:

$$\boldsymbol{T}_2 = \frac{1}{2}\begin{pmatrix}1\\1\end{pmatrix}^{\otimes 3} + \frac{1}{2}\begin{pmatrix}1\\-1\end{pmatrix}^{\otimes 3} + 2\begin{pmatrix}-1\\0\end{pmatrix}^{\otimes 3}$$

but we have in the complex field:

$$\frac{\imath}{2}\begin{pmatrix}-\imath\\1\end{pmatrix}^{\otimes 3} - \frac{\imath}{2}\begin{pmatrix}\imath\\1\end{pmatrix}^{\otimes 3}.$$

So inequality (6) may happen to be strict.

Can the symmetry constraint change rank? Constraining the CAND to be real can increase rank, as we have just seen. Similarly, since constraining the CAND can only increase rank, we have:

$$R(\boldsymbol{T}) \leq R_s(\boldsymbol{T}) \tag{7}$$

However, it has not been possible to provide examples of symmetric tensors for which the symmetric rank is strictly larger than the free rank. Put in other words, it seems that inequality (7) is always an equality. However, the result $R(\boldsymbol{T}) = R_s(\boldsymbol{T})$, $\forall \boldsymbol{T}$ symmetric, has only been partially proved in [4], and hence still remains a conjecture for most sub-generic ranks.

5. LOW-RANK APPROXIMATION

In this section, we are interested in finding r families of d vectors, $\{\mathbf{u}(p),\ \mathbf{v}(p), \dots\ \mathbf{w}(p)\}$, $1 \leq p \leq r$, such that the criterion below is minimized for a given rank value $r < R(\boldsymbol{T})$:

$$||\boldsymbol{T} - \sum_{p=1}^{r} \mathbf{u}(p) \otimes \mathbf{v}(p) \otimes \dots \otimes \mathbf{w}(p)||^2 \qquad (8)$$

5.1. Rank-1 approximation

It is well known that the set of rank-1 tensors is closed, as a determinantal variety [26, 27]. Yet, its computation still raises difficulties, due to the multimodal character of the criterion of the form (8), even for $r = 1$. If the Frobenius norm is used, one can think of running a power iteration [28]. Sufficient convergence conditions have been obtained for symmetric tensors in [29]. Also note that there exist a closed form solution for 3rd order tensors of dimension 2 [30].

5.2. Rank-r approximation

In applications, approximating a tensor by another of given lower rank, $r > 1$, is often what is actually sought for. However, this question is well posed only if the set $\mathcal{Y}_r$ of tensors of rank at most r is closed. But it turns out that $\mathcal{Y}_r$ is generally not closed, unless $r = 1$ or r is maximal (i.e. when $\mathcal{Y}_r$ is the whole space). Note that the proof that every set $\mathcal{Y}_r$ is not closed for $1 < r < R_{max}$ given in [4] is yet incomplete.

Example 10 *In order to prove that $\mathcal{Y}_r$ is not closed, it suffices to find a sequence of rank$-r$ tensors converging to a tensor of strictly higher rank. Take 3 linearly independent vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ and build the following sequence of 3rd order symmetric tensors (the principle extends to non symmetric tensors, with a somewhat more complicated notation):*

$$\begin{aligned}\boldsymbol{T}(n) &= n^2(\mathbf{x} + \frac{1}{n^2}\,\mathbf{y} + \frac{1}{n}\,\mathbf{z})^{\otimes 3} \\ &+ n^2(\mathbf{x} + \frac{1}{n^2}\,\mathbf{y} - \frac{1}{n}\,\mathbf{z})^{\otimes 3} - 2n^2\,\mathbf{x}^{\otimes 3}\end{aligned}$$

For every n, the above tensor is of rank 3, but it converges to the limit below, which is of rank 5 [18]:

$$\begin{aligned}\boldsymbol{T}(\infty) &= \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{y} + \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{x} + \mathbf{y} \otimes \mathbf{x} \otimes \mathbf{x} \\ &+ \mathbf{x} \otimes \mathbf{z} \otimes \mathbf{z} + \mathbf{z} \otimes \mathbf{x} \otimes \mathbf{z} + \mathbf{z} \otimes \mathbf{z} \otimes \mathbf{x}\end{aligned}$$

This tensor is in fact associated with the polynomial $3x^2y + 3xz^2$, belonging to the orbit of rank-5 ternary cubics (cf. Example 7).

