Decomposing tensors with structured matrix factors reduces to rank-1 approximations

Pierre Comon, Mikael Sørensen and Elias Tsigaridas
Lab. I3S, University of Nice, CNRS
Sophia-Antipolis, France

September 8, 2009

I3S report no. I3S/RR-2009-11

Abstract

Tensor decompositions are now known to permit to estimate in a deterministic way the parameters in a multi-linear model. Applications have been already pointed out in antenna array processing and digital communications, among others, and are extremely attractive provided some diversity at the receiver is available. Non iterative algorithms are proposed in this paper to compute the required tensor decomposition into a sum of rank-1 terms when some factor matrices enjoy some structure, such as block-Hankel, triangular, band, etc. The only condition is that the number of parameters characterizing the structure of a matrix should be significantly smaller than the number of its entries.

Acknowledgment

Work partially supported by contract ANR-06-BLAN-0074 "Decotes".

Contents

1 Motivation

This contribution is motivated by the fact that there exist very few efficient numerical algorithms for decomposing a tensor array into a sum of rank-1 terms. One can just mention the case of symmetric tensors of any order but of dimension 2, which can be decomposed with the help of a Sylvester's theorem [?], or the case of third order tensors having one dimension equal to 2, which can be handled by computing eigenvectors of a matrix pencil [?]. Even if the case of symmetric complex tensors has been partially solved in [?], the computational complexity is still significant, since a polynomial system of degree 2 often needs to be solved.

Yet, practical problems are encountered where the factor matrices have a structure, such as Toeplitz, which decreases the number of unknowns to be computed. We show in this paper that under certain conditions, the full decomposition can be computed almost surely within a finite number of operations (assuming that a matrix SVD can).

The Toeplitz structure has been already exploited in several contributions, e.g. [?], to speed up the ALS algorithm. But the algorithm is still iterative with unproved convergence. Only recently, some authors have attempted to build a non-iterative algorithm [?]; however, the latter works in three stages, and can only be applied for a single structured factor, which must be simultaneously Toeplitz lower triangular and banded, which is rather restrictive. It is suboptimal in the sense that the structure in incompletely exploited, so that it has to be recovered by projection in a third stage.

Tensor decompositions are very attractive in the fields of antenna array processing [?] and digital communications [?], when diversity is available at the receiver. But many other application areas exist [?]. Factor matrices appearing in the tensor decomposition can be structured [?] [?], and can have the very particular structure of banded triangular Toeplitz if Blind Identification of a SISO FIR channel is considered [?]. A contrario, the algorithms developed in the present paper exploit a structure that can be much less particular, since it is characterized by any linear space of reduced dimension. For instance, only one of the previous features is necessary, *e.g.* Toeplitz, or triangular, or banded, but not the three of them. We refer to this decomposition as "Structured Canonical Decomposition" (SCandD).

2 Notation

In order to ease the reading, array symbols are denoted with different fonts, depending on the number of indices. Plain font denotes scalar numbers, *e.g.* L, a_i or A_{ij}, boldface lowercases denote vectors, *e.g.* $\mathbf{x}$, or $\boldsymbol{\alpha}$, boldface uppercases denote matrices, *e.g.* $\mathbf{A}$, or $\mathbf{S}(\ell)$, and tensor arrays of order higher than 2 are represented by bold italic letters, *e.g.* $\boldsymbol{T}$, or $\boldsymbol{I}$. In the remainder, $\boldsymbol{I}$ will always denote the tensor array having ones on its diagonal, and zeros everywhere else.

Tensors are objects defining maps from a product of linear spaces to another. Once the bases of theses spaces are fixed, they are represented by arrays of coordinates. A tensor of order d is represented by an array with d indices. For simplicity, tensors are often (somewhat abusively) assimilated with their array representation, as done in the present paper.

Tensor arrays are modified in a multi-linear manner when bases are changed linearly. To make it simple, let $\boldsymbol{T}$ be a 3rd order tensor, and let $\mathbf{A}$ (resp. $\mathbf{B}$ and $\mathbf{C}$) be linear transforms acting in the first (resp. 2nd and 3rd) linear space. Then the new array representing the tensor can be written as

$$T'_{ijk} = \sum_{\ell mn} A_{i\ell} B_{jm} C_{kn}\, T_{\ell mn}$$

which can be conveniently written in a more compact form:

$$\boldsymbol{T}' = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \boldsymbol{T}$$

This way of denoting a multi-linear transformation is more and more used in the scientific community.

