Pascal—Lab. I3S R-PC-2004-06

Canonical Tensor Decompositions

Presented at *ARCC Workshop on Tensor Decompositions*,
American Institute of Mathematics, Palo Alto, CA, July 18–24 2004

Pierre Comon, Laboratoire I3S, Sophia Antipolis, France

Université Nice SOPHIA ANTIPOLIS

CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Canonical Tensor Decompositions

P. Comon

Lab. I3S, 2000 route des Lucioles, B.P.121,
F-06903 Sophia-Antipolis cedex, France

I3S Report **RR-2004-17**
PASCAL Report **R-PC-2004-02**

June 17, 2004; revised July 20, 2004

Abstract

The Singular Value Decomposition (SVD) may be extended to tensors at least in two very different ways. One is the High-Order SVD (HOSVD), and the other is the Canonical Decomposition (CanD). Only the latter is closely related to the *tensor rank*. Important basic questions are raised in this short paper, such as the maximal achievable rank of a tensor of given dimensions, or the computation of a CanD. Some questions are answered, and it turns out that the answers depend on the choice of the underlying field, and on tensor symmetry structure, which outlines a major difference compared to matrices.

Key words : Tensor, Rank, Canonical Decomposition, Factor Analysis, Blind Identification, Independent Component Analysis, Under-Determined Systems, Alternate Least Squares.

Acknowledgment : This work has been supported in part by the American Institute of Mathematics (AIM), Palo Alto, CA, and by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors' views.

1 Introduction

This paper aims at serving as a basis of discussion at the tensor workshop to be held at AIM, Palo Alto, in July 2004. It should not be considered as a polished paper, but rather as a first draft, containing more questions than answers.

The key tool that we would like to have at our disposal is the analogue of the SVD, but for tensors of order higher than 2. There are actually two possible directions to extend this tool.

Since tensors can have a rank larger than their dimensions, the first idea is to search for rectangular full rank matrices instead of unitary matrices. This idea can be traced back to the seventies, and is often referred to as "Canonical Decompositions" (CanD, or CanDecomp) [37] [7] [21]. It defines clearly the *tensor rank*, through a non unitary congruent diagonalization. Unfortunately, we have few theoretical results on *tensor ranks*, and even fewer efficient *practical algorithms*.

The second idea consists of keeping the unitary constraint in the change of coordinates, but of replacing the diagonal structure of the core tensor by an all-orthogonal structure [36] [16]. This decomposition, now often referred to as the HOSVD, is computable via an Alternate Least Squares (ALS) algorithm. However, it is not related to tensor rank, but to the so-called $n-$mode ranks [17].

In the present discussion, we shall be interested mainly in the tensor rank (subsequently defined) induced by the CanD. Then basic questions can be raised, such as: what are the *generic* and the *maximal* tensor ranks, as a function of the order and dimensions? Is the rank of a given real tensor the same if understood in the real and the complex fields? Does Symmetry bring any restriction on the results? Some papers in the literature seem to already give different answers to these questions. It will be explained in this paper why there is actually no contradiction.

As already pointed out, one difficulty lies in the fact that quite few numerical algorithms have been proposed to compute the CanD of a tensor. Obviously, the solution is not unique in a number of cases (in particular even orders, including the case of matrices); so what kind of minimal constraints may one impose in order to restore uniqueness? The so-called ParaFac algorithms impose a strong constraint on the rank itself, far below its maximal value.

Another problem, only tackled in the present paper, is to approximate a tensor by another of lower (given) tensor rank. Contrary to the HOSVD, the CanD cannot be truncated to yield a LS approximation, whereas the SVD can (a brief analysis of a Courant-Fisher characterization has been given in [10]).

There has been sometimes in the literature (including this author's work) too fast or unjustified assimilations of ranks in $\mathbb{R}$ and $\mathbb{C}$; ranks of free (asymmetric) tensors and symmetric tensors have also been sometimes undistinguished. This may be understood because the distinction is irrelevant for matrices. However, this author's view is that this was a mistake. Indeed, these ranks might be all

different for some values of order and dimensions. This paper is hopefully a first step towards the definition of a more rigorous framework.

2 Notation

2.1 Arrays

Arrays will more than one index will be denoted in bold uppercase; vectors are one-way arrays, and will be denoted in bold lowercase. Plain uppercases will be mainly used to denote dimensions. For our purpose, only a few notations related to arrays [16] [11] are necessary. In this paper, the outer product of two arrays of order M and N is denoted $\boldsymbol{C} = \boldsymbol{A} \circ \boldsymbol{B}$ and is an array of order $M+N$:

$$C_{ij..\ell\, ab..d} = A_{ij..\ell}\, B_{ab..d} \tag{1}$$

For instance, the outer product of two vectors, $\boldsymbol{u} \circ \boldsymbol{v}$, is a matrix. Conversely, the mode$-p$ inner product between two arrays having the same pth dimension is denoted $\boldsymbol{A} \bullet_p \boldsymbol{B}$, and is obtained by summing over the pth index. More precisely, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are of orders M and N respectively, this yields for $p = 1$:

$$\{\boldsymbol{A} \underset{1}{\bullet} \boldsymbol{B}\}_{i_2..i_M, j_2..j_N} = \sum_k A_{k i_2..i_M}\, B_{k j_2..j_N}$$