Examples in dimension 2 have already be reported as well [4, section 6] [25]. For instance, one can build a sequence of 2-dimensional cubic tensors of order d and rank 2 converging to a tensor of rank d. Note that the first example of lack of closeness was probably due to Bini et al. in 1979; see [25].

In general, the approximation problem is ill-posed, which means that not every tensor will have a low-rank approximate. However, there exist well-posed tensor approximation problems [31], as now shown.

Restriction to orthogonal transforms. Consider the case where columns $\mathbf{u}(p)$, (resp. $\mathbf{v}(p)$ and $\mathbf{w}(p)$) involved in (8) are constrained to be orthogonal. Then $\boldsymbol{T}$ must be cubic and r cannot exceed its dimension, which is a strong limitation. But on the other hand, according to Weierstrass theorem, a continuous function defined on a compact set reaches its extrema. Yet, the orthogonal group is compact, because bounded and closed in a finite dimensional space. On the other hand, a function such as (8) is continuous. Hence, the minimum always exists. This is why approximate orthogonal tensor diagonalization has been the subject of numerous researches for the last fifteen years [9]; Independent Component Analysis is one instance of this diagonalization problem.

Regularization of the criterion. A well-known way to deal with ill-posed problems is to add regularization terms; in [32] for instance, the regularized criterion consists of function (8) with additional terms proportional to the norm of each unknown. Yet, from the constrained optimization theory, we know that, under some regularity conditions, minimizing a function $f(x_1, \dots x_k)$ under the constraints $||x_i||^2 = \mu_i, i \in \{1, \dots, k\}$ is equivalent to minimizing the functional $f(x_1, \dots x_k) + \sum_i \lambda_i ||x_i||^2$, where the reals λ_i should be chosen so as to satisfy the constraints. Yet, in a finite dimensional space, the unit sphere is compact, and so is the set defined by $||x_i||^2 = \mu_i, i \in \{1, \dots, k\}$. Thus by the same argument as above, the continuous function $f(x)$ reaches its extrema on the latter set.

Restriction to real positive tensors. In a number of contexts, the tensors considered have real nonnegative entries, as well as the rank-1 terms appearing in the CAND. This is the case for instance of tensors built on power spectral densities or images (see Section 7). Now criterion (8) is to be optimized over the cone of vectors with real positive entries. It is shown in [31] that this is a well-posed problem for any r.

Border rank Another more natural way of facing the problem of ill-posedness is to search the closure $\bar{\mathcal{Y}}_r$ of $\mathcal{Y}_r$ instead of $\mathcal{Y}_r$ itself. By doing so, the rank of the approximation may be as large as the *border rank* of $\mathcal{Y}_r$; see [33, 3, 25] for this terminology. However, we solve a different problem, which may not be satisfactory in practice.

5.3. Subtraction of the best rank-1 approximate

It is now known that one cannot estimate the tensor rank by subtracting successive rank-1 approximates [29]. The reason is that the tensor rank may not decrease this way. This makes sense, otherwise the rank$-r$ approximation problem would be well-posed, since the rank-1 was. But it can be shown that subtracting the best rank-1 approximation generally even increases tensor rank [24], which is more surprising.

Example 11 *The tensor*

$$\boldsymbol{T} = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \end{array}\right]$$

is of rank 2. Its best rank-1 approximate is [24]:

$$\boldsymbol{Y} = \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{array}\right]$$

And one checks out that the difference

$$\boldsymbol{T} - \boldsymbol{Y} = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

is of rank 3, because in lies in the orbit of Example 4.