Given two matrices $\mathbf{A}$ and $\mathbf{B}$, one defines the Kronecker product:

$$\mathbf{A} \otimes \mathbf{B} \stackrel{\text{def}}{=} \begin{pmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \cdots \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & \cdots \\ \vdots & \vdots & \end{pmatrix},$$

If the latter matrices have the same number of columns, one also defines the column-wise Kronecker product, often referred to as the Khatri-Rao product:

$$\mathbf{A} \odot \mathbf{B} \stackrel{\text{def}}{=} \left[\ \mathbf{a}(1) \otimes \mathbf{b}(1) \quad \mathbf{a}(2) \otimes \mathbf{b}(2) \quad \cdots\ \right].$$

Yet another ingredient we shall need is the operation allowing to store a matrix in vector form, $\mathbf{x} \stackrel{\text{def}}{=} \mathbf{vec}\{\mathbf{X}\}$, and the inverse operation[1], $\mathbf{X} = \mathbf{Unvec}\{\mathbf{x}\}$. To fix the ideas, let $\mathbf{X}$ be a $I \times J$ matrix. We choose the $\mathbf{vec}\{\cdot\}$ map defined by $x_{(i-1)J+j} = X_{ij}$. With this definition, we have the property that $\mathbf{vec}\{\mathbf{x}\,\mathbf{y}^{\mathsf{T}}\} = \mathbf{x} \otimes \mathbf{y}$, for any pairs of vectors $\mathbf{x}$ and $\mathbf{y}$, or equivalently $\mathbf{Unvec}\{\mathbf{x} \otimes \mathbf{y}\} = \mathbf{x}\,\mathbf{y}^{\mathsf{T}}$.

Similarly, tensor arrays can be unfolded into the so-called "unfolding matrices", or "flattening matrices". In the case of order 3 tensors, there are 3 such matrices. Given a tensor $\boldsymbol{T}$ of dimensions $I \times J \times K$, represented by an arry T_{ijk}, the first unfolding matrix is defined as

$$\mathbf{T}^{(1)} = \begin{bmatrix} \mathbf{T}_{1::} \\ \vdots \\ \mathbf{T}_{i::} \\ \vdots \\ \mathbf{T}_{I::} \end{bmatrix}$$

where $\mathbf{T}_{i::}$ denotes the $J \times K$ matrix slice obtained by fixing the 1st index to i in the tensor array.

Finally, on the linear space of rectangular matrices, one defines the Hermitean scalar product $\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{trace}\{\mathbf{A}^{\mathsf{H}}\mathbf{B}\}$; the latter induces the Frobenius norm.

3 Structured tensor decomposition

We first define the minimal polyadic decomposition of a tensor, which will be referred to as Canonical Decomposition (CAND)[2]. The definition is given in the case of a 3rd order tensor, but it extends to any order in an obvious manner. It can be seen as a definition of the tensor rank.

Definition 1 *Let $\boldsymbol{T}$ be a tensor of order 3 and rank R. Then $\boldsymbol{T}$ can be written as a multilinear transform of the diagonal tensor $\boldsymbol{I}$:*

$$\boldsymbol{T} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \boldsymbol{I}$$

[1]In order for the operator $\mathbf{Unvec}\{\cdot\}$ to be unambiguously defined, the number of columns of the output should be given as an input argument. However, for the sake of simplicity, it shall be omitted since it will be always clear from the context.

[2]Note that this decomposition has been named "Parafac" in some communities, *i.e.* Psychometry.

where the factor matrices $\mathbf{A}$*,* $\mathbf{B}$ *and* $\mathbf{C}$ *have* R *columns.*

Note that the tensor rank and the CanD may be defined independently of any bases, by using the tensor product '$\otimes$':

$$\boldsymbol{T} = \sum_{p=1}^{R} \mathbf{a}(p) \otimes \mathbf{b}(p) \otimes \mathbf{c}(p)$$

but this abstract definition will not be used herein.

In Definition **??**, it is known that factor matrices are not defined in a unique way. In fact, each of them can be post-multiplied by a permutation $\mathbf{\Pi}$ and an invertible diagonal matrix, so that $\boldsymbol{T} = (\mathbf{A\Pi\Lambda}_A, \mathbf{B\Pi\Lambda}_B, \mathbf{C\Pi}\Lambda_C) \cdot \boldsymbol{I}$, provided $\mathbf{\Lambda}_A\mathbf{\Lambda}_B\mathbf{\Lambda}_C = \mathbf{I}$. The lemma below will be useful to choose the permutation and scaling matrices, whenever this indetermination is still present.

Lemma 2 *If a matrix* $\mathbf{N}$ *is invertible, then there exist a permutation* $\mathbf{\Pi}$ *and a diagonal invertible matrix* $\mathbf{\Lambda}$ *such that matrix* $\mathbf{N\Pi\Lambda}$ *has ones on its diagonal.*

It is worth noting that if factor matrices are structured, full scaling indeterminacies may disappear, and reduce to a mere scalar scale factor (this is what happens for Toeplitz factors, as addressed subsequently). This is a significant advantage of the SCanD over the CanD. We shall also need the well known results below, that we recall without proof.

