

Combined Deflation and Alternating Projections for Column-Wise Orthogonal Canonical Polyadic Tensor Decomposition

Alex P da Silva, Pierre Comon, André L F de Almeida

▶ **To cite this version:**

Alex P da Silva, Pierre Comon, André L F de Almeida. Combined Deflation and Alternating Projections for Column-Wise Orthogonal Canonical Polyadic Tensor Decomposition. 2016. hal-01567411

HAL Id: hal-01567411
https://hal.science/hal-01567411

Preprint submitted on 23 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Combined Deflation and Alternating Projections for Column-Wise Orthogonal Canonical Polyadic Tensor Decomposition

Alex P. da Silva[1,2], Pierre Comon[2], and André L. F. de Almeida[3]

[1]*CEA, 17 rue des Martyrs 38054 Grenoble cedex 9, France*
[2]*GIPSA-Lab, 11 rue des Mathematiques F-38402 Saint Martin d'Hères cedex, France*
[3]*Federal University of Ceará, Campus do Pici, CP 6005 60455-970, Fortaleza, CE, Brazil*
Email: alex.pereiradasilva@cea.fr

Abstract

The method of alternating projections is an efficient solution to a wide range of optimisation problems. Some important results on its local convergence motivate its usage in a number of applications. We propose an algorithm, combining deflation and alternating projections, to compute the column-wise orthogonal CP decomposition for three-way tensors. We present convergence results based on transversality and non-tangential concepts. To assess its performance, we draw numerical experiments for real and complex scenarios.

Keywords: alternating projection, deflation, column-wise orthogonal tensor, traversal manifold, non-tangential points, CP decomposition.

1. Introduction

The method of alternating projections consists of projecting iteratively a point onto manifolds, one after the other, in order to find a point in their intersection. To describe the method, let $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_n$ be n manifolds in some space $\mathcal{E}$. The alternating projection principle is given as follows

$$\mathbf{x}^{(k+1)} = \pi_{\mathcal{M}_n} \cdots \pi_{\mathcal{M}_2} \pi_{\mathcal{M}_1}(\mathbf{x}^{(k)}), \ \forall k = 0, 1, \ldots \tag{1}$$

where $\pi_{\mathcal{M}_i}$, $1 \leq i \leq n$ is the orthogonal projection operator onto $\mathcal{M}_i$, and k is the iteration index.

This method was originally proposed by von Neumann in 1930's for two manifolds, and he also proved that the sequence of alternating projections onto two linear manifolds converges to a point in their intersection [1].

Method (1) has also been much emphasized in the literature when the problem has two convex manifolds. We highlight, for instance, the articles [2, 3, 4, 5] for some discussions and application of the alternating convex projection method.

The extension of alternating projections to nonconvex manifolds plays an important role in several applications. We cite, for instance, those in control design [6] and in image processing [7].

For $n = 2$, some important theoretical results on the convergence of the alternating nonconvex projection method are highlighted in the following. In [8], it is shown that if the sequence $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ is bounded and the distance to $\mathcal{M}_1 \cap \mathcal{M}_2$ is strictly decreasing, then there exists a convergent subsequence to a point in $\mathcal{M}_1 \cap \mathcal{M}_2$. This result is known as the *Zangwill's global convergence theorem*. However, it is a weak result since it does not provide any information about the convergence of the complete sequence.

The authors in [9] introduce the concept of transversal manifolds and show that a linear convergence of the alternating projection method can be ensured if the manifolds are transversal at some point in $\mathcal{M}_1 \cap \mathcal{M}_2$ and the starting point $\mathbf{x}^{(0)}$ is close enough to $\mathcal{M}_1 \cap \mathcal{M}_2$. However, this approach presents a restriction on the dimensions of the associated manifolds. In order to drop the limitation of the transversal concept, the authors in [10] proposed a new concept called *non-tangential intersection point*, which generalizes that of transversality.

In this paper, both the transversality and non-tangential concepts will be studied in the context of column-wise orthogonal tensor decomposition. Since the convergence of tensor algorithms is a topic of interest in tensor research,

we reformulate this decomposition in order to tackle it by alternating projections and, thus, make use of the strong results associated with both concepts.

This paper is organised as follows. In Section 2, we describe the column-wise orthogonal tensor descomposition. Section 3 presents the proposed algorithm that combines alternating projection and deflation to decompose a column-wise orthogonal tensor. In Section 4, we study the convergence of our algorithm under some conditions, including the transversality and the non-tangential concepts for points at the intersection of manifolds. Section 5, reports a set of results supporting the advantages of our algorithm. Finally, Section 6 draws some conclusions.

Notation: Scalar numbers, vectors, matrices and tensors are denoted by lowercase, boldface lowercase, boldface capital and calligraphic letters, respectively. Plain capitals are used to denote array dimensions. Greek letters also denote scalars. The symbols $\odot$, $\boxtimes$ and $\otimes$ denote the Khatri-Rao, Kronecker and tensorial products, respectively, and $^+$, *, $^{\mathsf{H}}$ and $^{\mathsf{T}}$ denote matrix pseudo-inversion, conjugate, conjugate transpose and transpose. The Euclidean scalar product between tensors is denoted by $\langle \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{U}} \rangle = \sum_{i_1 \cdots i_N} T_{i_1 \cdots i_N} U^*_{i_1 \cdots i_N}$, and $\| \cdot \|$ then denotes the Frobenius norm induced by the previous scalar product. We shall also use the ker and range operators, which correspond to the null and the range space of a matrix. The operator vec is the vectorization operator that stacks the columns of a matrix into a long column vector, and Unvec is its reverse operator. Finally, $\mathbb{K}$ is either the real or the complex field.

2. Column-wise orthogonal tensor decomposition

A rank-R tensor can be written as $\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{X}}_1 + \boldsymbol{\mathcal{X}}_2 + \cdots + \boldsymbol{\mathcal{X}}_R$, where $\boldsymbol{\mathcal{X}}_i$ are rank-1 (*i.e.* decomposable) tensors. A tensor is orthogonal if its rank-1 components are orthogonal to one another, i.e., if $\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{X}}_j \rangle = 0, \ \forall i \neq j$. The notion of orthogonality between a pair of decomposable tensors depends on the level of coupling between their factors. To be more precise, consider two three-way decomposable tensors $\boldsymbol{\mathcal{X}}_1 = \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1$ and $\boldsymbol{\mathcal{X}}_2 = \mathbf{a}_2 \otimes \mathbf{b}_2 \otimes \mathbf{c}_2$. The tensors $\boldsymbol{\mathcal{X}}_1$ and $\boldsymbol{\mathcal{X}}_2$ are completely orthogonal if $\mathbf{a}_1 \perp \mathbf{a}_2, \mathbf{b}_1 \perp \mathbf{b}_2$ and $\mathbf{c}_1 \perp \mathbf{c}_2$. Actually, all factors do not need to be orthogonal to one another to ensure the orthogonality of the tensor. For instance, if only $\mathbf{a}_1 \perp \mathbf{a}_2$, then $\langle \boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{X}}_2 \rangle = 0$. Indeed, $\langle \boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{X}}_2 \rangle = \langle \mathbf{a}_1, \mathbf{a}_2 \rangle \langle \mathbf{b}_1, \mathbf{b}_2 \rangle \langle \mathbf{c}_1, \mathbf{c}_2 \rangle$. Herein, for a rank-$R$ tensor, we assume that the factor matrix $\mathbf{A}$ of a tensor $\boldsymbol{\mathcal{T}}$ is semi-unitary[1], which implies that $\langle \mathbf{a}_i, \mathbf{a}_j \rangle = \delta(i-j)$, ensuring the orthogonality of $\boldsymbol{\mathcal{T}}$. We call *column-wise orthogonal tensor decomposition* the decomposition of a tensor whose at least one of the factors matrices is semi-unitary. Notice that if all factor matrices are semi-unitary then the tensor is of unit norm and completely orthogonal. For other notions of orthogonality, see [11].