For instance, the standard matrix-vector product can be written as $\boldsymbol{A}\,\boldsymbol{u} = \boldsymbol{A}^{\mathrm{T}} \bullet_1 \boldsymbol{u}$. Note that some authors [36] [16] denote this contraction product as $\boldsymbol{A} \times_p \boldsymbol{B}$, which we find much less convenient.

We shall say that a $d-$way array is *square* if all its d dimensions are identical. A square array will be called *symmetric* if its entries do not change by any permutation of its d indices. The linear space of square $d-$way arrays of size N is of dimension N^d, whereas the space of symmetric $d-$way arrays of same size is of dimension $\binom{N+d-1}{d}$.

In this framework, only $d-$way arrays that enjoy the *multilinearity property* by linear change of coordinates will be considered; they will be referred to as *tensors* [23]. To illustrate this property, let $\boldsymbol{T}$ be a tensor of third order of dimensions $P_1 \times P_2 \times P_3$, and let $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ be three matrices of size $K_1 \times P_1$, $K_2 \times P_2$, and $K_3 \times P_3$, respectively (in general $K_i = P_i$ and matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are invertible, but this is actually not mandatory in most of our discussion). Then tensor $\boldsymbol{T}$ is transformed by the multi-linear map $\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}\}$ into a tensor $\boldsymbol{T}'$ defined by:

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

which may be written in compact form as $\boldsymbol{T}' = \boldsymbol{T} \bullet_1 \boldsymbol{A} \bullet_2 \boldsymbol{B} \bullet_3 \boldsymbol{C}$.

2.2 Polynomials

Any symmetric tensor of order d and dimension K can be associated with a unique homogeneous polynomial of degree d in K variables via the expression

[11]:

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{j}} T_{\boldsymbol{j}}\ \boldsymbol{x}^{\boldsymbol{f}(\boldsymbol{j})} \tag{3}$$

where for any integer vector $\boldsymbol{j}$ of dimension d, one associates bijectively the integer vector $\boldsymbol{f}(\boldsymbol{j})$ of dimension K, each entry f_k of $\boldsymbol{f}(\boldsymbol{j})$ being equal to the number of times index k appears in $\boldsymbol{j}$. We have in particular $|\boldsymbol{f}(\boldsymbol{j}))| = d$. We also assume the following conventions (rather standard in algebraic geometry): $\boldsymbol{x}^{\boldsymbol{j}} \stackrel{\text{def}}{=} \prod_{k=1}^{K} x_k^{j_k}$ and $|\boldsymbol{f}| \stackrel{\text{def}}{=} \sum_{k=1}^{K} f_k$, where $\boldsymbol{j}$ and $\boldsymbol{f}$ are integer vectors. The converse is true as well, and the correspondence between tensors and homogeneous polynomials is obviously bijective.

Now for asymmetric tensors, the same association is not possible. In order to connect tensor spaces with algebraic geometry, tensors are associated with multilinear maps [35]. This justifies the use of the Zariski topology.

3 Ranks

3.1 Definitions

A tensor of order d and dimensions $(K_1, \ldots, K_d)$ can always be decomposed into elementary vector outer products as:

$$\boldsymbol{T} = \sum_{i=1}^{r} \boldsymbol{u}(i) \circ \boldsymbol{v}(i) \circ \ldots \boldsymbol{w}(i) \tag{4}$$

The smallest possible value of r for which this decomposition is possible is called the *tensor rank*, and denoted $r(\boldsymbol{T})$. The decomposition obtained this way is referred to as the Canonical Decomposition (CanD) of $\boldsymbol{T}$.

At this stage, we can make several observations. This definition does not need tensors to be defined in a field, and a ring is sufficient. Next, the CanD of a multilinear transform of $\boldsymbol{T}$, CanD$\{\mathcal{L}(\boldsymbol{T})\}$, is obtained as the multilinear transform of the CanD, $\mathcal{L}($CanD$\{\boldsymbol{T}\})$.

Given a linear space of tensors, $\mathcal{T}$, of given order and dimensions, several questions are relevant: (i) what is the maximal achievable tensor rank in $\mathcal{T}$, and (ii) what is the *generic* tensor rank in $\mathcal{T}$? Genericity refers to properties that are *almost always* true, which depends on the topology assumed (the Zariski topology in our case). The answers to these two questions turn out to depend on the field in which the problem is immersed (basically the real or complex fields, $\mathbb{R}$ and $\mathbb{C}$), and the symmetry constraints (symmetric tensors or tensors with free entries).