Lemma 3 *Let* $\boldsymbol{T}$ *be a tensor, whose* CanD *is defined in Def.***??***. Then its first unfolding matrix can be written as*

$$\mathbf{T}^{(1)} = (\mathbf{A} \odot \mathbf{B})\mathbf{C}^{\mathsf{T}} \tag{1}$$

Lemma 4 *Denote* $\mathbf{T}^{(1)} = \mathbf{U\Sigma V}^{\mathsf{H}}$ *the SVD of* $\mathbf{T}^{(1)}$*, where* $\mathbf{\Sigma}$ *is* $R \times R$*, and* $\mathbf{V}^{\mathsf{H}}\mathbf{V} = \mathbf{I}_R$*. Then, if* $\mathbf{A} \odot \mathbf{B}$ *and* $\mathbf{C}$ *are full rank in (***??***), there always exist* $R \times R$ *invertible matrices* $\mathbf{M}$ *and* $\mathbf{N}$ *such that*

$$\mathbf{U\Sigma M}^{-1} = \mathbf{A} \odot \mathbf{B} \quad \textit{and} \quad \mathbf{M}\,\mathbf{V}^{\mathsf{H}} = \mathbf{C}^{\mathsf{T}} \tag{2}$$

$$\mathbf{UN} = \mathbf{A} \odot \mathbf{B} \quad \textit{and} \quad \mathbf{N}^{-1}\mathbf{\Sigma V}^{\mathsf{H}} = \mathbf{C}^{\mathsf{T}} \tag{3}$$

If one or several factor matrices appearing in the CanD are imposed to be structured (cf. definition below), we say that we are dealing with a "structured CanD" (SCanD). We shall be concerned by the classes of structured matrices that form linear spaces. Let $\{\mathbf{S}(\ell), 1 \leq \ell \leq IR\}$ be an orthonormal basis of $I \times R$ matrices. We state the following:

Definition 5 *An $I \times R$ matrix $\mathbf{S}$ is said to be structured if there exists an orthonormal basis of matrices $\mathbf{S}(\ell)$ such that*

$$\mathbf{S} = \sum_{\ell=1}^{\omega(\mathbf{S})} \alpha(\ell)\, \mathbf{S}(\ell) \tag{4}$$

where $\omega(\mathbf{S}) < IR$ is given. Such linear spaces will be denoted $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ for factor matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, respectively.

For instance, strictly lower triangular matrices are structured in the above sense, as well as Toeplitz or Hankel matrices, skew-symmetric matrices, and certain band matrices. Assume only matrix $\mathbf{C}$ is structured in (**??**); then we have:

Lemma 6 *If the linear space $\mathcal{C}$ is stable by post-multiplication by invertible diagonal matrices, then matrix $\mathbf{M}$ defined in Lemma **??** can be imposed to have ones on its diagonal. Otherwise, one can always impose $M_{r1} = 1$, or $\gamma(r) = 1$ for some r.*

Proof. From equation (**??**), it is clear that pre-multiplication of $\mathbf{C}^{\mathsf{T}}$ by $\mathbf{\Lambda}\mathbf{\Pi}^{\mathsf{T}}$ implies pre-multiplication of $\mathbf{M}$ by the same factor. Then by Lemma **??**, one can choose scaling matrix $\mathbf{\Lambda}$ and permutation $\mathbf{\Pi}$ so that $\mathbf{Diag}\{\mathbf{M}\} = \mathbf{I}$. Now when $\mathbf{C}\mathbf{\Pi}\mathbf{\Lambda}$ does not always belong to $\mathcal{C}$, such a choice is not possible. But a scalar scale factor always subsists in (**??**), such that any non-zero entry of $\mathbf{M}$ can be set to 1, and there is at least one in the 1st column. Another choice consists of using this scale factor to impose a nonzero parameter in the right hand side of (**??**) to be equal to 1, e.g. $\gamma(1) = 1$. □

With a similar reasoning one can state the following

Lemma 7 *If both $\mathbf{A}$ and $\mathbf{B}$ are structured in (**??**), but not $\mathbf{C}$, one can always impose either $M_{11} = 1$, or $\alpha(1) = 1 = \beta(1)$.*

4 Non-iterative solutions

In this section, results allowing to deflate the SCanD to matrix SVD's will be stated.

Proposition 8 *Let $\boldsymbol{T}$ be a tensor of dimensions $I \times J \times K$ and rank $R > 1$, admitting the* SCANDD *below:*

$$\boldsymbol{T} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \boldsymbol{I}.$$

*where matrix $\mathbf{C}$ is structured (according to Def. **??**) with $\omega(\mathbf{C}) \leq K^2/4$. Then, the calculation of the three matrix factors may be achieved* generically *by solving a linear system followed by R matrix rank-1 approximations, provided the rank of $\boldsymbol{T}$ is not too large, namely*

$$R^2 - KR + \omega(\mathbf{C}) - 1 \leq 0 \tag{5}$$

Proof. Since $\mathbf{C}$ is structured, it can be written as $\mathbf{C} = \sum_{\ell=1}^{\omega(\mathbf{C})} \gamma(\ell)\, \mathbf{C}(\ell)$ where matrices $\mathbf{C}(\ell)$ are known. Since $\boldsymbol{T}$ is of rank $R > 1$, $\mathbf{C}$ is nonzero, and from Lemma **??**, we may set $\gamma(1) = 1$. Then from Lemma **??**, matrix $\mathbf{M}$ must satisfy:

$$\mathbf{M}\,\mathbf{V}^{\mathsf{H}} \;=\; \mathbf{C}(1) + \sum_{\ell=2}^{\omega(\mathbf{C})} \gamma(\ell)\, \mathbf{C}(\ell)^{\mathsf{T}}, \tag{6}$$