6. COMPUTATION

The computation of the CAND of a given tensor remains a difficult task, even if the rank is known and the CAND essentially unique. The most often encountered way to compute the CAND is to resort to standard optimization algorithms such as the conjugate gradient [34], Newton [35, 36], Alternating Least squares [37, 38, 36, 18], or Levenberg-Marquardt [35, 18] algorithms. All these algorithms have a common limitation: they can lead to a local minimum of (8) even for $r = R(\boldsymbol{T})$, that is, to a nonzero error, if a good initial guess is not available.

A global line search can be combined with any of these iterative algorithms in order to avoid local minima; since the global minimum is found only in a one- or two-dimensional space for complexity reasons, the technique has been called *Enhanced Line Search* (ELS) [39, 40].

Nevertheless, in the case of 2-dimensional symmetric tensors, a theorem by Sylvester allows to compute the CAND, either in the real or complex field, and only requires to compute the kernel of some Hankel matrix [17, 9]. This algorithm has been recently extended to symmetric tensors of larger dimensions, but only in the complex field [8], and basically works for ranks below generic, and when the CAND admits a finite number of solutions.

For tensors of order 3 with one dimension equal to 2, techniques based on matrix pencils have been shown to be very attractive [15]. The CAND can also be calculated with the help of the Generalized Schur Decomposition [41, 42]. Other algorithms have been proposed for tensors with specific shapes, in particular in [21] for 3rd order tensors when the rank is smaller than the largest dimension.

7. APPLICATIONS

As already said in introduction, the CAND may be seen as a discrete version of the functional equation

$$f(x,y,z) = \sum_p u_p(x)\, v_p(y)\, w_p(z) \qquad (9)$$

A1. Limiting the sum (9) to a finite number of terms allows to reduce the computational complexity in several problems, for instance in the approximation of integral or differential operators [43].
A2. Another interest in the CAND in mathematics is the evaluation of the asymptotic arithmetic complexity of the product of two matrices [44, 3, 45, 46]. The tensor to be decomposed is then of order 3, and represents this bilinear map.
A3. The CAND (3) is widely utilized in Multi-way Factor Analysis [37]. It appears in several quite different branches of engineering [38, 47, 48, 9, 49]. Let's review some of them.
A4. In fluorescence spectroscopy for instance, an optical excitation at wavelength λ_e produces a fluorescence emission at wavelength λ_f with a fluorescence intensity that can be modeled as indicated below, if the Beer-Lambert law can be linearized (that is, at low concentrations):

$$I(\lambda_f, \lambda_e, k) = I_o \sum_\ell \gamma_\ell(\lambda_f)\, \epsilon_\ell(\lambda_e)\, c_{k,\ell}$$

where k denotes an experiment index, which may correspond to different concentrations or temperatures. We recognize a tensor CAND, if wavelengths are discretized [38]. Note that the tensors are here real and *nonnegative*.
A5. Another direct application occurs in probability, when writing the joint distribution $q(\mathbf{x})$ of a $d-$dimensional random variable $\mathbf{X}$ conditionally to a discrete variable Θ taking r possible values θ_ℓ with probability $\mu(\theta_\ell)$. If random variables X_n are conditionally independent, we have, applying Bayes rule:

$$q(\mathbf{x}) = \sum_{\ell=1}^{r} \mu(\theta_\ell)\; q_1(x_1 \mid \theta_\ell)\, q_2(x_2 \mid \theta_\ell) \cdots q_n(x_n \mid \theta_\ell)$$