3.2 Generic and maximal ranks

Define the set of tensors $\mathcal{Y}_r = \{\boldsymbol{T} \in \mathcal{T} : r(\boldsymbol{T}) \leq r\}$ with values in $\mathbb{C}$. Also denote $\bar{\mathcal{Y}}_r$ its closure, and $\mathcal{Z}_r = \{\boldsymbol{T} \in \mathcal{T} : r(\boldsymbol{T}) = r\}$. Then we obviously have $\forall r$, $\mathcal{Y}_{r-1} \subseteq \mathcal{Y}_r$.

One can show the following [35]: (i) $\mathcal{Z}_r$ is dense in $\bar{\mathcal{Y}}_r$ as long as $r < \bar{R}$; In other words, if $\bar{\mathcal{Y}}_{r-1} \subset_{\neq} \bar{\mathcal{Y}}_r$, then $\bar{\mathcal{Z}}_r = \bar{\mathcal{Y}}_r$. (ii) if $\bar{\mathcal{Y}}_{r-1} = \bar{\mathcal{Y}}_r$, then $r \geq \bar{R}$.

Next, even if we know that $\mathcal{Y}_1$ is closed as a determinantal variety, $\mathcal{Y}_r$ are generally not closed for $r > 1$. This is another major difference with matrices, for which all $\mathcal{Y}_r$ are closed. One can actually prove a more accurate statement, from known results borrowed from algebraic geometry [14]. The varieties $\bar{\mathcal{Z}}_r$ can be ordered by inclusion as follows:

$$\text{if } r_1 < r_2 < \bar{R} < r_3 \leq R, \text{ then } \bar{\mathcal{Z}}_{r_1} \subset \bar{\mathcal{Z}}_{r_2} \subset \bar{\mathcal{Z}}_{\bar{R}} \supset \bar{\mathcal{Z}}_{r_3} \supseteq \bar{\mathcal{Z}}_R \tag{5}$$

This will be proved in a forthcoming paper. I also conjecture that $R = \bar{R} + 1$.

These statements extend previous results [3], and prove that there can be only *one* subset of non empty interior, and the latter is dense in $\mathcal{T}$; this result needs however an algebraically closed field (e.g. the field of complex numbers), and is thus not valid in the real field. The conjecture of Kruskal [29] according to which there could be several generic ranks for given order and dimensions may consequently hold true for real CanD of real tensors.

Now what we just described holds also true for symmetric tensors of $\mathcal{T}_s$ immersed in the complex field.

As an example, it has been shown [13] [25] that symmetric tensors of order 3 and dimension 3 have a generic rank $R_s = 4$ but a maximal rank of $R = 5$. This means that only $\mathcal{Z}_4$ is dense in $\bar{\mathcal{Y}}_4 = \bar{\mathcal{Y}}_5$, and $\mathcal{Z}_3$ and $\mathcal{Z}_5$ are not closed and of empty interior. On the other hand, $\mathcal{Z}_1$ is closed.

3.3 Asymmetric tensors

Let's start with tensors without symmetry structure. Then the generic rank $\bar{R}$ in $\mathcal{T}$ can be lower-bounded [35]. For tensors of order d and dimensions $(K_1, ..K_d)$, this bound can be written as:

$$\bar{R} \geq \left\lceil \frac{\prod_{i=1}^{d} K_i}{1 + \sum_{i=1}^{d} (K_i - 1)} \right\rceil \tag{6}$$

which reduces to the more readable value $K^d/(d\,K - d + 1)$ is the square case. It must be stressed that the generic rank $\bar{R}$ may of course be much larger than the smallest dimension. Next, the maximal achievable rank in $\mathcal{T}$, denoted R, can be itself larger than $\bar{R}$.

For instance in the case of complex square tensors order 3 and dimension $K \neq 3$, according to Lickteig [35], the bound would be reached:

$$\bar{R} = \left\lceil \frac{K^3}{3K - 2} \right\rceil \tag{7}$$

Unfortunately, for general values of order and dimensions, only bounds on generic and maximal ranks are known [3] [23] [32].

Now for real tensors, if the CanD is sought in $\mathbb{R}$, the rank can be found to be larger than the value found in $\mathbb{C}$, as pointed out in [29]. In other words, we

have actually the (rather natural) inequality, for any tensor $\boldsymbol{T}$:

$$r^{\mathbb{C}}(\boldsymbol{T}) \leq r^{\mathbb{R}}(\boldsymbol{T}) \tag{8}$$

It seems that not much attention has been paid to this up to now.