The column-wise orthogonal tensor decomposition plays an important role in several applications, such as blind source separation [12], array processing[13], and wireless communication systems [14]. Let?s point out some related results available in the literature. In [15], the authors show that the best orthogonal low rank approximation of a tensor, with one semi-unitary factor matrix, always exists. Additionally, for almost all tensors, they also show that the ALS algorithm converges globally. In [16], the authors discuss the uniqueness and low-rank approximation properties of a CP decomposition with a semi-unitary matrix factor. In the following, we apply the results of alternating projection methods to solve this kind of decomposition.

3. Combined deflation and alternating projections algorithm

Let $\mathbf{A} \in \mathbb{K}^{I \times R}, \mathbf{B} \in \mathbb{K}^{J \times R}$ and $\mathbf{C} \in \mathbb{K}^{K \times R}$ be the factor matrices forming the CP decomposition of a tensor $\boldsymbol{\mathcal{T}}$ with entries in some field $\mathbb{K}$: in other words, if $\mathbf{a}_r$ (resp. $\mathbf{b}_r$ and $\mathbf{c}_r$) denote the columns of $\mathbf{A}$ (resp. $\mathbf{B}$ and $\mathbf{C}$), $1 \leq r \leq R$, tensor $\boldsymbol{\mathcal{T}}$ can be decomposed as $\boldsymbol{\mathcal{T}} = \sum_{r=1}^{R} \boldsymbol{\mathcal{X}}_r$, with $\boldsymbol{\mathcal{X}}_r = \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r$. We assume that $\text{rank}\{\boldsymbol{\mathcal{T}}\} = R$ is known, and that one of the factor matrices is column-wise orthogonal, say $\mathbf{A}$. This implies $R \leq I$. Actually, without loss of generality, $\mathbf{A}$ can be viewed as a semi-unitary matrix due to scalar elementary indeterminacies (unitary if $R = I$).

Let $\mathbf{T} = \mathbf{A} (\mathbf{C} \odot \mathbf{B})^{\mathsf{T}}$ be the mode-1 unfolding of tensor $\boldsymbol{\mathcal{T}}$, as defined for instance in [17]. Since $\mathbf{A}$ is semi-unitary, $\mathbf{A}^{\mathsf{H}}\mathbf{A} = \mathbf{I}$. Therefore, after elementary operations

$$\mathbf{T}^{\mathsf{H}}\mathbf{A} = \mathbf{C}^* \odot \mathbf{B}^* \tag{2}$$

[1] The upcoming reasoning is identical if $\mathbf{B}$ or $\mathbf{C}$ were semi-unitary.

For a specific column of $\mathbf{A}$, namely $\mathbf{a}_r$, we derive from equation (2) that

$$\mathbf{T}^{\mathsf{H}}\mathbf{a}_r = \mathbf{c}_r^* \boxtimes \mathbf{b}_r^*, \tag{3}$$

where $\mathbf{b}_r$ and $\mathbf{c}_r$ are the r-th column of matrices $\mathbf{B}$ and $\mathbf{C}$, respectively.

We propose an algorithm called CAPD, which stands for *Combined Alternating Projections and Deflation*. CAPD recovers all factor matrices of a given orthogonal tensor $\mathcal{T}$ having one semi-unitary factor. The idea is to solve first the equation (3) with an alternating projection algorithm in order to estimate the factors $\mathbf{a}_r$, $\mathbf{b}_r$ and $\mathbf{c}_r$, and to perform the deflation of the rank-1 component $\mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r$ from $\mathcal{T}$, obtaining thus a rank- $(R-1)$ tensor $\mathcal{T} - \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r$. The process is repeated[2] $R-2$ times until all rank-1 components are estimated. It is important to remind that the deflation procedure does not generally work for non-orthogonal tensors [18].

Let $\mathbf{T}_{[r]}$ be the mode-1 unfolding of a rank-$(R-r+1)$ tensor in $\mathbb{K}^{I\times J\times K}$ obtained by deflation, $1 \le r \le R$; we have in particular $\mathbf{T}_{[1]} = \mathbf{T}$. We define the following linear subspace

$$\mathcal{M}_r = \{\mathbf{x} \in \mathbb{K}^{JK} : \left(\mathbf{T}_{[r]}^{\mathsf{H}}\mathbf{T}_{[r]}^{\mathsf{H}+} - \mathbf{I}\right)\mathbf{x} = \mathbf{0}\}, \tag{4}$$

and the manifold

$$\mathcal{N} = \{\mathbf{x} \in \mathbb{K}^{JK} : \mathbf{x} = \mathbf{c} \boxtimes \mathbf{b}, \|\mathbf{x}\| = 1, \forall \mathbf{b} \in \mathbb{K}^J \setminus \{\mathbf{0}\} \text{ and } \forall \mathbf{c} \in \mathbb{K}^K \setminus \{\mathbf{0}\}\ \}. \tag{5}$$

The former is the column space of $\mathbf{T}_{[r]}^{\mathsf{H}}\mathbf{T}_{[r]}^{\mathsf{H}+}$, that is, the space spanned by rows of $\mathbf{T}_{[r]}$, and the latter is the subset of unit Kronecker vectors in $\mathbb{K}^{JK}$, respectively. By applying the Unvec operator on $\mathbf{x}$, $\mathcal{N}$ can be actually viewed as the set of $J \times K$ unit-norm rank-1 matrices. In the proposed CAPD algorithm, the factors $\mathbf{a}_r$, $\mathbf{b}_r$ and $\mathbf{c}_r$ are computed by using the alternating projection onto these two manifolds. We set initially $\mathbf{T}_{[1]} = \mathbf{T}$. CAPD is described in Algorithm 1. Upper indices (k) denote the k-th iteration of the algorithm, and $\pi_{\mathcal{M}_r}$ and $\pi_{\mathcal{N}}$ are the projection operators onto $\mathcal{M}_r$ and $\mathcal{N}$, respectively. Notice that $\pi_{\mathcal{M}_r} = \mathbf{T}_{[r]}^{\mathsf{H}}\mathbf{T}_{[r]}^{\mathsf{H}+}$.

The algorithm works as follows. For every $r \in \{1, 2, \ldots, R-1\}$, the vectors $\mathbf{b}_r^{(0)}$ and $\mathbf{c}_r^{(0)}$ are randomly initialized so that $\mathbf{c}_r^{(0)} \boxtimes \mathbf{b}_r^{(0)} \notin \ker(\mathbf{T}_{[r]})$, otherwise $\mathbf{x}_r^{(0)} = \mathbf{0}$ and the algorithm would not extract the r-th rank-1 component. The vector $\mathbf{x}_r^{(0)}$ is introduced into a *repeat* loop that consists of the alternating projections phase. If $\mathbf{x}_r^{(k)}$ converges to a point in $\mathcal{M}_r \cap \mathcal{N}$, then the estimated factor $\widehat{\mathbf{a}}_r$ is computed by solving in the least square sense equation (3) for a given $\mathbf{x}_r^{(k)} = \mathbf{c}_r^{(k)} \boxtimes \mathbf{b}_r^{(k)}$. Clearly, the factors $\widehat{\mathbf{b}}_r$ and $\widehat{\mathbf{c}}_r$ are directly obtained from the singular vectors of $\text{Unvec}(\mathbf{x}_r^{(k)})$ as shown in Algorithm 1.

After the estimation of $\widehat{\mathbf{a}}_r, \widehat{\mathbf{b}}_r$, and $\widehat{\mathbf{c}}_r$, we deflate the original tensor by the rank-1 component $\widehat{\mathbf{a}}_r(\widehat{\mathbf{c}}_r \boxtimes \widehat{\mathbf{b}}_r)^{\mathsf{T}}$, and perform again $R-2$ times the alternating projection procedure in order to find the other components of the factor matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$.