This linear system contains KR equations and $R^2 + \omega(\mathbf{C}) - 1$ unknowns, $\gamma(k)$ and M_{ij}. Since it has more equations than unknowns, according to our assumptions, it generically admits one solution. In fact, the condition $\omega(\mathbf{C}) \leq K^2/4$ ensures that inequality (**??**) admits a non empty set of solutions for R. We have thus obtained matrices $\mathbf{M}$, and $\mathbf{C}$.

On the other hand, we have from (**??**) that $\mathbf{F} \stackrel{\text{def}}{=} \mathbf{U}\boldsymbol{\Sigma}\mathbf{M}^{-1} = \mathbf{A} \odot \mathbf{B}$. The last operation remaining to perform is the calculation of matrices $\mathbf{A}$ and $\mathbf{B}$, which can be done in a standard way column by column. For doing this, one notices that the rth column of matrix $\mathbf{F}$, $\mathbf{f}(r)$, is ideally equal to $\mathbf{a}(r) \otimes \mathbf{b}(r)$, whose matrix unvectorization form is $\mathbf{a}(r)\mathbf{b}(r)^{\mathsf{T}}$. So estimates of columns $\mathbf{a}(r)$ and $\mathbf{b}(r)$ of $\mathbf{A}$ and $\mathbf{B}$ can indeed be obtained by computing the best rank-1 approximate of matrix $\mathbf{Unvec}\{\mathbf{f}(r)\}$. □

Proposition 9 *Let $\boldsymbol{T}$ be a tensor of dimensions $I \times J \times K$ and rank R, admitting the* SCANDD*: $\boldsymbol{T} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \boldsymbol{I}$, where matrices $\mathbf{A}$ and $\mathbf{B}$ are structured, possibly with different structures, with $\omega(\mathbf{A})\omega(\mathbf{B}) \leq I^2 J^2/4$. Then, the calculation of the three matrix factors may be achieved* generically *by solving* *a linear system followed by one matrix rank-1 approximation, provided the rank of* $\boldsymbol{T}$ *satisfies the necessary condition:*

$$R^2 - IJ\,R + \omega(\mathbf{A})\omega(\mathbf{B}) - 1 \leq 0. \tag{7}$$

Proof. Since matrices $\mathbf{A}$ and $\mathbf{B}$ are structured, we have

$$\mathbf{U}\,\mathbf{N} = \sum_{i=1}^{\omega(\mathbf{A})} \sum_{j=1}^{\omega(\mathbf{B})} \alpha(i)\,\beta(j)\,(\mathbf{A}(i) \odot \mathbf{B}(j)) \tag{8}$$

which contains IJR equations, using the notation of (**??**). System (**??**) can be seen as a linear system of IJR equations in the $\omega(\mathbf{A})\omega(\mathbf{B})$ unknowns $X(i,j) \stackrel{\text{def}}{=} \alpha_i\beta_j$. Next, from Lemma **??**, if we choose to impose $N_{11} = 1$, we also have $R^2 - 1$ unknowns N_{ij}. Note that if spaces $\mathcal{A}$ and $\mathcal{B}$ are stable by diagonal scaling, we impose instead $\mathbf{Diag}\{\mathbf{N}\} = \mathbf{I}$, and we have $R^2 - R$ remaining unknowns; but let's concentrate on the less favorable case $N_{11} = 1$.

Hence, linear system (**??**) contains IJR equations in $R^2 - 1 + \omega(\mathbf{A})\omega(\mathbf{B})$ unknowns. It generically suffices to determine matrices $\mathbf{N}$ and consequently $\mathbf{C}$, as well as matrix $\mathbf{X}$, because R is not too large, by hypothesis.

The last step consists of computing the best rank-1 approximate of matrix $\mathbf{X}$, $\boldsymbol{\alpha}\,\boldsymbol{\beta}^{\mathsf{T}}$, as in the proof of Proposition **??**, which will yield $\mathbf{A}$ and $\mathbf{B}$. □

Now if all three factor matrices are structured, one can show that their estimation can be carried out with the help of rank-1 approximates as now pointed out.