Example: In order to demonstrate that the equality does not always hold, define the square symmetric real tensor $\boldsymbol{T}$ of order 3 and dimension 2 as:

$$\boldsymbol{T}(:,:,1) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{T}(:,:,2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

If decomposed in $\mathbb{R}$, it is of rank 3:

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

whereas it admits a CanD of rank 2 in $\mathbb{C}$:

$$\boldsymbol{T} = \frac{\jmath}{2} \begin{pmatrix} -\jmath \\ 1 \end{pmatrix}^{\circ 3} - \frac{\jmath}{2} \begin{pmatrix} \jmath \\ 1 \end{pmatrix}^{\circ 3}$$

with $\jmath \stackrel{\text{def}}{=} \sqrt{-1}$. These decompositions can be obtained with the help of the algorithm described in [13], for instance. Alternatively, this tensor is associated with the homogeneous polynomial in two variables

$$p(x,y) = 3\,x\,y^2 - x^3,$$

which can be decomposed in $\mathbb{R}$ into

$$p(x,y) = \frac{1}{2}\,(x+y)^3 + \frac{1}{2}\,(x-y)^3 - 2\,x^3$$

Now, if this tensor is associated with a multilinear map (which is possible even for non symmetric tensors) as mentioned in section 2.2, then one would get the bilinear form:

$$b(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}) = \begin{pmatrix} y_1 y_2 - x_1 x_2 \\ x_1 y_2 + x_2 y_1 \end{pmatrix}$$

3.4 Symmetric tensors

Through the bijection (3), the CanD (4) of symmetric tensors can be transposed to homogeneous polynomials (also called *quantics*), as pointed out in [13]. This allows to talk indifferently either about CanD of tensors or quantics.

As for the asymmetric case, symmetric tensors may have only a unique generic rank $\bar{R}_s$, and this rank is not necessarily maximal [13]. A lower bound can be derived in a similar manner:

$$\bar{R}_s \geq \left\lceil \frac{\binom{K+d-1}{d}}{K} \right\rceil \tag{9}$$

An upper bound has also been derived for real [32] or complex [33] tensors:

$$\bar{R}_s \leq \binom{K+d-2}{d-1} \tag{10}$$

Because the space of symmetric tensors, $\mathcal{T}_s$, is included in the subspace of $\mathcal{T}$ of square tensors, maximal and generic ranks are related for every fixed order d and dimension K by:

$$\bar{R} \geq \bar{R}_s, \quad \text{and} \quad R \geq R_s \tag{11}$$

On the other hand, given a symmetric tensor $\boldsymbol{S}$, one can compute its CanD either in $\mathcal{T}_s$ or in $\mathcal{T}$. Since the CanD in $\mathcal{T}_s$ is constrained, we obviously also have the inequality between tensor ranks:

$$\forall \boldsymbol{S} \in \mathcal{T}_s,\ r(\boldsymbol{S}) \leq r_s(\boldsymbol{S}) \tag{12}$$

Note that (11) and (12) are in reverse order, but there is no contradiction: the spaces are not the same in (11). All this remains to be put in more formal words.

A question can then be raised: does there exist a symmetric tensor of "symmetric rank" r_s, but of "asymmetric rank" $r < r_s$? A numerical example remains to be found.

The value of the generic rank R_s has been reported in [13] for several values of order and dimension. There does not seem to be an explicit expression for it. For instance, as pointed out by Reznick [32], Clebsh proved that even when the numbers of free parameters are the same on both sides of the CanD, then the generic rank may not be equal to $\binom{K+d-1}{d}/K$; it is precisely the case for $(K, d) = (3, 4)$, which yields 15 degrees of freedom but $R_s = 6 \neq 5 = \binom{6}{4}/3$. This still holds true in the complex case [19].

On the other hand, the case of cubics ($d = 3$) is much better known. The classification is known since 1964, but a constructive algorithm to compute the symmetric CanD has been only recently proposed [25]. The case of binary quantics ($K = 2$) is much simpler, and is also completely known [37] [13] [30], and has already been utilized in real world problems [12].

3.5 Uniqueness

We have seen above that it is not easy to determine the rank $r(\boldsymbol{T})$ of a tensor $\boldsymbol{T}$. But even if its rank is given, is its CanD unique?

The answer is rather simple. Uniqueness is ensured if the number of parameters to determine is smaller than the number of free entries in $\boldsymbol{T}$. This means that the CanD is essentially unique (understand unique up to scale and permutation) if $r < \bar{R}$. On the other hand, uniqueness is not generally guaranteed if $r = \bar{R}$, and probably never when $r = R$. The dimensionality of the fiber of solutions has been given in [13] for generic symmetric tensors of order and dimension smaller than or equal to 8, and it turns out that the CanD is rarely unique for $r(\boldsymbol{T}) = \bar{R}_s$.

The so-called ParaFac algorithm [22] [28] [4] leads to an essentially unique solution because it assumes that $r(\boldsymbol{T}) \leq P(K,d)$, where $P(K,d)$ is an integer depending entirely on the numerical algorithm. Unfortunately, it turns out that $P(K,d) < \bar{R}$ for almost all values of (K,d). In other words, ParaFac is able to compute the CanD on a null-measure subset of $\mathcal{T}$, which may seem quite restrictive at first glance. Another ALS implementation is proposed in appendix 1.