Remark 1. *Neither the convergence of the CAPD algorithm to a point $\mathbf{x}_r^{\infty} \in \mathcal{M}_r \cap \mathcal{N}$ nor the estimation of the right factors $\mathbf{a}_r, \mathbf{b}_r$ and $\mathbf{c}_r$ can be ensured. However, theoretical results in the next section guarantee local convergence and the extraction of the suitable factors.*

Remark 2. *Instead of estimating one-by-one the rank-1 components, we could try to estimate all components all-at-once from equation (2). However, the orthogonality of the estimated factor $\widehat{\mathbf{A}}$ would not be ensured anymore unless a constraint is imposed. Actually, all columns of the matrix factors could be collinear, as we will show later in Section 5. Moreover, the estimation of all factors all-at-once requires more computations.*

4. Convergence study

This section reports new results on the convergence of CAPD algorithm. We start with the proof of some essential lemmas, which will be important to demonstrate that the extraction of one rank-1 component of the tensor is ensured if the CAPD method converges to some point in $\mathcal{M}_r \cap \mathcal{N}$ for every $r \in \{1, 2, \ldots, R-1\}$. Finally, based on the ideas within [9], we show that CAPD algorithm has locally linear convergence properties.

[2] After the last deflation, the algorithm yields a rank-1 tensor, whose factors are easily obtained by computing an economic SVD.

input : $\mathcal{T} \in \mathbb{K}^{I \times J \times K}$: input data
output: $\widehat{\mathbf{A}}, \widehat{\mathbf{B}}$ and $\widehat{\mathbf{C}}$: factor matrices
$\mathbf{T} \leftarrow$ mode-1 unfolding of $\mathcal{T}$;
$\mathbf{T}_{[1]} \leftarrow \mathbf{T}$;
Define $\mathcal{N}$;
for $r = 1$ *to* $R - 1$ **do**
  Define the linear subspace $\mathcal{M}_r$ from $\mathbf{T}_{[r]}$;
  Initialize $\mathbf{b}_r^{(0)}$ and $\mathbf{c}_r^{(0)}$;
  $\mathbf{x}_r^{(0)} \leftarrow \pi_{\mathcal{M}_r}(\mathbf{c}_r^{(0)} \boxtimes \mathbf{b}_r^{(0)})$;
  $k \leftarrow 0$;
  repeat
    $k \leftarrow k + 1$;
    $\mathbf{x}_r^{(k)} \leftarrow \pi_{\mathcal{M}_r}\pi_{\mathcal{N}}(\mathbf{x}_r^{(k-1)})$;
  until *some stopping criteria are satisfied*;
  $\widehat{\mathbf{a}}_r \leftarrow \mathbf{T}_{[r]}^{\mathsf{H}+}\mathbf{x}_r^{(k)}$;
  $\widehat{\mathbf{b}}_r \leftarrow$ conjugate of the left singular vector of Unvec$(\mathbf{x}_r^{(k)})$;
  $\widehat{\mathbf{c}}_r \leftarrow$ right singular vector of Unvec$(\mathbf{x}_r^{(k)})$;
  $\widehat{\mathbf{A}}(:, r) \leftarrow \widehat{\mathbf{a}}_r$; $\widehat{\mathbf{B}}(:, r) \leftarrow \widehat{\mathbf{b}}_r$; $\widehat{\mathbf{C}}(:, r) \leftarrow \widehat{\mathbf{c}}_r$;
  $\mathbf{T}_{[r+1]} \leftarrow \mathbf{T}_{[r]} - \widehat{\mathbf{a}}_r(\widehat{\mathbf{c}}_r \boxtimes \widehat{\mathbf{b}}_r)^{\mathsf{T}}$;
end
$\mathbf{T}_{[R]}$ is the mode-1 unfolding of the last rank-1 component of $\mathcal{T}$.

Algorithm 1: CAPD algorithm

Lemma 4.1. *Let* $\mathbf{c}_r^{(k)} \boxtimes \mathbf{b}_r^{(k)} = \pi_{\mathcal{N}}(\mathbf{x}_r^{(k-1)})$ *be the best approximation of vector* $\mathbf{x}_r^{(k-1)}$ *in* $\mathcal{N}$ *at k-th iteration of CAPD algorithm. If* $\mathbf{c}_r^{(0)} \boxtimes \mathbf{b}_r^{(0)} \notin \ker(\mathbf{T}_{[r]})$ *then for every* $k \geq 1$, $\mathbf{c}_r^{(k)} \boxtimes \mathbf{b}_r^{(k)} \notin \ker(\mathbf{T}_{[r]})$ *either.*

Proof. $\mathbf{c}_r^{(0)} \boxtimes \mathbf{b}_r^{(0)} \notin \ker(\mathbf{T}_{[r]}) = \ker(\mathbf{T}_{[r]}^{\mathsf{H}+}) \implies \mathbf{x}_r^{(0)} \in \text{range}(\mathbf{T}_{[r]}^{\mathsf{H}})\backslash\{\mathbf{0}\}$. Since $\mathbf{c}_r^{(1)} \boxtimes \mathbf{b}_r^{(1)} \not\perp \mathbf{x}_r^{(0)}$, it follows that $\mathbf{c}_r^{(1)} \boxtimes \mathbf{b}_r^{(1)} \notin \ker(\mathbf{T}_{[r]})$. The same reasoning can be applied to the next iterations, and the proof is complete. ∎

Lemma 4.2. *Let* $\mathbf{a}_r^{(k)} = \mathbf{T}_{[r]}^{\mathsf{H}+}(\mathbf{c}_r^{(k)} \boxtimes \mathbf{b}_r^{(k)})$ *be the minimal norm solution at k-th iteration in CAPD algorithm. Then for all* $k \geq 0$, $\mathbf{a}_r^{(k)}$ *is a linear combination of the columns of* $\mathbf{A}$.

Proof. It follows directly from the mode-1 unfolding $\mathbf{T} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}}$ that the columns of $\mathbf{A}$ span range$(\mathbf{T})$. Yet, $\mathbf{a}_r^{(k)} \in \text{range}(\mathbf{T}_{[r]}^{\mathsf{H}+}) = \text{range}(\mathbf{T}_{[r]}) \subseteq \text{range}(\mathbf{T})$. ∎

Theorem 4.3. *Let* $\mathcal{T}$ *be a tensor whose factor* $\mathbf{A}$ *is a column-wise orthogonal matrix. Assume that* $\mathcal{T}$ *has an essentially unique CP decomposition, and* $\mathbf{c}_r^{(0)} \boxtimes \mathbf{b}_r^{(0)} \notin \ker(\mathbf{T}_{[r]})$, $1 \leq r \leq R - 1$. *If* $\mathbf{x}_r^{(k)}$, $k \geq 0$, *converges to a point* $\mathbf{x}_r^{\infty} \in \mathcal{M}_r \cap \mathcal{N}$ *in CAPD algorithm, then* $\widehat{\mathbf{a}}_r \otimes \widehat{\mathbf{b}}_r \otimes \widehat{\mathbf{c}}_r$ *is one of the rank-1 components of* $\mathcal{T}$ *in its CP decomposition.*

Proof. From Lemma 4.1 $\mathbf{x}_r^{\infty} \neq \mathbf{0}$ because $\mathbf{c}_r^{(0)} \boxtimes \mathbf{b}_r^{(0)} \notin \ker(\mathbf{T}_{[r]})$, which implies that $\widehat{\mathbf{a}}_r \neq \mathbf{0}$, $\widehat{\mathbf{b}}_r \neq \mathbf{0}$, and $\widehat{\mathbf{c}}_r \neq \mathbf{0}$. From Lemma 4.2, $\widehat{\mathbf{a}}_r$ can be written as a linear combination of the columns of factor $\mathbf{A}$. That is, $\widehat{\mathbf{a}}_r = \sum_{i=1}^{R} \alpha_i \mathbf{a}_i$.