Proposition 10 *Let* $\boldsymbol{T}$ *be a tensor of order* 3*, with* 3 *structured matrix factors. Then under the same conditions as in Prop.* **??***, its* SCanD *can be computed by solving two overdetermined linear systems, and by computing* R *rank-1 matrix approximates.*

Proof. If the three matrix factors are structured, the system of equations to solve is the following:

$$\mathbf{U}\,\mathbf{N} \;=\; \sum_{i=1}^{\omega(\mathbf{A})} \sum_{j=1}^{\omega(\mathbf{B})} \alpha(i)\,\beta(j)\mathbf{A}(i) \odot \mathbf{B}(j) \tag{9}$$

$$\mathbf{\Sigma}\,\mathbf{V}^{\mathsf{H}} \;=\; \sum_{k=1}^{\omega(\mathbf{C})} \gamma(k)\,\mathbf{C}(k)\,\mathbf{N} \tag{10}$$

The first system (**??**) can be solved for $\mathbf{N}$, $\alpha(i)$ and $\beta(j)$ under the same conditions as in Proposition **??**, with the help of R matrix rank-1 approximates. Then the value of $\mathbf{N}$ can be plugged back into (**??**), which then becomes linear in the unknowns $\gamma(k)$, and can be solved in the LS sense. □

There is however yet another case where the CanD can be computed by solely resorting to matrix SVD's, as shown in the proposition below.

Proposition 11 *Let $\boldsymbol{T}$ be a tensor of order 4, with 2 structured matrix factors. Then under the same conditions as in Prop.* **??***, its* SCanD *can be computed by solving an overdetermined linear system, and by computing $R+1$ rank-1 matrix approximates.*

Proof. Consider a $I \times J \times K \times L$ tensor, and its $IJ \times KL$ unfolding matrix:

$$\mathbf{T}^{(2,2)} = (\mathbf{A} \odot \mathbf{B})(\mathbf{C} \odot \mathbf{D})^{\mathsf{T}}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ all have R columns. We assume that both $\mathbf{A}$ and $\mathbf{B}$ are structured, which means that we have

$$\mathbf{T}^{(2,2)} = \sum_{i=1}^{\omega(\mathbf{A})} \sum_{j=1}^{\omega(\mathbf{B})} \alpha_i \beta_j \, (\mathbf{A}(i) \odot \mathbf{B}(j))(\mathbf{C} \odot \mathbf{D})^{\mathsf{T}}$$

As in the proof of Proposition **??**, we consider the SVD of matrix $\mathbf{T}^{(2,2)} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}$. Hence there exists a $R \times R$ invertible matrix, $\mathbf{N}$, such that $\mathbf{U}\,\mathbf{N} = \mathbf{A} \odot \mathbf{B}$ and $\mathbf{N}^{-1}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}} = \mathbf{C} \odot \mathbf{D}$. Our linear system contains IJR equations and $R^2 - 1 + \omega(\mathbf{A})\omega(\mathbf{B})$ unknowns, N_{ij} and X_{pq}, if we set $N_{11} = 1$. Once this over-determined system has been solved in the Least Squares (LS) sense, matrices $\mathbf{N}$ and $\mathbf{X}$ are known. Next we obtain again α_i and β_j via a rank-one approximation of matrix $\mathbf{X}$.

It remains to solve $\mathbf{N}^{-1}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}} = (\mathbf{C} \odot \mathbf{D})^{\mathsf{T}}$. Following the same lines as in the proof of Proposition **??**, denote $\mathbf{F} = \mathbf{V}^{*}\boldsymbol{\Sigma}\mathbf{N}^{-\mathsf{T}}$ and $\mathbf{f}(r)$ its columns, $1 \le r \le R$, each of dimension KL. The rest of the proof is similar to that of Proposition **??**, viz, the columns $\mathbf{a}(r)$ and $\mathbf{b}(r)$ are obtained by computing the best rank-1 approximation of the $K \times L$ matrices $\mathbf{Unvec}\{\mathbf{f}(r)\}$, respectively. □

Remark 1 *An alternative approach would be to use the orthogonality equations*

$$\langle \mathbf{M}\,\mathbf{V}^{\mathsf{H}}, \mathbf{C}(n)^{\mathsf{T}} \rangle = 0, \quad \omega(\mathbf{C}) < n \le KR \tag{11}$$

$$\langle \mathbf{A}(p) \odot \mathbf{B}(q), \mathbf{UN} \rangle = 0 \tag{12}$$

either for $\{1 \leq p \leq IR$ *and* $\omega(\mathbf{B}) < j \leq JR\}$ *or for* $\{\omega(\mathbf{A}) < i \leq IR$ *and* $1 \leq j \leq JR\}$.

5 Solutions requiring Higher Order rank-1 approximates

From the proofs derived in the previous section, it is clear that our propositions can be extended to tensors having three structured matrices, or to tensors of order larger than 3. One such instance is given below.