4 Algorithms for the Canonical Decomposition

4.1 Asymmetric Tensors

As already pointed out earlier, the goal is to find matrices $\boldsymbol{A}^{(i)}$, $1 \leq i \leq d$, so as to minimize

$$\Psi = ||\boldsymbol{T} - \boldsymbol{A}^{(1)} \underset{2}{\bullet} \boldsymbol{A}^{(2)} \underset{2}{\bullet} \ldots \underset{2}{\bullet} \boldsymbol{A}^{(d)}||^2 \tag{13}$$

This LS criterion can also be rewritten in the form of a joint diagonalization of matrix slices. Assume $d = 3$ for simplicity; then we look for three matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ so as to minimize [31]:

$$\Psi = \sum_{k=1}^{K_3} ||\boldsymbol{T}(:,:,k) - \boldsymbol{A}\, Diag(\boldsymbol{C}(k,:))\, \boldsymbol{B}^{\mathrm{T}}||^2 \tag{14}$$

This obviously holds for orders $d > 3$. At order 4 for instance, one can either express the LS criterion as the joint diagonalization of $K_3 K_4$ matrices of size $K_1 \times K_2$:

$$\Psi = \sum_{k=1}^{K_3} \sum_{\ell=1}^{K_4} ||\boldsymbol{T}(:,:,k,\ell) - \boldsymbol{A}\, Diag(\boldsymbol{C}(k,:))\, Diag(\boldsymbol{D}(\ell,:))\, \boldsymbol{B}^{\mathrm{T}}||^2 \tag{15}$$

or as a joint diagonalization of K_4 third order tensors of size $K_1 \times K_2 \times K_3$:

$$\Psi = \sum_{\ell=1}^{K_4} ||\boldsymbol{T}(:,:,:,\ell) - \boldsymbol{A} \underset{2}{\bullet} \boldsymbol{B} \underset{2}{\bullet} \boldsymbol{C} \underset{2}{\bullet} Diag(\boldsymbol{D}(\ell,:))||^2 \tag{16}$$

For fixed matrices $\boldsymbol{B}$ and $\boldsymbol{C}$, criterion Ψ is quadratic in $\boldsymbol{A}$ and thus yields easily a closed-form LS solution for $\boldsymbol{A}$ (involving a pseudo-inversion). Hence, there is a quite evident Alternate LS (ALS) algorithm to compute the CanD. This is precisely what the ParaFac algorithm is doing. Implementations based on (13) have been independently proposed in 1970 by Carroll [8] and Harshman [21] [28] [4]. However, some limitations can be pointed out. First, the ALS algorithm can be trapped in local minima depending on the initialization, and may also converge very slowly. But more importantly, it has been noticed that there exist an upper bound on the rank of tensors that can be handled, to ensure uniqueness [34] [29]:

$$\sum_{p=1}^{d} K_p \geq 2\, r(\boldsymbol{T}) + d - 1 \tag{17}$$

As pointed out earlier, in the square case $K_p = K, \forall p$, this rank is likely to be below the generic rank $\bar{R}$ for almost all values of dimension K and order d. This may be interpreted as a strong limitation, if the algorithm is able to yield an exact CanD only on a set of tensors of null measure. On the other hand, it is able to yield an approximation of the CanD of given rank. And as we have seen, imposing a lower rank also allows to restore uniqueness.

As long as the algorithm has not converged, the joint diagonailization is only approximate. Moreover, it may happen that the LS error never reaches zero, which will either indicate that the chosen rank is smaller than the actual rank, or that the algorithm is stuck in a local minimum, or that the series of tensors diverges.

4.2 Symmetric tensors

In the symmetric case, the problem is not quadratic anymore, which increases considerably its complexity. However, there exists a reliable algorithm in the binary case ($K = 2$) [13], based on a theorem by Sylvester [37], which permits the computation of the CanD with a low computational complexity, for any odd order. For even orders, the solution is not unique, and some tricks have been imagined (namely to run jointly the CanD of two different but related tensors) in order to make the CanD usable in practice [12]. These techniques cannot be extended to higher orders.

For third order tensors, a number of specific results are available, and yield special purpose algorithms [25], which cannot be extended to orders $d \neq 3$.

For orders and dimensions higher than two, a solution consists of imposing only part of the symmetry, for instance to only for 2 of the d factors. Take example (14) at order $d = 3$. This means that we ignore that $\boldsymbol{C}$ is supposed to be equal to $\boldsymbol{A}$ in the criterion below:

$$\Psi = \sum_{k=1}^{K_3} ||\boldsymbol{T}(:,:,k) - \boldsymbol{A}\, Diag(\boldsymbol{C}(k,:))\, \boldsymbol{A}^{\mathrm{T}}||^2 \tag{18}$$

As shown in appendix 1, this minimization is feasible with an ALS algorithm. Of course the same can be done at order 4, based on (15), or at higher orders based on the same principle. The approach is suboptimal because symmetry is only partially imposed.

