Convergence of the CP-AltLS algorithm for orthogonally and incoherently decomposable tensors

Nicholas Hu

The University of California, Los Angeles (UCLA)

April 14, 2026

Introduction

Introduction

Tensors are multidimensional arrays that can be used to represent and analyze data, with the different dimensions/modes corresponding to different data components (e.g., position, time, frequency, intensity, object, type).

$$\mathcal{X} = \left[\begin{bmatrix} x_{111} & x_{121} \\ x_{211} & x_{221} \\ x_{311} & x_{321} \\ x_{411} & x_{421} \end{bmatrix} \quad \begin{bmatrix} x_{112} & x_{122} \\ x_{212} & x_{222} \\ x_{312} & x_{322} \\ x_{412} & x_{422} \end{bmatrix} \quad \begin{bmatrix} x_{113} & x_{123} \\ x_{213} & x_{223} \\ x_{313} & x_{323} \\ x_{413} & x_{423} \end{bmatrix} \right] \in \mathbb{R}^{4 \times 2 \times 3}$$

Vectors (first-order tensors) will be represented by *bold lowercase* letters (e.g., $\mathbf{a}$), matrices (second-order tensors) by *bold uppercase* letters (e.g., $\mathbf{A}$), and general tensors by *bold uppercase script* letters (e.g., $\mathcal{X}$).

The $(i_1, \ldots, i_N)$-entry of $\mathcal{X}$ will be denoted by $x_{i_1 \cdots i_N}$ and the j^{th} column of $\mathbf{A}$ by $\mathbf{a}_j$.

Introduction

Tensor decompositions are structured representations of tensors that render them more amenable to storage, manipulation, and/or analysis.

The **CP decomposition** expresses a tensor as a sum of simpler component tensors, which can reveal patterns in or features of the underlying data (e.g., for a tensor whose entries are excitation-emission intensities of chemical samples, each component tensor corresponds to a chemical compound in the samples).

$$\mathcal{X} = \mathcal{X}_1 + \mathcal{X}_2 + \cdots + \mathcal{X}_R$$

Introduction: The CP decomposition

Recall that the eigendecomposition of a symmetric matrix $\mathbf{A} \in \mathbb{R}^{I \times I}$ is

$$\mathbf{A} = \sum_{i=1}^{I} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top,$$

where $\lambda_i \in \mathbb{R}$ and $\mathbf{U} \in \mathbb{R}^{I \times I}$ is *orthogonal*.

Recall that the singular value decomposition of a matrix $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$ is

$$\mathbf{A} = \sum_{i=1}^{\min\{I_1, I_2\}} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top,$$

where $\sigma_i \in \mathbb{R}_{\geq 0}$, and $\mathbf{U} \in \mathbb{R}^{I_1 \times \min\{I_1, I_2\}}$ and $\mathbf{V} \in \mathbb{R}^{I_2 \times \min\{I_1, I_2\}}$ have *orthonormal* columns.

Introduction: The CP decomposition

Definition (Outer product)

The **outer product** of $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ and $\mathcal{Y} \in \mathbb{R}^{J_1 \times \cdots \times J_M}$ is the tensor $\mathcal{X} \circ \mathcal{Y} \in \mathbb{R}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$ with entries

$$(\mathcal{X} \circ \mathcal{Y})_{i_1 \cdots i_N j_1 \cdots j_M} := x_{i_1 \cdots i_N} y_{j_1 \cdots j_M}.$$

Example $\mathbf{u} \in \mathbb{R}^I$, $\mathbf{v} \in \mathbb{R}^J$

$$\mathbf{u} \circ \mathbf{v} = \begin{bmatrix} u_1 \\ \vdots \\ u_I \end{bmatrix} \circ \begin{bmatrix} v_1 \\ \vdots \\ v_J \end{bmatrix} = \begin{bmatrix} u_1 v_1 & \cdots & u_1 v_J \\ \vdots & \ddots & \vdots \\ u_I v_1 & \cdots & u_I v_J \end{bmatrix} (= \mathbf{u}\mathbf{v}^\top) \in \mathbb{R}^{I \times J}$$

Introduction: The CP decomposition

Definition (CP decomposition and rank)

A **CP decomposition** of $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is a decomposition of the form

$$\mathcal{X} = [\![\boldsymbol{\lambda};\, \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] := \sum_{r=1}^{R} \lambda_r \mathbf{a}_r^{(1)} \circ \cdots \circ \mathbf{a}_r^{(N)},$$

where $\boldsymbol{\lambda} \in \mathbb{R}^R$ and $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ has *unit-norm* (but not necessarily orthonormal) columns. The λ_r are called **weights** and the $\mathbf{A}^{(n)}$ are called **factor matrices**.

We say that $\mathcal{X}$ has **(CP) rank** R if R is *minimal* (note that every tensor has a CP decomposition with $R = \prod_n I_n$.)

Tensor rank agrees with matrix rank for second-order tensors (matrices) by the SVD/rank factorization.