Since $\mathbf{x}_r^{(k)} \to \mathbf{x}_r^{\infty} \in \mathcal{M}_r \cap \mathcal{N}$, it follows that $\widehat{\mathbf{a}}_r, \widehat{\mathbf{b}}_r$ and $\widehat{\mathbf{c}}_r$ satisfy equation (3), so that $\mathbf{x}_r^{\infty} = \widehat{\mathbf{c}}_r^* \boxtimes \widehat{\mathbf{b}}_r^*$. Using twice (3), first for substituting $\widehat{\mathbf{a}}_r$, and then $\mathbf{a}_i$, we get:

$$\mathbf{T}^{\mathsf{H}}\widehat{\mathbf{a}}_r = \widehat{\mathbf{c}}_r^* \boxtimes \widehat{\mathbf{b}}_r^* \implies \sum_{i=1}^{R} \alpha_i \mathbf{T}^{\mathsf{H}}\mathbf{a}_i = \widehat{\mathbf{c}}_r^* \boxtimes \widehat{\mathbf{b}}_r^* \implies \widehat{\mathbf{c}}_r \boxtimes \widehat{\mathbf{b}}_r = \sum_{i=1}^{R} \alpha_i^*(\mathbf{c}_i \boxtimes \mathbf{b}_i).$$

Now, let $\alpha_j \neq 0$ for some $j \in \{1, 2, \ldots, R\}$, and write

$$\mathbf{c}_j \boxtimes \mathbf{b}_j = \frac{1}{\alpha_j^*}\left(\widehat{\mathbf{c}}_r \boxtimes \widehat{\mathbf{b}}_r - \sum_{\substack{i=1 \\ i\neq j}}^{R} \alpha_i^*(\mathbf{c}_i \boxtimes \mathbf{b}_i)\right).$$

Yet, a decomposition of $\boldsymbol{\mathcal{T}}$ along the mode-1 is given by

$$\mathbf{T} = \mathbf{a}_1(\mathbf{c}_1 \boxtimes \mathbf{b}_1)^{\mathsf{T}} + \cdots + \mathbf{a}_R(\mathbf{c}_R \boxtimes \mathbf{b}_R)^{\mathsf{T}}.$$

Thus, plugging the expression of $\mathbf{c}_j \boxtimes \mathbf{b}_j$ into that of $\mathbf{T}$ and reorganizing the factors we obtain

$$\mathbf{T} = \frac{1}{\alpha_j^*}\mathbf{a}_j(\widehat{\mathbf{c}}_r \boxtimes \widehat{\mathbf{b}}_r)^{\mathsf{T}} + \sum_{i\neq j}\left(\mathbf{a}_i - \frac{\alpha_i^*}{\alpha_j^*}\mathbf{a}_j\right)(\mathbf{c}_i \boxtimes \mathbf{b}_i)^{\mathsf{T}}.$$

It is important to mention that $\widehat{\mathbf{c}}_r \boxtimes \widehat{\mathbf{b}}_r$ is not in the column space of the vectors $\{\mathbf{c}_1 \boxtimes \mathbf{b}_1, \ldots, \mathbf{c}_{j-1} \boxtimes \mathbf{b}_{j-1}, \mathbf{c}_{j+1} \boxtimes \mathbf{b}_{j+1}, \ldots, \mathbf{c}_R \boxtimes \mathbf{b}_R\}$, otherwise $\operatorname{rank}\{\boldsymbol{\mathcal{T}}\} < R$.

Since the first factor matrix is column-wise orthogonal and the decomposition is essentially unique, the vectors

$$\left\{\mathbf{a}_1 - \frac{\alpha_1^*}{\alpha_j^*}\mathbf{a}_j, \ldots, \mathbf{a}_{j-1} - \frac{\alpha_{j-1}^*}{\alpha_j^*}\mathbf{a}_j, \frac{1}{\alpha_j^*}\mathbf{a}_j, \mathbf{a}_{j+1} - \frac{\alpha_{j+1}^*}{\alpha_j^*}\mathbf{a}_j, \ldots, \mathbf{a}_R - \frac{\alpha_R^*}{\alpha_j^*}\mathbf{a}_j\right\}$$

are orthogonal to one another. In particular, for $i \in \{1, 2, \ldots, R\} - \{j\}$

$$\left(\mathbf{a}_i - \frac{\alpha_i^*}{\alpha_j^*}\mathbf{a}_j\right)^{\mathsf{H}}\left(\frac{1}{\alpha_j^*}\mathbf{a}_j\right) = 0 \implies \alpha_i = 0,$$

which implies that $\widehat{\mathbf{a}}_r = \alpha_j\mathbf{a}_j$ and $\widehat{\mathbf{c}}_r \boxtimes \widehat{\mathbf{b}}_r = \alpha_j^*(\mathbf{c}_j \boxtimes \mathbf{b}_j)$. As $\widehat{\mathbf{c}}_r \boxtimes \widehat{\mathbf{b}}_r$ is a unit vector, it follows that $|\alpha_j| = 1/\|\mathbf{c}_j \boxtimes \mathbf{b}_j\|$.

To eliminate scalar indeterminacies, assume $\|\mathbf{c}_j \boxtimes \mathbf{b}_j\| = 1$, which leads to $|\alpha_j| = 1$. Thus,

$$\widehat{\mathbf{a}}_r \otimes \widehat{\mathbf{b}}_r \otimes \widehat{\mathbf{c}}_r = |\alpha_j|\mathbf{a}_j \otimes \mathbf{b}_j \otimes \mathbf{c}_j = \mathbf{a}_j \otimes \mathbf{b}_j \otimes \mathbf{c}_j,$$

and the proof is complete. ∎

Corollary 4.4. $\mathbf{x}_r^\infty \in \mathcal{M}_r \cap \mathcal{N} \implies \mathbf{x}_r^\infty$ *is a column of the Khatri-Rao product* $\mathbf{C}^* \odot \mathbf{B}^*$.

Proof. Up to scalar elementary indeterminances, it follows from Theorem 4.3 that $\mathbf{x}_r^\infty = \widehat{\mathbf{c}}_r^* \boxtimes \widehat{\mathbf{b}}_r^* = \alpha_j(\mathbf{c}_j^* \boxtimes \mathbf{b}_j^*)$, for some $j \in \{1, 2, \ldots, R\}$. ∎

Theorem 4.3 ensures that one rank-1 component of tensor $\boldsymbol{\mathcal{T}}$ can be extracted if CAPD converges to a limit point $\mathbf{x}_r^\infty \in \mathcal{M}_r \cap \mathcal{N}$. Actually, the convergence is not globally ensured since $\mathcal{N}$ is a non-convex manifold, which means that CAPD can get stuck if the starting point is badly chosen. Therefore, we focus the convergence study of the CAPD algorithm on starting points close to $\mathcal{M}_r \cap \mathcal{N}$, in which case we can draw some important results in the real field. Before, we introduce two basic definitions on manifolds.

Definition 4.5. *[9] Let $\mathcal{A}$ and $\mathcal{B}$ be two C^k-manifolds around a point $\mathbf{x} \in \mathcal{A} \cap \mathcal{B}$ in some space $\mathcal{E}$. The manifolds $\mathcal{A}$ and $\mathcal{B}$ are transverse at $\mathbf{x}$ if*

$$T_{\mathcal{A}}(\mathbf{x}) + T_{\mathcal{B}}(\mathbf{x}) = \mathcal{E},$$

where $T_{\mathcal{A}}(\mathbf{x})$ and $T_{\mathcal{B}}(\mathbf{x})$ are the tangent spaces to $\mathcal{A}$ and $\mathcal{B}$ at $\mathbf{x}$, respectively.

Definition 4.6. *[10] Let $\mathcal{A}$ and $\mathcal{B}$ be two C^k-manifolds around a point $\mathbf{x} \in \mathcal{A}\cap\mathcal{B}$ in some space $\mathcal{E}$. $\mathbf{x}$ is a non-tangential point if and only if*

$$T_{\mathcal{A}}(\mathbf{x}) \cap T_{\mathcal{B}}(\mathbf{x}) = T_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x}),$$

where $T_{\mathcal{A}}(\mathbf{x})$, $T_{\mathcal{B}}(\mathbf{x})$ and $T_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x})$ are the tangent spaces to $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{A} \cap \mathcal{B}$ at $\mathbf{x}$, respectively

4.1. Convergence under transversality condition

According to [9], if two smooth manifolds are transverse, and the initialization is close enough to their intersection, then the alternating projections method converges linearly to a point at the intersection of the manifolds. This is stated in the following theorem.