If variables X_n are themselves all discrete, the rule takes the form of the CAND of a dth order real *nonnegative* tensor [31].
A6. Image processing is also an example where tensors have real nonnegative entries [50, 51, 52, 53].
A7. Modeling the signals received on an array of antennas generally leads to a matrix or tensor formulation. For instance in digital communications, the signal received on the antenna array takes the form

$$T_{ijp} = \sum_\ell \sum_q a_{i\ell q} \sum_k h_{\ell qkp} s_{j\ell k}$$

where $i,\ j,\ k,\ \ell,\ q$ denote the space, time, symbol time, transmitter and path indices, respectively; $\mathbf{a}$ characterizes the receiver geometry, $\mathbf{h}$ the global channel impulse response, and

$\mathbf{s}$ the transmitted signal. Index p appears only in the presence of additional *diversity*, induced by oversampling, polarization, coding, geometrical invariance, wide frequency band, nonstationarity... See e.g. [54, 55] for an introduction.

A8. In some problems however, the diversity is lacking in the measurements. This may occur for instance when the matrix slices of a third order tensor are proportional. In such a case, the CAND is not unique, because the tensor decomposition reduces to a matrix one. In statistical signal processing, there are several ways to face this problem, by taking advantage of another property.

Consider the simplest linear statistical model

$$\mathbf{x} = \mathbf{H}\,\mathbf{s}$$

where realizations of vector $\mathbf{x}$ are observed. In other words, a matrix X_{ij} is given. The goal is to identify matrix $\mathbf{H}$, which is assumed to mix statistically independent random variables s_i. The second joint characteristic function Ψ_x of variables x_i may be shown to satisfy the core equation

$$\Psi_x(\mathbf{u}) = \sum_\ell \Psi_{s_\ell}\left(\sum_q u_q\, H_{q\ell}\right)$$

where Ψ_{s_ℓ} denotes the 2nd characteristic functions of variable s_ℓ. As pointed out in Section 1, the dth order derivative of such a function yields a symmetric tensor of order d. At order 3 for instance, this tensor can be modeled as

$$T_{ijk}(\mathbf{u}) = \sum_\ell \mu_\ell(\mathbf{u})\, H_{i\ell}\, H_{j\ell}\, H_{k\ell}$$

where $\mu_\ell(\mathbf{u})$ stands for the third derivative of Ψ_{s_ℓ} taken at $\sum_q u_q\, H_{q\ell}$. At $\mathbf{u} = 0$, this is a cumulant matching equation, and we face the CAND of a *3rd order symmetric* tensor. If a grid of values is chosen for $\mathbf{u}$, then we have the CAND of a *4th order non symmetric* tensor [56].

In all the applications mentioned above, it was necessary to resort to tensors in order to solve the problem. In fact, if stated in matrix form, the above problems would have infinitely many solutions. From the engineer point of view, tensors have allowed to restore identifiability.

8. REFERENCES

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

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

[3] T. LICKTEIG, "Typical tensorial rank," *Linear Algebra Appl.*, vol. 69, pp. 95–120, 1985.

[4] P. COMON, G. GOLUB, L-H. LIM, and B. MOURRAIN, "Symmetric tensors and symmetric tensor rank," *SIAM Jour. Matrix Ana. Appl.*, vol. 30, no. 3, pp. 1254–1279, 2008.

[5] F. L. HITCHCOCK, "Multiple invariants and generalized rank of a p-way matrix or tensor," *J. Math. and Phys.*, vol. 7, no. 1, pp. 40–79, 1927.

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

[7] D. A. WEINSTEIN, "Canonical forms for symmetric tensors," *Linear Algebra Appl.*, pp. 271–282, 1984.

[8] J. BRACHAT, P. COMON, B. MOURRAIN, and E. TSIGARIDAS, "Symmetric tensor decomposition," *Linear Algebra and its Applications*, 2009, submitted. hal:inria-00355713, arXiv:0901.3706.

[9] P. COMON, "Tensor decompositions, state of the art and applications," in *Mathematics in Signal Processing V*, J. G. McWhirter and I. K. Proudler, Eds., pp. 1–24. Clarendon Press, Oxford, UK, 2002, arXiv:0905.0454v1.