However, tensor rank behaves very differently from matrix rank for higher-order tensors! For instance, we can have $\operatorname{rank}(\mathcal{X}) > \min_n I_n$.

Example

$$\mathcal{X} = \left[\overbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}^{\mathbf{X}^{(1)}} \; \overbrace{\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}}^{\mathbf{X}^{(2)}} \right] \in \mathbb{R}^{2\times 2\times 2}$$

If $\mathcal{X} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \mathbf{a}_2 \circ \mathbf{b}_2 \circ \mathbf{c}_2$, then $\mathbf{X}^{(i)} = \mathbf{A} \begin{bmatrix} c_{i1} & 0 \\ 0 & c_{i2} \end{bmatrix} \mathbf{B}^\top$.

In particular, since $\mathbf{X}^{(1)}$ is invertible, we would have $\mathbf{X}^{(2)}(\mathbf{X}^{(1)})^{-1} = \mathbf{A} \begin{bmatrix} c_{21}/c_{11} & 0 \\ 0 & c_{22}/c_{12} \end{bmatrix} \mathbf{A}^{-1}$. But $\mathbf{X}^{(2)}(\mathbf{X}^{(1)})^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ is not diagonalizable, so $\operatorname{rank}(\mathcal{X}) > 2$.

Introduction: The CP decomposition

- The rank of a tensor over $\mathbb{R}$ may be greater than its rank over $\mathbb{C}$ even when its entries are real, as the previous example shows (Kruskal 1983).
- Determining whether the rank of a tensor is at most r is NP-complete in any finite field and NP-hard in any field $\mathbb{F} \supseteq \mathbb{Q}$, even for third-order tensors (Håstad 1990; Hillar and Lim 2013).
  [What is the rank (over $\mathbb{R}$) of the tensor in the previous example?]
  - In practice, the rank of a tensor is estimated by fitting CP decompositions (with varying R) to the tensor.
  - The best rank-r approximation of a tensor (in the Frobenius norm, for instance) may not exist (Paatero 2000).
- The maximum rank of a tensor with given dimensions is unknown in general.
  [What is the maximum rank (over $\mathbb{R}$) of a $2 \times 2 \times 2$ (real) tensor?]

Introduction: The CP decomposition

$$\mathcal{X} = [\![\boldsymbol{\lambda}\,;\, \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \in \mathbb{R}^{I_1 \times \cdots \times I_N} \quad (\boldsymbol{\lambda} \in \mathbb{R}^R, \mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R})$$

Definition (Orthogonally decomposable tensor)

A tensor is **orthogonally decomposable** if it admits a CP decomposition whose factor matrices have *orthonormal* columns.

Such a decomposition must have $R \leq \min_n I_n$, so not all tensors are odeco!

Definition (Incoherently decomposable tensor)

The **coherence** of a matrix $\mathbf{A}$ with *unit-norm* columns is $\mu(\mathbf{A}) := \max_{i \neq j} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle| \in [0, 1]$. A tensor is **$\mu$-coherently decomposable** if it admits a CP decomposition whose factor matrices have *coherence at most μ*.

Introduction: The CP-AltLS algorithm

$$\mathcal{X} = [\![\boldsymbol{\lambda}\,;\, \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}]\!] \in \mathbb{R}^{I_1 \times \cdots \times I_N} \quad (\boldsymbol{\lambda} \in \mathbb{R}^R, \mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R})$$

Given $\mathcal{X}$ and initial guesses for the $\mathbf{A}^{(n)}$, the **CP-alternating least squares** algorithm approximates each factor matrix successively by solving a least squares problem in which all other factor matrices are fixed:

$$\mathbf{A}^{(n)} = \operatorname*{argmin}_{\mathbf{A}} \left\| \mathcal{X} - [\![\boldsymbol{\lambda}\,;\, \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(n-1)}, \mathbf{A}, \mathbf{A}^{(n+1)}, \dots, \mathbf{A}^{(N)}]\!] \right\|.$$

Definition (Inner product and norm of tensors)

The **(Frobenius) inner product** of tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is $\langle \mathcal{X}, \mathcal{Y} \rangle := \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} x_{i_1 \cdots i_N} y_{i_1 \cdots i_N}$ and the **(Frobenius) norm** of $\mathcal{X}$ is $\|\mathcal{X}\| := \langle \mathcal{X}, \mathcal{X} \rangle^{\frac{1}{2}}$.

Introduction: The CP-AltLS algorithm

To solve these tensor least squares problems, we “reshape” the tensors into matrices.