Another algorithm has been proposed in the complex case for tensors of even orders enjoying Hermitean symmetries. The so-called Biome algorithm attempts to exploit as many redundancies as possible, which are present in the original tensor [2] [1]. In that algorithm, not all symmetries are imposed however, and the highest achievable rank is $K^{d/2}$, which is probably still smaller than R_s.

4.3 Deflation

Using successive rank-1 approximates in order to compute all the rank-one tensors entering in the whole (exact) decomposition is not obvious, for it would

require a Courant-Fisher characterization that is not available [10] [11]. This is not what might have been understood from the early paper of Carroll and Chang in the seventies [8].

5 Algorithms for Orthogonal Decompositions

Orthogonal decompositions differ fundamentally from the CanD in the fact that matrices that we are looking for are unitary, and thus invertible. In the CanD, these matrices are rectangular, and do not define an invertible change of coordinate system.

One major advantage of this type of decomposition (here referred to as HOSVD) is that it allows to approximate a tensor by another of lower rank by merely truncating the full-size HOSVD.

Beside this Eckart-Young approximation ability, there is another advantage in searching for unitary decompositions. Indeed, unitary matrices can always be searched for in the form of products of Givens rotations [20], which reduces the number of unknowns to find in each iteration to a single angle (but may also increase the vulnerability to to local minima).

5.1 Asymmetric Tensors

As is well known, a tensor $\boldsymbol{T}$ is not always orthogonally diagonalizable [9] [15] [26]. That's why the HOSVD attempts to reduce the tensor into an all-orthogonal core tensor, and not to a diagonal one [16] [36] [8]. The HOSVD is an exact decomposition, as CanD is. In addition, as for the CanD, the decomposition is not always unique.

Originally, the so-called Tucker3 (older) algorithm devised by Tucker was dedicated to approximating a tensor by another of lower rank [27]. The case of rank-1 approximates is particular, and can be calculated with specific algorithms [24] [39].

Other types of tensor approximation are also possible, namely of lower $n-$mode ranks [17]. $n-$mode ranks are related to tensor rank only through inequalities; the set of $n-$mode ranks may also be sometimes called *dimensionality vector*, e.g. by Kruskal.

5.2 Symmetric tensors

A problem widely addressed in the Signal Processing community is the approximation of a square tensor of order d and dimension K by another of rank K [9]. As in the Jacobi-type approaches for matrices, the Frobenius norm does not change by unitary change of coordinates, and it is thus equivalent to maximize the diagonal terms of the resulting tensor in the new coordinate system.

To my knowledge, the Jacobi sweeping for tensors was first proposed by Comon in 1989 [9]. The solution in dimension 2 involves the computation of a single unknown, which can be done in several manners of various complexities;

see [9] [10] [11] [15] and references therein. We give in appendix 2 another less complex solution for fourth order complex tensors, when the diagonal tensor form is known to have real diagonal entries with the same sign.

There exist also algorithms based on slabs. The algorithm presented in [6] diagonalizes approximately a set of Hermitean matrices; the idea is to consider a tensor of even order $d = 2m$ as a linear mapping from the space of mth order tensors onto itself. As such, it can be represented by a Hermitean matrix, whose eigenvectors should also yield Hermitean mth order tensors, when unfolded [5]. In other words, the idea is again the same: one tries to impose as many symmetries as possible. This Joint Approximate Diagonalization (JAD) is thus not minimizing the original criterion, but only an approximation of it.

The STOTD algorithm [18] is dedicated to 3rd order tensors, but also works for higher orders. The fourth (or higher) order is solved in [18] by building 3rd order tensor slices. The optimal Givens rotation is obtained at each iteration by computing the rank-1 approximate of a 3×3 matrix. One can see this algorithm as a recursion on the slab order.

6 Concluding remarks

To conclude, I list below some problems that seem important to me, and which are still open, to my current knowledge, despite their fairly basic character:

- The generic rank may take different values in the four cases: symmetric/asymmetric and real/complex. It would be useful to build the four tables, at least in the square case, for orders $3 \leq d \leq 4$ and dimensions $3 \leq K \leq 8$. I don't know whether these tables have many different entries, but they definitely may have some.

- The same question can be raised for the maximal achievable rank, and is apparently much harder to respond.

- Find some simple examples for which these ranks are different; actually 8 examples are necessary if the (generic or maximal) ranks differ in the 4 cases.

- Check out if there exist a numerical algorithm to compute the CanD in each of the 4 situations, even for moderate orders and dimensions, e.g. $3 \leq (d, K) \leq 4$.