[10] L-H. LIM, "Singular values and eigenvalues of tensors: a variational approach," in *IEEE Int. Workshop on Comput. Adv. Multi-Sensor Adapt. Proc.*, Puerto Vallarta, Mexico, 13-15 Dec. 2005.

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

[12] L. QI, "Rank and eigenvalues of a symmetric tensor, the multivariate homogeneous polynomial and the algebraic hypersurface it defines," *Jour.Symbolic Computation*, vol. 41, pp. 1309–1327, 2006.

[13] R. EHRENBORG and G. C. ROTA, "Apolarity and canonical forms for homogeneous polynomials," *European Jour. Combinatorics*, vol. 14, pp. 157–181, 1993.

[14] A. IARROBINO, "Inverse system of a symbolic power. II. the Waring problem for forms," *J. Algebra*, vol. 174, pp. 1091–1110, 1995.

[15] J. M. F. ten BERGE, "The typical rank of tall three-way arrays," *Psychometrika*, vol. 65, no. 5, pp. 525–532, Sept. 2000.

[16] P. COMON, J. M. F. ten BERGE, L. De LATHAUWER, and J. CASTAING, "Generic and typical ranks of multi-way arrays," *Linear Algebra Appl.*, vol. 430, no. 11–12, pp. 2997–3007, June 2009.

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

[18] P. COMON, X. LUCIANI, and A. L. F. de ALMEIDA, "Tensor decompositions, alternating least squares and other tales," *Jour. Chemometrics*, vol.23, pp.393–405, Aug. 2009

[19] N. D. SIDIROPOULOS and R. BRO, "On the uniqueness of multilinear decomposition of N-way arrays," *Jour. Chemo.*, vol. 14, pp. 229–239, 2000.

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

[21] L. De LATHAUWER, "A link between canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM Jour. Matrix Ana. Appl.*, vol. 28, no. 3, pp. 642–666, 2006.

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

[23] I. A. KOGAN and M. M. MAZA, "Computation of canonical forms for ternary cubics," in *ISSAC*. 2000, pp. 151–160, ACM.

[24] A. STEGEMAN and P. COMON, "Subtracting a best rank-1 approximation does not necessarily decrease tensor rank," in *EUSIPCO'09*, Glasgow, Scotland, Aug. 24-28 2009.

[25] V. De SILVA and L-H. LIM, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM Jour. Matrix Ana. Appl.*, vol. 30, no. 3, pp. 1084–1127, 2008.

[26] J. HARRIS, *Algebraic Geometry, a first course*, vol. 133 of *Graduate Texts in Math.*, Springer, 1992.

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

[28] L. de LATHAUWER, 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] E. KOFIDIS and P. REGALIA, "One the best rank-1 approximation of higher-order supersymmetric tensors," *SIAM Jour. Matrix Ana. Appl.*, vol. 23, pp. 863–884, 2002.

[30] L. de LATHAUWER, B. de MOOR, and J. VANDEWALLE, "On the best rank-1 and rank-(R1,R2,...RN) approximation of high-order tensors," *SIAM Jour. Matrix Ana. Appl.*, vol. 21, no. 4, pp. 1324–1342, Apr. 2000.

[31] L-H. LIM and P. COMON, "Nonnegative approximations of nonnegative tensors," *Jour. Chemometrics*, vol.23, pp.432–441, Aug. 2009

[32] P. PAATERO, "Construction and analysis of degenerate Parafac models," *Jour. Chemometrics*, vol. 14, pp. 285–299, 2000.

[33] D. BINI, G. LOTTI, and F. ROMANI, "Approximate solutions for the bilinear form computational problem," *SIAM Jour. Comput.*, pp. 692–697, 1980.

[34] P. PAATERO, "The multilinear engine: A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model," *Journal of Computational and Graphical Statistics*, vol. 8, no. 4, pp. 854–888, Dec. 1999.