Theorem 4.7. *[9] Let $\mathcal{A}$ and $\mathcal{B}$ be two transverse manifolds around a point $\bar{\mathbf{x}} \in \mathcal{A} \cap \mathcal{B}$ in some space $\mathcal{E}$. If the initial point $\mathbf{x}_0 \in \mathcal{E}$ is close to $\bar{\mathbf{x}}$, then the method of alternating projections*

$$\mathbf{x}^{(k+1)} = \pi_{\mathcal{A}}\pi_{\mathcal{B}}(\mathbf{x}^{(k)}),\ (k = 0, 1, 2, \ldots)$$

is well-defined, and the distance $d_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x}_k)$ from the iterate $\mathbf{x}_k$ to the intersection $\mathcal{A} \cap \mathcal{B}$ decreases Q-linearly[3] to zero.

In our case, $\mathcal{M}_r$ are linear subspaces and $\mathcal{N}$ is locally smooth in the Euclidean space (see Example 2 in [9]). Thus, under particular conditions stated below, the manifolds $\mathcal{M}_r$ and $\mathcal{N}$ are transverse for every $r \in \{1, 2, \ldots, R-1\}$.

Lemma 4.8. *Let $\mathcal{R} = \{\mathbf{x} \in \mathbb{R}^{JK} : \mathbf{x} = \mathbf{c} \boxtimes \mathbf{b}, \forall \mathbf{b} \in \mathbb{R}^J \setminus \{\mathbf{0}\}$ and $\forall \mathbf{c} \in \mathbb{R}^K \setminus \{\mathbf{0}\}\}$. Then $\forall \mathbf{x} \in \mathcal{R}$ it follows that $T_{\mathcal{R}}(\mathbf{x}) = T_{\mathcal{N}}(\mathbf{x}/\|\mathbf{x}\|)$.*

Proof. Note that $\mathcal{R}$ can be viewed as the set of $J \times K$ rank-1 matrices. According to [9], the tangent space of $\mathcal{R}$ is given by

$$T_{\mathcal{R}}(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^{JK} : \mathbf{u}_i^{\mathsf{T}} \operatorname{Unvec}(\mathbf{y})\mathbf{v}_j = 0, \forall (i, j), 1 < i \leq J, 1 < j \leq K\}$$

where $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_J\}$ and $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K\}$ are the sets of left and right singular vectors of the matrix $\operatorname{Unvec}(\mathbf{x}) = \sigma_1\mathbf{u}_1\mathbf{v}_1^{\mathsf{T}},\ \sigma_1 > 0$.

Define $\mathcal{U} = \{\mathbf{x} \in \mathbb{R}^{JK} : \|\mathbf{x}\| = 1\}$. Since the tangent space does not depend on the singular value σ_1, and $\mathcal{R} \cap \mathcal{U} = \mathcal{N}$, it follows that $T_{\mathcal{R}}(\mathbf{x}) = T_{\mathcal{R}}(\mathbf{x}/\|\mathbf{x}\|) = T_{\mathcal{N}}(\mathbf{x}/\|\mathbf{x}\|)$. ∎

Proposition 4.9. *Let $\mathcal{T} \in \mathbb{R}^{I\times J\times K}$ be a rank-R tensor with factor $\mathbf{A}$ column-wise orthogonal. Assume neither $\mathbf{C}$ nor $\mathbf{B}$ have collinear columns. In the CAPD algorithm, $\mathcal{M}_r$ and $\mathcal{N}$ are transverse manifolds at any point $\mathbf{x} \in \mathcal{M}_r \cap \mathcal{N}$ if $r \leq R - (J-1)(K-1)$, for $r \in \{1, 2, \ldots, R-1\}$.*

Proof. Let $\mathbf{x} \in \mathcal{M}_r \cap \mathcal{N}$, for some $r \in \{1, 2, \ldots, R-1\}$. According to Lemma 4.8, the tangent space of $\mathcal{N}$ at $\mathbf{x}$ is given by

$$T_{\mathcal{N}}(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^{JK} : \mathbf{u}_i^{\mathsf{T}} \operatorname{Unvec}(\mathbf{y})\mathbf{v}_j = 0, \forall (i, j), 1 < i \leq J, 1 < j \leq K\},$$

for left and right singular vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_J\}$ and $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K\}$ of the matrix $\operatorname{Unvec}(\mathbf{x}) = \mathbf{u}_1\mathbf{v}_1^{\mathsf{T}}$ of dimension $J \times K$.

Let $\mathcal{K}(\mathbf{x}) = \{\operatorname{vec}(\mathbf{u}_1\mathbf{v}_1^{\mathsf{T}}), \operatorname{vec}(\mathbf{u}_1\mathbf{v}_2^{\mathsf{T}}), \ldots, \operatorname{vec}(\mathbf{u}_1\mathbf{v}_K^{\mathsf{T}}),$
$\operatorname{vec}(\mathbf{u}_2\mathbf{v}_1^{\mathsf{T}}), \ldots, \operatorname{vec}(\mathbf{u}_J\mathbf{v}_1^{\mathsf{T}})\}$. Note that all vectors in $\mathcal{K}(\mathbf{x})$ are orthogonal to one another so that they span a $J + K - 1$ subspace in $\mathbb{R}^{JK}$.

Let also $\mathcal{D}(\mathbf{x}) = \{\mathbf{z} \in \mathbb{R}^{JK} : \mathbf{z} = \operatorname{vec}(\mathbf{u}_i\mathbf{v}_j^{\mathsf{T}}),\ \ \forall 1 < i \leq J, 1 < j \leq K\}$. Clearly, $\mathcal{D}(\mathbf{x}) \cap T_{\mathcal{N}}(\mathbf{x}) = \emptyset$. Actually, $\mathcal{D}(\mathbf{x})$ is a set of orthogonal vectors that is the complement of $T_{\mathcal{N}}(\mathbf{x})$. Indeed, $\mathcal{D}(\mathbf{x}) + \mathcal{K}(\mathbf{x}) = \mathbb{R}^{JK} \implies \mathcal{D}(\mathbf{x}) + T_{\mathcal{N}}(\mathbf{x}) = \mathbb{R}^{JK}$ because $\mathcal{K}(\mathbf{x}) \subseteq T_{\mathcal{N}}(\mathbf{x})$. Thus, we can conclude that $J + K - 1$ is the dimension of $T_{\mathcal{N}}(\mathbf{x}) = \operatorname{Span}\{\mathcal{K}(\mathbf{x})\}$.

Now, define $\mathcal{H}$ the set composed of the possible columns of $\mathbf{C} \odot \mathbf{B}$. Since neither $\mathbf{C}$ nor $\mathbf{B}$ have collinear columns,

$$\left\{\mathcal{K}(\mathbf{x}) - \{\operatorname{vec}(\mathbf{u}_1\mathbf{v}_1^{\mathsf{T}})\}\right\} \cap \mathcal{H} = \emptyset,$$

and thereby $\mathcal{K}(\mathbf{x}) - \{\operatorname{vec}(\mathbf{u}_1\mathbf{v}_1^{\mathsf{T}})\} \nsubseteq \mathcal{M}_r = T_{\mathcal{M}_r}(\mathbf{x})$.

[3] The distance decreases Q-linearly to zero if $\lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{x}\|/\|\mathbf{x}_{k+1} - \mathbf{x}\| = 0$, for some $\mathbf{x} \in \mathcal{A} \cap \mathcal{B}$.