Proposition 12 *Let* $\boldsymbol{T}$ *be a tensor of order* d *and dimensions* K_μ, $1 \leq \mu \leq d$, *with* m *structured matrix factors,* $m > 0$, *and* $d - m$ *unstructured. Then its* SCAND *can be computed by solving an overdetermined linear system, and by computing rank-1 approximations of one tensor of order* m *and* R *tensors of order* $d - m$, *provided the necessary conditions below are met:*

$$\prod_{\mu=1}^{m} \omega(\mathbf{A}^{(\mu)}) - \frac{1}{4} \prod K_\mu^2 \;\leq\; 0 \tag{13}$$

$$R^2 - (\prod_{\mu=1}^{m} K_\mu) R + \prod_{\mu=1}^{m} \omega(\mathbf{A}^{(\mu)}) - 1 \;\leq\; 0 \tag{14}$$

Proof. Tensor $\boldsymbol{T}$ is unfolded into a $K_1 \dots K_m \times K_{m+1} \dots K_d$ unfolding matrix, denoted $\mathbf{T}$, and denote $\mathbf{A}^{(\mu)}$ the μth factor matrix of size $K_\mu \times R$, each defined with $\omega(\mathbf{A}^{(\mu)})$ parameters. Write the SVD of $\mathbf{T}$ as $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}$, and consider the identity

$$\mathbf{U}\mathbf{N} = \mathbf{A}^{(1)} \odot \dots \odot \mathbf{A}^{(m)}$$

where $\mathbf{N}$ is an unknown $R \times R$ invertible matrix. Similarly to (**??**), there are $K_1 \dots K_m R^m$ equations, allowing to compute the $R^2 - 1 + \prod_{\mu=1}^{m} \omega(\mathbf{A}^{(\mu)})$ unknowns N_{ij} and $X_{i_1..i_m} \stackrel{\text{def}}{=} \alpha_{i_1}^{(1)} \dots \alpha_{i_m}^{(m)}$. Compute the coefficient vectors $\boldsymbol{\alpha}^{(i)}$, $1 \leq i \leq m$, via a rank-1 approximation of tensor $\boldsymbol{X}$, whose entries are $X_{i_1..i_m}$.

On the orther hand, as in the proof of Proposition **??**, use equation $\mathbf{V}^*\boldsymbol{\Sigma}\mathbf{N}^{-\mathsf{T}} = \mathbf{A}^{(m+1)} \odot \dots \odot \mathbf{A}^{(d)}$ column-wise, in order to obtain the columns of the remaining factor matrices. Each column is obtained with the help of a rank-1 approximation of a tensor of order $d - m$. □

Now, a better bound on rank R can be obtained by isolating two structured matrices on one side (in fact the two having the strongest structure).

Proposition 13 *Let* $\boldsymbol{T}$ *be a tensor of order* d *and dimensions* K_μ, $1 \leq \mu \leq d$, *with* m *structured matrix factors,* $m > 0$, *and* $d - m$ *unstructured. Assume that modes are sorted in such a way that* $\omega(\mathbf{A}^{(1)} \leq \omega(\mathbf{A}^{(1)} \leq \cdots \leq \omega(\mathbf{A}^{(d)}$. *Then* $\mathbf{A}^{(1)}$ *and* $\mathbf{A}^{(2)}$ *can be computed by solving an overdetermined linear system, and by computing* R *rank-1 matrix approximations, provided the necessary conditions below are met:*

$$\omega(\mathbf{A}^{(1)})\omega(\mathbf{A}^{(2)}) - \frac{1}{4}K_1^2K_2^2 \;\leq\; 0 \tag{15}$$

$$R^2 - (K_1K_2)R + \omega(\mathbf{A}^{(1)})\omega(\mathbf{A}^{(2)}) - 1 \;\leq\; 0 \tag{16}$$

These conditions may be less restrictive on R, if $\omega(\mathbf{A}^{(1)})$ and $\omega(\mathbf{A}^{(2)})$ are small compared to the other $\omega(\mathbf{A}^{(\mu)})$. The proof is similar to that of Proposition **??**.

6 Solutions with refinements

The solutions that we provided above are valid when the model is exact. Most of these solutions still work when the model is inaccurate, or when data are corrupted by noise. However, it might be useful in some cases to run an iterative refinement, starting with the solution obtained by the previous algorithms, by iteratively minimizing the mismatch criterion

$$||\mathbf{T}^{(1)} - (\mathbf{A} \odot \mathbf{B})\mathbf{C}^{\mathsf{T}}||^2$$

by a local descent algorithm. In particular, two cases can take advantage of such refinements.

First the case when all matrix factors are structured, seen in Proposition **??**. Second, the case of linked matrix factors. Consider for instance a model where

$$f_q(\mathbf{A}, \mathbf{B}, \mathbf{C}) = 0,\ 1 \leq q \leq Q$$

for known differentiable functions f_q. If the noise is not too large, one can assume to be in a neighborhood of the set defined by these equations and resort to the implicit function theorem to obtain a set of linear equations in the perturbations of matrices $(\mathbf{A}, \mathbf{B}, \mathbf{C})$.

The simplest instance of linked matrices is encountered with symmetric tensors, i.e. when $\mathbf{A} = \mathbf{B} = \mathbf{C}$. The idea is thus to ignore this dependency, to compute the three factors, and to run afterwards a local refinement by minimizing the polynomial $||\mathbf{T}^{(1)} - (\mathbf{A} \odot \mathbf{A})\mathbf{A}^{\mathsf{T}}||^2$. This approach has shown very satisfactory results.