[35] G. TOMASI and R. BRO, "A comparison of algorithms for fitting the Parafac model," *Comp. Stat. Data Anal.*, vol. 50, pp. 1700–1734, 2006.

[36] R. BRO and C. A. ANDERSSON, "Improving the speed of multiway algorithms. Part II: Compression," *Chemometrics and Intelligent Laboratory Systems*, vol. 42, no. 1-2, pp. 105–113, 1998.

[37] P. KROONENBERG, *Applied Multiway Data Analysis*, Wiley, 2008.

[38] A. SMILDE, R. BRO, and P. GELADI, *Multi-Way Analysis*, Wiley, 2004.

[39] M. RAJIH, P. COMON, and R. HARSHMAN, "Enhanced line search : A novel method to accelerate Parafac," *SIAM Jour. Matrix Ana. Appl.*, vol. 30, no. 3, pp. 1148–1171, Sept. 2008.

[40] D. NION and L. DeLATHAUWER, "An enhanced line search scheme for complex-valued tensor decompositions. application in ds-cdma," *Signal Processing*, vol. 88, no. 3, pp. 749–755, Mar. 2008.

[41] A. STEGEMAN and L. De LATHAUWER, "A method to avoid diverging components in the Candecomp/Parafac model for generic IxJx2 arrays," *SIAM J. Matrix Ana. Appl.*, vol. 30, pp. 1614–1638, 2009.

[42] L. De LATHAUWER, B. De MOOR, and J. VANDEWALLE, "Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition," *SIAM Jour. Matrix Ana. Appl.*, vol. 26, no. 2, pp. 295–327, 2004.

[43] W. HACKBUSCH, B. N. KHOROMSKIJ, and E. E. TYRTYSHNIKOV, "Hierarchical Kronecker tensor-product approximation," *Jour. Numer. Math.*, vol. 13, pp. 119–156, 2005.

[44] V. STRASSEN, "Rank and optimal computation of generic tensors," *Linear Algebra Appl.*, vol. 52, pp. 645–685, July 1983.

[45] P. BÜRGISSER, M. CLAUSEN, and M. A. SHOKROLLAHI, *Algebraic Complexity Theory*, vol. 315, Springer, 1997.

[46] A. IARROBINO and V. KANEV, *Power sums, Gorenstein algebras and determinantal loci*, Springer, 1999.

[47] B. W. BADER and T. KOLDA, "Algorithm 862: Matlab tensor classes for fast algorithm prototyping," *ACM Trans. Math. Soft.*, vol. 32, pp. 635–653, 2006.

[48] M. MORUP and L. K. HANSEN et alterae, "Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG," *NeuroImage*, vol. 29, pp. 938–947, 2006.

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

[50] A. CICHOCKI, R. ZDUNEK, and S-I. AMARI, "Nonnegative matrix and tensor factorization," *IEEE Sig. Proc. Mag.*, pp. 142–145, 2008.

[51] M. WELLING and M. WEBER, "Positive tensor factorization," *Pattern Recognition Letters*, vol. 22, no. 12, pp. 1255–1261, 2001.

[52] M. PLUMBLEY, "Conditions for nonnegative independent component analysis," *Sig. Proc. Letters*, vol. 9, pp. 177–180, June 2002.

[53] D. D. LEE and H. S. SEUNG, "Learning the parts of objects by non-negative matrix factorization.," *Nature*, vol. 401, pp. 788–791, Oct. 1999.

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

[55] N. D. SIDIROPOULOS, G. B. GIANNAKIS, and R. BRO, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. on Sig. Proc.*, vol. 48, no. 3, pp. 810–823, Mar. 2000.

[56] P. COMON and M. RAJIH, "Blind identification of under-determined mixtures based on the characteristic function," *Signal Processing*, vol. 86, no. 9, pp. 2271–2281, Sept. 2006.