Note that $\text{vec}(\mathbf{u}_1\mathbf{v}_1^\intercal)$ is the only vector of $\mathcal{K}(\mathbf{x})$ in $T_{\mathcal{M}_r}(\mathbf{x})$ such that any linear combination of at least two vectors in $\mathcal{K}(\mathbf{x})$ does not lie in $T_{\mathcal{M}_r}(\mathbf{x})$. In other words, $\text{Span}\{\mathcal{K}(\mathbf{x})\} \cap T_{\mathcal{M}_r}(\mathbf{x}) = T_{\mathcal{N}}(\mathbf{x}) \cap T_{\mathcal{M}_r}(\mathbf{x}) = \{\beta\text{vec}(\mathbf{u}_1\mathbf{v}_1^\intercal)\}$ for $\beta \in \mathbb{R}$. Yet, the dimension of $T_{\mathcal{M}_r}(\mathbf{x})$ is equal to $R - r + 1$ since $\text{rank}\{\mathbf{T}_{[r]}^\intercal \mathbf{T}_{[r]}^{\intercal+}\} = R - r + 1$. Hence,

$$\begin{aligned}\dim\{T_{\mathcal{N}}(\mathbf{x}) + T_{\mathcal{M}_r}(\mathbf{x})\} &= \dim\{T_{\mathcal{N}}(\mathbf{x})\} + \dim\{T_{\mathcal{M}_r}(\mathbf{x})\} - \dim\{T_{\mathcal{N}}(\mathbf{x}) \cap T_{\mathcal{M}_r}(\mathbf{x})\} \\ &= J + K + R - 1 - r.\end{aligned}$$

Thus, in order to ensure transversality between the manifolds, we should have $J + K + R - 1 - r \geq JK \iff r \leq R - (J-1)(K-1)$. ■

Although Proposition 4.9 ensures the transversality between the manifolds $\mathcal{M}_r$ and $\mathcal{N}$ and, thus, the linear convergence of our algorithm, the condition $r \leq R - (J-1)(K-1)$ is very restrictive. Indeed, to ensure the convergence of the CAPD algorithm when $r = R - 1$, we must have $R - 1 \leq R - (J-1)(K-1) \implies (J-1)(K-1) \leq 1$, which is true only when $J \leq 2, K \leq 2$. In other cases, Proposition 4.9 can only ensure, at best, the convergence for a few values of r. For some scenarios, Table 1 shows how many rank-1 components can be ensured by CAPD using the transversality concept.

(J, K)	R	# of rank-1 components
$(2, 2)$	2	2
$(3, 3)$	2	0
$(3, 3)$	3	0
$(5, 5)$	3	0
$(5, 5)$	20	4
$(6, 8)$	20	0
$(6, 8)$	40	5
$(10, 15)$	100	0
$(20, 20)$	100	0

Table 1: Number of estimated components ensured by CAPD under the transversality concept.

Notice that both the dimension I (omitted in the table) and the rank R must be much larger than J and K in general, in order to estimate some rank-1 components with convergence guarantees. Even for rank-100 tensors with dimensions $I \times 20 \times 20$, with $I \geq 100$ for ensuring the column-wise orthogonality of factor $\mathbf{A}$, the convergence to any component cannot be ensured by our algorithm. Thus, the table evinces the limitation of Proposition 4.9.

4.2. Convergence under non-tangentiality condition

Instead of considering transversality between manifolds, we can evaluate the convergence of CAPD under the non-tangential concept presented in Definition 4.6. According to [10], non-tangential points at the intersection of two manifolds is less restrictive than the transversality concept. Actually, the latter requires that the individual tangent spaces of the manifolds in a space must generate the whole space, which is not necessary for non-tangential points. The authors show that the convergence to a point at the intersection of two smooth manifolds can be ensured if the initialization of the alternating projection algorithm is close to a non-tangential point in the intersection. This result is stated in the theorem below.

Theorem 4.10. *[10] Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{A} \cap \mathcal{B}$ be C^2-manifolds, and let $\mathbf{x} \in \mathcal{A} \cap \mathcal{B}$ be a non tangential intersection point of $\mathcal{A}$ and $\mathcal{B}$. Given $\varepsilon > 0$ and $\cos(\alpha(\mathbf{x})) < c < 1$, where $\alpha(\mathbf{x})$ is the minimal angle between $\mathcal{A}$ and $\mathcal{B}$ at $\mathbf{x}$, there exists a $\rho > 0$ such that for any $\mathbf{x}^{(0)} \in \mathcal{B}_\rho(\mathbf{x})$, where $\mathcal{B}_\rho(\mathbf{x})$ is the open ball of radius ρ around $\mathbf{x}$, the sequence of alternating projections*

$$\mathbf{x}^{(k+1)} = \pi_{\mathcal{A}}\pi_{\mathcal{B}}(\mathbf{x}^{(k)}),\ (k = 0, 1, 2, \ldots)$$

(i) converges to a point $\mathbf{x}^\infty \in \mathcal{A} \cap \mathcal{B}$,
(ii) $\|\mathbf{x}^\infty - \pi_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x}^{(0)})\| \leq \varepsilon\|\mathbf{x}^{(0)} - \pi_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x}^{(0)})\|$,
(iii) $\|\mathbf{x}^\infty - \mathbf{x}^{(k)}\| \leq \text{cte} \cdot c^k\|\mathbf{x}^{(0)} - \pi_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x}^{(0)})\|$.

We are mainly interested in the result *(iii)*. Indeed, $\mathbf{x}^{(k)} \to \mathbf{x}^{\infty}$ as $k \to \infty$. In order to apply the result of [10] to our context, we must first prove under some conditions that there exists at least a non-tangential point $\mathbf{x}_r \in \mathcal{M}_r \cap \mathcal{N}$, for every $r \in \{1, 2, \ldots, R-1\}$. This is ensured thanks to the following proposition.

Proposition 4.11. *Let $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ be a rank-R tensor with factor $\mathbf{A}$ column-wise orthogonal. If $\mathcal{T}$ has an essentially unique decomposition then any point $\mathbf{x}_r \in \mathcal{M}_r \cap \mathcal{N}$, $r \in \{1, 2, \ldots, R-1\}$, is non-tangential.*

Proof. Since the decomposition is essentially unique, the matrix $\mathbf{C} \odot \mathbf{B}$ is full column rank, which implies that the set $\mathcal{M}_r \cap \mathcal{N}$ is composed of $R - r + 1$ straight lines passing through the origin (See Collorary 4.4).

Thus, it turns out that $\exists \rho > 0 : \mathcal{B}_\rho(\mathbf{x}_r) \cap \mathcal{M}_r \cap \mathcal{N} = \{\beta \mathbf{x}_r\},\ \forall \beta \in \mathbb{R}$ such that $|\beta| < \rho$. Therefore, we can conclude that $T_{\mathcal{M}_r \cap \mathcal{N}}(\mathbf{x}_r) = \{\gamma \mathbf{x}_r\},\ \forall \gamma \in \mathbb{R}$.

Yet, since $T_{\mathcal{M}_r}(\mathbf{x}_r) = \mathcal{M}_r$, the uniqueness of $\mathbf{x}_r \in \mathcal{M}_r \cap \mathcal{N}$ within $\mathcal{B}_\rho(\mathbf{x}_r)$ also ensures that $T_{\mathcal{N}}(\mathbf{x}_r) \cap T_{\mathcal{M}_r}(\mathbf{x}_r) = \{\gamma \mathbf{x}_r\},\ \forall \gamma \in \mathbb{R}$. Hence, $T_{\mathcal{M}_r}(\mathbf{x}_r) \cap T_{\mathcal{N}}(\mathbf{x}_r) = T_{\mathcal{M}_r \cap \mathcal{N}}(\mathbf{x}_r)$, at any point $\mathbf{x}_r \in \mathcal{M}_r \cap \mathcal{N}$, for every $r \in \{1, 2, \ldots, R-1\}$. ■

Proposition 4.11 brings up a strong result that ensures the convergence of CAPD algorithm. The only imposed constraint is the uniqueness of the decomposition, which is desired in practical applications.

5. Numerical Results

In this section, we evaluate the performance of the CAPD algorithm by numerical experiments. First, we show how the distance between the iterates $\mathbf{x}_r^{(k)}$ and $\mathbf{x}_r^{(k+1)}$ of the alternating projection part of our algorithm decreases as the number of iterations increases. Second, we will see that the convergence of a random tensor with a column-wise orthogonal factor matrix is log-linear. Finally, we corroborate the advantage of CAPD with respect to the alternating projections method that computes the factor matrices all-at-once.