7 Examples and computer results

In this section, we provide more precise results for particular structures, and especially the triangular case in Section **??**. In addition, we report results of computer simulations. Three matrix factors are randomly generated, among which one or two are structured. The trial is repeated several times (at least 5 times) and the median of the relative error

$$||\mathbf{T}^{(1)} - (\mathbf{A} \odot \mathbf{B})\mathbf{C}^{\mathsf{T}}||/||\mathbf{T}^{(1)}||.$$

serves as a performance index. In the noiseless case, the experience is conducted for R ranging from 8 to 17 or 18. By construction, R is the rank of tensor $\boldsymbol{T}$. In the noisy case, R is fixed to 5, and the noise level is varied.

7.1 One banded Toeplitz factor

Assume the $K \times R$ matrix factor $\mathbf{C}$ is Toeplitz lower triangular with bandwidth $\omega(\mathbf{C}) = K - R + 1$. If the first matrix $\mathbf{C}(1)$ in the basis is the identity, assuming $\gamma(1) = 1$ means that we can assume $\mathbf{C}$ has ones on its diagonal. Beside the identity matrix $\mathbf{I}$, the next basis matrices $C(\ell)$ have ones on their ℓth subdiagonal, and are null elsewhere: $C(\ell)_{ij} = \delta(i - j - \ell)$, $i > j$, where $\delta(\cdot)$ denotes the Kronecker delta. In other words, we have

$$\mathbf{C} = \mathbf{I} + \sum_{\ell=2}^{\omega(\mathbf{C})} \gamma(\ell)\, \mathbf{C}(\ell)$$

In other words, we have assumed the constraint $\gamma(1) = 1$ in Lemma **??**. The remaining $KR - \omega(\mathbf{C}) - 1$ basis matrices may be obtained in a non unique manner by completion, under the orthonormality constraint. The rank condition (**??**) becomes, as a function of $\omega(\mathbf{C})$:

$$\omega(\mathbf{C})^2 - (K+1)\,\omega(\mathbf{C}) + 2\,K \leq 0$$

It can be checked that this condition admits solutions only for $K > 5$. For instance, for $K = 20$, the above condition becomes $2 < \omega(\mathbf{C}) < 19$.

Figure 1: Relative reconstruction error for a $20 \times 20 \times 20$ tensor. Left: one Lower triangular Toeplitz banded factor; Right: one Hankel factor.

7.2 One Hankel factor

Assume the $K \times R$ matrix factor $\mathbf{C}$ is Hankel. It is characterized by $\omega(\mathbf{C}) = K+R-1$ free parameters. The ℓth basis matrix $1 \le \ell \le \omega(\mathbf{C})$ is Hankel with a single nonzero antidiagonal, that is: $C(\ell)_{ij} = \delta(i+j-\ell-1)$, $1 \le i \le K$, $1 \le j \le R$. The remaining $KR - \omega(\mathbf{C})$ basis matrices are obtained in a non unique manner by completion, under the orthonormality constraint. The exact way they are obtained is irrelevant; what is important is to have a full orthonormal basis. According to Lemma **??**, we have hence assumed the constraint $\gamma(1) = 1$, which means here $C_{11} = 1$.

7.3 Two banded Toeplitz factors

We considered next a more general banded Toeplitz case. We took two factors with the same structure, i.e. $I = J$ and $\omega(\mathbf{A}) = \omega(\mathbf{B}) = \omega$. The rank condition (**??**) can be expressed as a function of ω as $R^2 - I^2\, R + \omega^2 - 1 \le 0$.

Figure 2: Relative reconstruction error for a $20 \times 20 \times 20$ tensor. Bottom: one Toeplitz factor; Top: two Toeplitz factors.

Simulations have been run for $I = 20$, $8 \leq R \leq 19$, and $\omega = 18$, with 12 subdiagonals and 5 superdiagonals. The computer experiments reported in Fig. **??** have been executed under these conditions, with one or two structured matrices, the other factors being drawn randomly.

7.4 Triangular factors

In the case of triangular matrix factors, the bound given in Proposition **??** can be improved. In fact, the upper bound $\omega(\mathbf{A})\omega(\mathbf{B})$ on the dimension of the space spanned by $\mathbf{A}(i) \odot \mathbf{B}(j)$ is too loose.

Similarly to Lemma **??**, we have

Lemma 14 *Let $\boldsymbol{T}$ be a tensor of dimensions $I \times J \times K$ and rank $R > 1$, admitting the* CanD $\boldsymbol{T} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \boldsymbol{I}$. *where matrices $\mathbf{A}$ and $\mathbf{B}$ are respectively lower and upper triangular, and $\mathbf{C}$ is full. then one can choose any two constraints among the three below:*

$$\mathbf{Diag}\{\mathbf{N}\} = \mathbf{I},\ \mathbf{Diag}\{\mathbf{A}\} = \mathbf{I},\ \mathbf{Diag}\{\mathbf{B}\} = \mathbf{I},$$

Proposition 15 *Consider the tensor of Lemma* **??**. *Then, the calculation of the three matrix factors may be calculated by solving R linear systems,* *provided the necessary condition below is satisfied:*

$$2 \leq I \leq R \leq J$$

Moreover, this condition is generically sufficient.