- Determine for every numerical algorithm the maximal rank it can handle. Then check out if this rank reaches the generic value or not.

- If an algorithm does not reach the generic rank, it means that it will generally compute only an approximate CanD. Check that this is useful in practice: my understanding is that even the dominant factors will be disturbed if the CanD is approximated. In what cases is the HOSVD more appropriate as far applications are concerned?

References

[1] L. ALBERA, P. COMON, P. CHEVALIER, A. FERREOL, "Blind identification of underdetermined mixtures based on the hexacovariance", in *ICASSP'04*, Montreal, May 17–21 2004, vol. II, pp. 29–32.

[2] L. ALBERA, A. FERREOL, P. COMON, P. CHEVALIER, "Blind identification of overcomplete mixtures of sources (BIOME)", *Lin. Algebra Appl.*, 2004, submitted in 2003, accepted March, 2004.

[3] M. D. ATKINSON, LLOYD S, "Bounds on the ranks of some 3-tensors", *Linear Algebra Appl.*, vol. 31, pp. 19–31, 1980.

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

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

[6] J. F. CARDOSO, A. SOULOUMIAC, "Jacobi angles for simultaneous diagonalization", *SIAM Jour. matrix Analysis*, vol. 17, no. 1, pp. 161–164, Jan. 1996.

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

[8] J. D. CARROLL, 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.

[9] 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.

[10] 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.

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

[12] P. COMON, "Blind identification and source separation in 2x3 underdetermined mixtures", *IEEE Trans. Signal Processing*, pp. 11–22, Jan. 2004, submitted in Sep. 2001, accepted in Apr. 2003.

[13] 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.

[14] D. COX, J. LITTLE, D. O'SHEA, *Using Algebraic Geometry*, Graduate Texts in Mathematics. Springer Verlag, New York, 1998.

[15] L. de LATHAUWER, *Signal Processing based on Multilinear Algebra*, Doctorate, Katholieke Universiteit Leuven, Sept. 1997.

[16] L. de LATHAUWER, B. de MOOR, J. VANDEWALLE, "A multilinear singular value decomposition", *SIAM Jour. Matrix Ana.*, vol. 21, no. 4, pp. 1253–1278, 2000.

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

[18] L. de LATHAUWER, B. de MOOR, J. VANDEWALLE, "Independent Component Analysis and (simultaneous) third-order tensor diagonalization", *IEEE Trans. Sig. Proc.*, pp. 2262–2271, Oct. 2001.

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

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

[21] R. A. HARSHMAN, "Foundations of the Parafac procedure: Models and conditions for an explanatory multimodal factor analysis", *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

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

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

[24] 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.

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

[26] T. G. KOLDA, "Orthogonal tensor decomposition", *SIAM Jour. Matrix Ana.*, vol. 23, no. 1, pp. 243–255, 2001.

[27] P. M. KROONENBERG, J. de LEEUW, "Principal component analysis of three-mode data", *Psychometrika*, vol. 45, pp. 69–97, 1980.

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

[29] J. B. KRUSKAL, "Rank, decomposition, and uniqueness for 3-way and n-way arrays", in *Multiway Data Analysis*, R. Coppi, S. Bolasco, Eds., pp. 7–18. Elsevier Science, North-Holland, 1989.

[30] J. KUNG, G. ROTA, "The invariant theory of binary forms", *Bulletin of the American Mathematical Association*, pp. 27–85, 1984.

[31] 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.

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

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

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

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

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

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

[38] A. YEREDOR, "Non-orthogonal joint diagoinalization in the LS sense with application in blind source separation", *IEEE Trans. Sig. Proc.*, vol. 50, no. 7, pp. 1545–1553, 2002.

[39] T. ZHANG, G. H. GOLUB, "Rank-one approximation to high-order tensors", *SIAM Jour. Matrix Ana.*, vol. 23, no. 2, pp. 534–550, 2001.

Appendix 1

Two practical algorithms are described in this section, that are able to compute the CanD and the symmetric CanD. The former can be seen to be an alternative description of the well-known Parafac algorithm, and the latter turns out to have strong similarities with a work of Yeredor [38], developed for diagonalizing a set of square matrices by an invertible transform, i.e. applicable only for $r_s(\boldsymbol{T}) \leq K$; however, the algorithm described here also works for square tensors $r_s(\boldsymbol{T}) \geq K$.

Both are Alternate Least Squares (ALS) algorithms, and attempt to approximate a given tensor by another of given rank. We have previously seen that if the rank is chosen large, the solution may not be unique, whereas if the rank is chosen smaller than that of the actual tensor, the approximation may not be possible because of the lack of closure of the subset of tensors of smaller rank. It is thus hopeless to run a large number of iterations in order to reach a stationary point; on the contrary, a reasonable number of iterations can still yield a usable result for practical purposes.