5.1. Performances for real tensors

We consider a sample of 1000 real tensors with factor matrices generated as follows

- The entries of $\mathbf{B}$ and $\mathbf{C}$ are distributed according to a uniform measure in $[-1, 1]$;
- For $\mathbf{A}$, we first generate a random matrix with uniformly distributed entries in $[-1, 1]$, and then set $\mathbf{A}$ as the left singular matrix of that random matrix, ensuring the orthogonality of one factor.

Let $\Delta_k = \|\mathbf{x}_r^{(k+1)} - \mathbf{x}_r^{(k)}\|$ be the norm of the difference between two successive iterates in the CAPD algorithm for some $1 \leq r \leq R-1$ at iteration $k \geq 0$. Since we have 1000 tensors, we replace $\mathbb{E}\{\Delta_k\}$ by the average value at iteration k. We assume the tensors have rank 4, so that we evaluate the algorithm by estimating 3 rank-1 components (the last component is obtained directly by means of deflation, as shown in Algorithm 1) for four scenarios.

According to Figure 1, in scenarios with K=3 the average distance between iterates decreases as the rank of the tensor is reduced by deflation. For the $20 \times 20 \times 20$ scenario, the same phenomenon is noted for $10^{-2} < \mathbb{E}\{\Delta_k\} < 10^{-6}$, which covers most of the range of the other scenarios. We also note from the last scenario that for tensors with all dimensions much larger than the rank, the CAPD algorithm converges in a few iterations. Indeed, the iterates converge to zero[4] in 60, 55, and 40 iterations approximately, for r equal to 1, 2 and 3, respectively.

Now, we draw randomly one tensor to show how the iterative procedure converges as k increases. We choose the $5 \times 4 \times 3$ scenario. In Figure 2, we see that the distance between iterates of the CPAD algorithm converges log-linearly to zero. Although we do not show other examples here, this behaviour can be noted for most of the tensors of our sample. Moreover, the log-linear convergence also depends on the starting point $\mathbf{x}_r^{(0)}$.

Contrary to the averaged result, we note that $\|\mathbf{x}_r^{(k+1)} - \mathbf{x}_r^{(k)}\|$, for some values of k, does not decrease when the rank decreases. Indeed, the convergence is slower when the rank of the tensor is only 2 ($r = 3$).

[4]It is the virtual zero or the floating-point relative accuracy of Matlab, whose value is 2^{-52}.

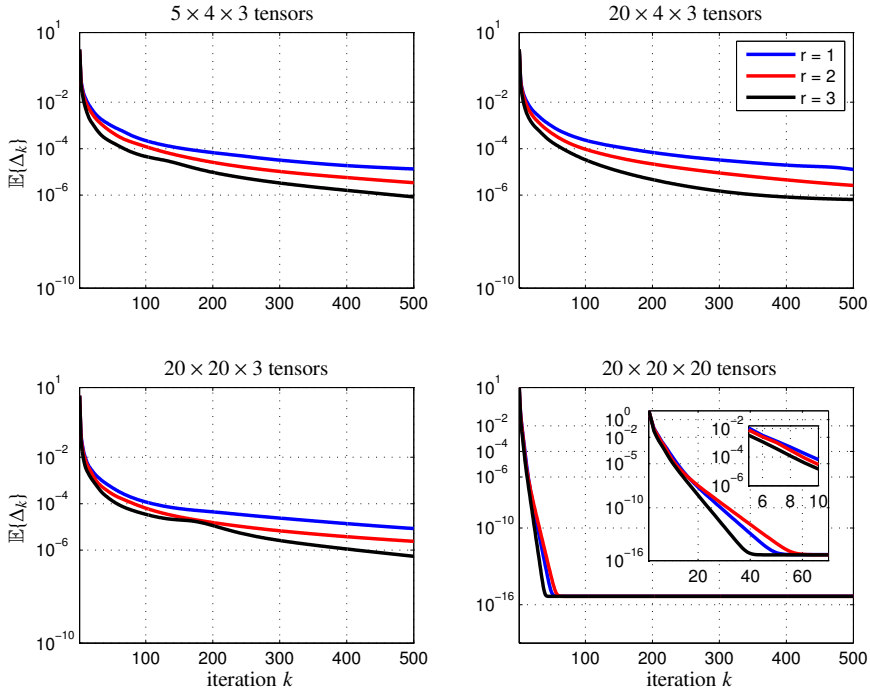


Figure 1: Convergence of the iterates of CAPD algorithm for rank-4 tensors.

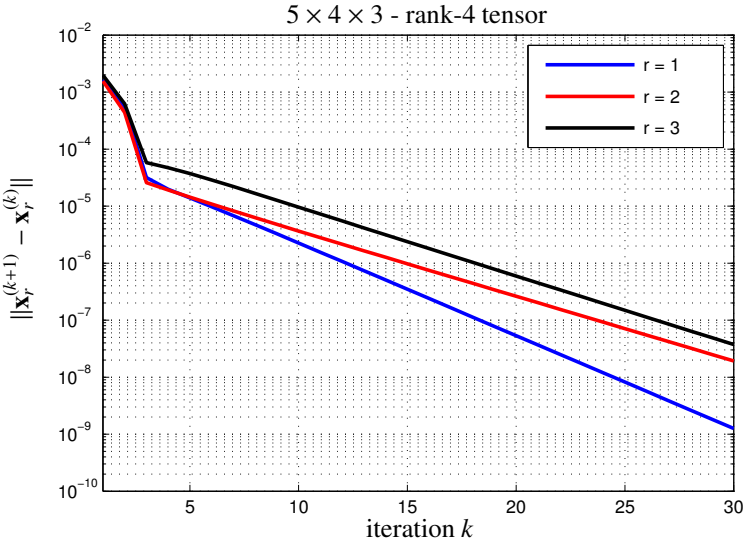


Figure 2: Convergence of the iterates in CAPD algorithm for a single rank-4 tensor.

5.2. Performance for complex tensors

Herein, we evaluate the performance of CAPD in terms of the computational time and the percentage of successful decompositions denoted by $\%p$. We assume that CAPD delivers a correct decomposition for a tensor $\mathcal{T}$ if $\|\mathcal{T} - \widehat{\mathcal{T}}\| \leq 10^{-6}$, where $\widehat{\mathcal{T}}$ is the estimated tensor. We generate a sample of 500 complex tensors. The real and imaginary parts are uniformly distributed in $[-1, 1]$, and the column-wise factor $\mathbf{A}$ is the left singular matrix of a random complex matrix with uniform entries generated as before. We simulate the same four scenarios as in the previous subsection.

Table below summarizes the results.

Scenario	Average time (seconds)	$\%p$
$5 \times 4 \times 3$	0.7231	99.6
$20 \times 4 \times 3$	0.7131	99
$20 \times 20 \times 3$	0.7942	100
$20 \times 20 \times 20$	0.2723	100

Table 2: Time and percentage of correct CAPDs for rank-4 tensors.

Notice that the average time of simulations for $20\times20\times20$ tensors corroborates with the convergence performance depicted in Figure 1, where we have shown that CAPD algorithm converges very fast when the dimensions are much larger than the rank. Here, it results in a small computational time compared to the other scenarios.

5.3. Simultaneous estimation of factors

As mentioned in Remark 2, the all-at-once estimation of the factor matrices might be performed via an alternating projection algorithm based on equation (2). Instead of initializing the vector $\mathbf{c}_r^{(0)} \boxtimes \mathbf{b}_r^{(0)}$, we could initialize directly a Khatri-Rao matrix $\mathbf{C}^{(0)} \odot \mathbf{B}^{(0)}$, and thus we would have the following alternating projection problem

$$\mathbf{X}^{(k+1)} = \pi_{\mathcal{M}}\pi_{\mathcal{N}_R}(\mathbf{X}^{(k)}),\ (k = 0, 1, 2, \ldots),$$

where

$$\mathcal{M} = \left\{\mathbf{X} \in \mathbb{K}^{JK\times R} : \left(\mathbf{T}^{\mathsf{H}}\mathbf{T}^{\mathsf{H}+} - \mathbf{I}\right)\mathbf{X} = \mathbf{0}\right\},$$

and

$$\mathcal{N}_R = \{\mathbf{X} \in \mathbb{K}^{JK\times R} : \mathbf{X} = \mathbf{C} \odot \mathbf{B}, \|\mathbf{X}\| = 1, \forall \mathbf{B} \in \mathbb{K}^{J\times R} \setminus \{\mathbf{0}\} \text{ and } \forall \mathbf{C} \in \mathbb{K}^{K\times R} \setminus \{\mathbf{0}\}\ \}.$$

If the algorithm converged to a point in $\mathcal{M} \cap \mathcal{N}_R$ so that the right factors $\mathbf{B}$ and $\mathbf{C}$ could be properly estimated, then the factor $\mathbf{A}$ would be estimated by solving the matrix linear equation (2), whose the least square solution is $\mathbf{A} = \mathbf{T}^{\mathsf{H}+}(\mathbf{C} \odot \mathbf{B})$.