Proof. From Lemma **??**, we may assume that diagonal terms of both $\mathbf{A}$ and $\mathbf{B}$ are equal to 1. From lemma **??**, we use the equation $\mathbf{UN} = \mathbf{A} \odot \mathbf{B}$. In the latter, there is no coupling between the unknown in various columns. So consider the rth column, $1 \leq r \leq R$. We have IJ equations with R unknowns in $\mathbf{N}$, $I - r$ unknowns in $\mathbf{A}$ and $r - 1$ in $\mathbf{B}$; as a consequence, the number of unknowns is constant and equal to $I + R - 1$.

Among those equations, $(r-1)(J-r)$ have a zero right hand side, $I + J - 2$ have linear terms in $\mathbf{A}$ and $\mathbf{B}$, and one is equal to 1. So we have $(r-1)(J-r) + I + J - 1$ linear equations. In the worst case, we have only $I + J - 1$ linear equations (when $r = 1$, or $r = J$ if this can happen). In this worst cases (first and last columns), the number of linear equations is at least as large as the number of unknowns if $I + J - 1 \geq I + R - 1$. Yet, this is clearly satisfied when $J \geq R$, which is true by hypothesis in our proposition. As a conclusion, if $J > R$, matrices $\mathbf{N}$, $\mathbf{A}$ and $\mathbf{B}$ can be obtained by solving R independent overdetermined linear systems. If $J = R$, the system is over determined for all columns of $\mathbf{N}$ except for the first and the last. □

7.5 Presence of noise

In order to have an idea of the influence of noise, which destroys the structure that is exploited by the various algorithms presented, Monte Carlo simulations have been carried out as follows.

- Generate randomly the nonzero entries of factor matrices $\mathbf{A}$,. $\mathbf{B}$ and $\mathbf{C}$ according to the same uniform distribution.
- Compute tensor $\boldsymbol{T} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \boldsymbol{I}$.
- Draw randomly the entries of a full tensor $\boldsymbol{E}$ of same dimensions as $\boldsymbol{T}$ according to a zero-mean normal distribution.
- Normalize tensors $\boldsymbol{T}$ and $\boldsymbol{E}$ by their Frobenius norm
- for various noise levels, k, decompose the tensor $\boldsymbol{T}_\varepsilon = \boldsymbol{T} + k\,\boldsymbol{E}$.

Figure 3: Relative reconstruction error for a $5 \times 5 \times 5$ tensor of rank 5, with one lower triangular factor, and on upper triangular factor, and one full factor, in the presence of noise.

The relative error $||\boldsymbol{T}_\varepsilon - \boldsymbol{T}||/||\boldsymbol{T}||$ is reported in figures **??** and **??** as a function of the noise level k, for one single Toeplitz band factor, and two triangular factors, respectively.

Note that these performances can be dramatically improved by running a refinement as described in Section **??**.

7.6 Concluding remarks

When two matrices are structured, their identification conditions are easier to meet than in the case when only one is structured. Mainly necessary conditions have been given in this paper. But in some cases (like in the Toeplitz banded case, or the case with two triangular upper and lower factors), they are also sufficient. Sufficiency proofs will be provided in a forthcoming full paper. In addition, the presentation has been made in the real field for the sake of simplicity, but it extends straightforwardly to the complex field.

Maple and Matlab computer codes will be made available after publication.

Figure 4: Relative reconstruction error for a $20 \times 20 \times 20$ tensor of rank 5, with one Toeplitz Band factor, and 2 full factors, in the presence of noise.

References

[1] 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.

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

[3] J. Brachat, P. Comon, B. Mourrain, and E. Tsigaridas, "Symmetric tensor decomposition," *Linear Algebra and its Applications*, 2009, submitted, arXiv:0901.3706.

[4] D. Nion and L. De Lathauwer, "A block component model based blind DS-CDMA receiver," *IEEE Trans. Signal Proc.*, vol. 56, no. 11, pp. 5567–5579, 2008.

[5] A. Kibangou and G. Favier, "Non iterative solution for Parafac with a Toeplitz factor," in *Eusipco'09*, Glasgow, UK, Aug. 24-28 2009, Eurasip.

[6] 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.

[7] 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.

[8] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis*, Wiley, 2004.

[9] A. L. F. de Almeida, G. Favier, and J. C. M. Mota, "Parafac-based unified tensor modeling for wireless communication systems with application to blind multiuser equalization," *Signal Processing*, vol. 87, no. 2, pp. 337–351, Feb. 2007.