Asymmetric Canonical Decomposition

Given a set of possibly rectangular matrices $\boldsymbol{T}[k]$ (typically tensor slabs), the algorithm aims at minimizing

$$\Upsilon = \sum_k ||\boldsymbol{T}[k] - \boldsymbol{B}\,\boldsymbol{\Lambda}[k]\,\boldsymbol{C}^\dagger||^2 \tag{19}$$

with respect to matrices $\boldsymbol{B}$ and $\boldsymbol{C}$, where matrices $\boldsymbol{\Lambda}[k]$ are diagonal. This criterion can alternatively be written in the form of a distance between vectors as

$$\Upsilon = \sum_k ||\boldsymbol{t}[k] - \sum_n \boldsymbol{\lambda}_n[k]\,\boldsymbol{c}[n]^* \otimes \boldsymbol{b}[n]||^2$$

where $\boldsymbol{t}[k] = \mathbf{vec}(\boldsymbol{T}[k])$, $\boldsymbol{b} = \mathbf{vec}(\boldsymbol{B})$, and $\boldsymbol{c} = \mathbf{vec}(\boldsymbol{C})$. A more compact form is, with appropriate definition of matrix $\mathcal{M}$ (which depends on $\boldsymbol{B}$ and $\boldsymbol{C}$):

$$\Upsilon = \sum_k ||\boldsymbol{t}[k] - \mathcal{M}\,\boldsymbol{\lambda}[k]||^2 \tag{20}$$

Stationary values of $\boldsymbol{B}$ and $\boldsymbol{C}$ are given by

$$\boldsymbol{B} = \{\sum_k \boldsymbol{T}[k]\boldsymbol{C}\boldsymbol{\Lambda}[k]\}\{\sum_\ell \boldsymbol{\Lambda}[\ell]\boldsymbol{C}^\dagger\boldsymbol{C}\boldsymbol{\Lambda}[\ell]\}^{-1} \tag{21}$$

$$\boldsymbol{C} = \{\sum_k \boldsymbol{T}[k]^\dagger\boldsymbol{B}\boldsymbol{\Lambda}[k]\}\{\sum_\ell \boldsymbol{\Lambda}[\ell]\boldsymbol{B}^\dagger\boldsymbol{B}\boldsymbol{\Lambda}[\ell]\}^{-1} \tag{22}$$

whereas stationary values of the diagonal of $\boldsymbol{\Lambda}[k]$ are given by the vectors

$$\boldsymbol{\lambda}[k] = \{\mathcal{M}^\dagger\mathcal{M}\}^{-1}\mathcal{M}^\dagger\boldsymbol{t}[k] \tag{23}$$

The ALS algorithm consists of executing alternatively (23), (21), and (22). When matrices involved in a system solution are singular, a LS solution is computed.

Symmetric Canonical Decomposition

In the asymmetric case, things are more complicated because the optimization criterion is not quadratic anymore in the unknown rectangular matrix. Again, two writings are derived in order to obtain stationary values with respect to the rectangular matrix and to the diagonal one:

$$\Upsilon = \sum_k ||\boldsymbol{T}[k] - \boldsymbol{B}\boldsymbol{\Lambda}[k]\boldsymbol{B}^\dagger||^2 \tag{24}$$

and

$$\Upsilon = \sum_k ||\boldsymbol{t}[k] - \mathcal{B}\,\boldsymbol{\lambda}[k]||^2 \tag{25}$$

Some manipulations would show that the stationary values $\boldsymbol{\lambda}[k]$ are given by

$$\boldsymbol{\lambda}[k] = \{\mathcal{B}^\dagger\,\mathcal{B}\}^{-1}\mathcal{B}^\dagger\,\boldsymbol{t}[k] \tag{26}$$

Last, the stationary value of each column $\boldsymbol{b}[\ell]$ of matrix $\boldsymbol{B}$ is the dominant eigenvector of the Hermitian matrix

$$\boldsymbol{P}[\ell] = \frac{1}{2}\sum_k \lambda_\ell[k]\{\tilde{\boldsymbol{T}}[k;\ell]^\dagger + \tilde{\boldsymbol{T}}[k;\ell]\} \tag{27}$$

where $\tilde{\boldsymbol{T}}[k;\ell] \stackrel{\text{def}}{=} \boldsymbol{T}[k] - \sum_{n\neq\ell}\lambda_n[k]\boldsymbol{b}[n]\boldsymbol{b}[n]^\dagger$.

The ALS algorithm consists of executing successively (26) and the calculation of the dominant eigenvector of the K matrices (27). As before, a LS solution is computed when matrices involved are singular.

Note that even when tensor $\boldsymbol{T}$ is square and symmetric, it is still possible to compute the asymmetric CanD. If the solution obtained is such that $\boldsymbol{B} = \boldsymbol{C}$ and $\boldsymbol{\Lambda}[k] = \text{Diag}(\boldsymbol{b}[k])$, then it yields a symmetric CanD.