This approach presents two crucial drawbacks:

- The complexity of this *all-at-once* alternating projections algorithm is larger than that of the CAPD algorithm. Indeed, the employment of that method to solve equation (2) is equivalent to employ it to solve R equations (3) simultaneously. Notice, however, that CAPD solves only $R - 1$ times the equation (3) .
- There is no guarantee that the estimated columns of the factors $\widehat{\mathbf{A}}$, $\widehat{\mathbf{B}}$, and $\widehat{\mathbf{C}}$ be the same of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. As the initialization $\mathbf{C}^{(0)} \odot \mathbf{B}^{(0)}$ is actually R rank-1 initializations $\mathbf{c}_r^{(0)} \boxtimes \mathbf{b}_r^{(0)}$ (of R parallel problems), some of those initializations can converge to the same point, which means that the estimated matrices might have collinear columns. Indeed, this happens more often that one can imagine.

In order to show the limitation of computing the factors all-at-once using the alternating projection method, we count how many column factors are properly estimated for the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. We evaluate the following real and complex scenarios for a sample of 100 tensors: $15 \times 10 \times 5$, $20 \times 10 \times 5$, $20 \times 20 \times 5$ and $20 \times 20 \times 20$ tensors. For each tensor, we perform a random initialization.

Figure 3 shows that the all-at-once estimation of all factors using the alternating projection method happens in a very few cases. In fact, the best scenario is the one with $20 \times 20 \times 5$ real tensors, in which only for 7 tensors were sufficient to extract completely the factor matrices. In all scenarios, the algorithm recovers 3 or 4 column factors of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ in general. To increase the chances of estimating the right columns, we could test more initializations. However we would run again the algorithm, increasing though the number of operations to be computed. Therefore, CAPD is more efficient than the all-at-once alternating projection computation.

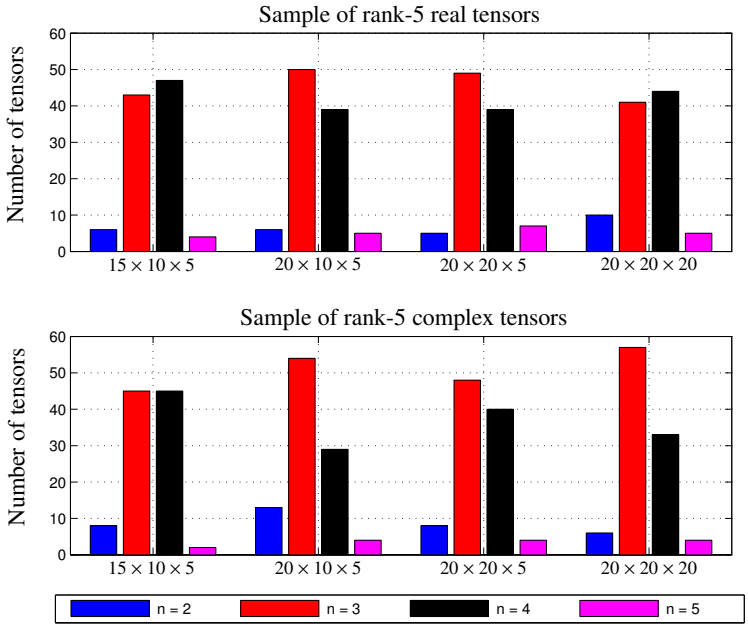


Figure 3: Tensors for which n column factors are correctly extracted with the *all-at-once* algorithm.

6. Conclusion

Herein, we have introduced a new algorithm, called CAPD, to compute the exact column-wise orthogonal CP decomposition. We showed that our algorithm extracts the exact components of the factor matrices when the iterates converge to a point at the intersection of two manifolds. We also applied the transversality and the non-tangential concepts to the column-wise orthogonal tensor context to prove local convergence. Finally, by means of numerical experiments, we showed the performance of CAPD in terms of convergence rate, and the advantage regarding the all-at-once estimation of factor matrices with alternating projections.

Acknowledgements

This work has been funded by the European Research Council under the 7th Framework Program FP7/2007-2013 Grant Agreement no. 320594, and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) Grant Agreement no. 237778/2012-2.

References

[1] J. Von Neumann, Functional operators: The geometry of orthogonal spaces, Princeton University Press, 1950.
[2] P. Combettes, H. J. Trussell, Method of successive projections for finding a common point of sets in metric spaces, Journal of optimization theory and applications 67 (3) (1990) 487–507.
[3] N. J. Higham, Computing the nearest correlation matrix - a problem from finance, IMA journal of Numerical Analysis 22 (3) (2002) 329–343.
[4] L. Gubin, B. Polyak, E. Raik, The method of projections for finding the common point of convex sets, USSR Computational Mathematics and Mathematical Physics 7 (6) (1967) 1–24.
[5] L. Breg, Finding the common point of convex sets by the method of sucessive projections, Dokl. Akad. Mousk SSSR (1965) 487–490.
[6] K. M. Grigoriadis, E. B. Beran, Alternating projection algorithms for linear matrix inequalities problems with rank constraints, Society for Industrial and Applied Mathematics, 1999.
[7] H. H. Bauschke, P. L. Combettes, D. R. Luke, Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization, JOSA A 19 (7) (2002) 1334–1345.
[8] W. I. Zangwill, Non-linear programming via penalty functions, Management science 13 (5) (1967) 344–358.
[9] A. S. Lewis, J. Malick, Alternating projections on manifolds, Mathematics of Operations Research 33 (1) (2008) 216–234.
[10] F. Andersson, M. Carlsson, Alternating projections on nontangential manifolds, Constructive Approximation 38 (3) (2013) 489–525.
[11] T. G. Kolda, Orthogonal tensor decompositions, SIAM Journal on Matrix Analysis and Applications 23 (1) (2001) 243–255.

[12] L. De Lathauwer, J. Castaing, J.-F. Cardoso, Fourth-order cumulant-based blind identification of underdetermined mixtures, Signal Processing, IEEE Transactions on 55 (6) (2007) 2965–2973.
[13] S. Miron, X. Guo, D. Brie, Doa estimation for polarized sources on a vector-sensor array by parafac decomposition of the fourth-order covariance tensor, in: Signal Processing Conference, 2008 16th European, IEEE, 2008, pp. 1–5.
[14] M. Sørensen, L. De Lathauwer, L. Deneire, Parafac with orthogonality in one mode and applications in ds-cdma systems, in: Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, IEEE, 2010, pp. 4142–4145.
[15] L. Wang, M. T. Chu, B. Yu, Orthogonal low rank tensor approximation: Alternating least squares method and its global convergence, SIAM Journal on Matrix Analysis and Applications 36 (1) (2015) 1–19.
[16] M. Sørensen, L. D. Lathauwer, P. Comon, S. Icart, L. Deneire, Canonical polyadic decomposition with a columnwise orthonormal factor matrix, SIAM Journal on Matrix Analysis and Applications 33 (4) (2012) 1190–1213.
[17] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, SIAM review 51 (3) (2009) 455–500.
[18] P. Comon, Tensors: a brief introduction, IEEE Sig. Proc. Magazine 31 (3) (2014) 44–53, hal-00923279.