Example

$$\mathcal{X} = \left[\begin{bmatrix} x_{111} & x_{121} \\ x_{211} & x_{221} \\ x_{311} & x_{321} \\ x_{411} & x_{421} \end{bmatrix} \begin{bmatrix} x_{112} & x_{122} \\ x_{212} & x_{222} \\ x_{312} & x_{322} \\ x_{412} & x_{422} \end{bmatrix} \begin{bmatrix} x_{113} & x_{123} \\ x_{213} & x_{223} \\ x_{313} & x_{323} \\ x_{413} & x_{423} \end{bmatrix}\right] \in \mathbb{R}^{4\times 2\times 3}$$

$$\mathbf{X}_{(1)} = \begin{bmatrix} \mathbf{x}_{:,1,1} & \mathbf{x}_{:,2,1} & \cdots & \mathbf{x}_{:,2,3} \end{bmatrix} = \begin{bmatrix} x_{111} & x_{121} & \cdots & x_{123} \\ x_{211} & x_{221} & \cdots & x_{223} \\ x_{311} & x_{321} & \cdots & x_{323} \\ x_{411} & x_{421} & \cdots & x_{423} \end{bmatrix} \in \mathbb{R}^{4\times 6}$$

$$\mathbf{X}_{(2)} = \begin{bmatrix} \mathbf{x}_{1,:,1} & \mathbf{x}_{2,:,1} & \cdots & \mathbf{x}_{4,:,3} \end{bmatrix} = \begin{bmatrix} x_{111} & x_{211} & \cdots & x_{413} \\ x_{121} & x_{221} & \cdots & x_{423} \end{bmatrix} \in \mathbb{R}^{2\times 12}$$

Introduction: The CP-AltLS algorithm

The matrix $\mathbf{X}_{(n)}$ is called the **mode-n matricization** of $\mathcal{X}$.

Example $\mathcal{X} = [\![\boldsymbol{\lambda}\,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \in \mathbb{R}^{4\times 2\times 3}$

$$x_{ijk} = \sum_{r=1}^{R} \lambda_r a_{ir} b_{jr} c_{kr}$$

$$\mathbf{x}_{:,j,k} = \sum_{r=1}^{R} \lambda_r \mathbf{a}_r b_{jr} c_{kr} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_R \end{bmatrix} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_R \end{bmatrix}}_{\boldsymbol{\Lambda}} \begin{bmatrix} b_{j1} c_{k1} \\ \vdots \\ b_{jR} c_{kR} \end{bmatrix}$$

$$\mathbf{X}_{(1)} = \mathbf{A}\boldsymbol{\Lambda} \begin{bmatrix} c_{11}\mathbf{b}_1^\top & c_{21}\mathbf{b}_1^\top & c_{31}\mathbf{b}_1^\top \\ \vdots & \vdots & \vdots \\ c_{1R}\mathbf{b}_R^\top & c_{2R}\mathbf{b}_R^\top & c_{3R}\mathbf{b}_R^\top \end{bmatrix} = \mathbf{A}\boldsymbol{\Lambda} \begin{bmatrix} (\mathbf{c}_1 \otimes \mathbf{b}_1)^\top \\ \vdots \\ (\mathbf{c}_R \otimes \mathbf{b}_R)^\top \end{bmatrix}.$$

Introduction: The CP-AltLS algorithm

Definition (Khatri–Rao product)

The **Khatri–Rao product** of $\mathbf{A} \in \mathbb{R}^{I \times K}$ and $\mathbf{B} \in \mathbb{R}^{J \times K}$ is the matrix $\mathbf{A} \odot \mathbf{B} \in \mathbb{R}^{IJ \times K}$ given by

$$\mathbf{A} \odot \mathbf{B} := \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \cdots & \mathbf{a}_K \otimes \mathbf{b}_K \end{bmatrix} = \begin{bmatrix} a_{11}\mathbf{b}_1 & \cdots & a_{1K}\mathbf{b}_K \\ \vdots & \ddots & \vdots \\ a_{I1}\mathbf{b}_1 & \cdots & a_{IK}\mathbf{b}_K \end{bmatrix}.$$

<u>Example</u> $\mathcal{X} = [\![\boldsymbol{\lambda};\, \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!] \in \mathbb{R}^{I_1 \times \cdots \times I_N}$

$$\boxed{\mathbf{X}_{(n)} = \mathbf{A}^{(n)} \boldsymbol{\Lambda} (\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \cdots \odot \mathbf{A}^{(1)})^\top}$$

Introduction: The CP-AltLS algorithm

Example $\mathbf{A} \in \mathbb{R}^{I \times K}$, $\mathbf{B} \in \mathbb{R}^{J \times K}$

$$(\mathbf{A} \odot \mathbf{B})^\top (\mathbf{A} \odot \mathbf{B}) = \begin{bmatrix} \mathbf{a}_1^\top \otimes \mathbf{b}_1^\top \\ \vdots \\ \mathbf{a}_K^\top \otimes \mathbf{b}_K^\top \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \cdots & \mathbf{a}_K \otimes \mathbf{b}_K \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{a}_1^\top \mathbf{a}_1 \otimes \mathbf{b}_1^\top \mathbf{b}_1 & \cdots & \mathbf{a}_1^\top \mathbf{a}_K \otimes \mathbf{b}_1^\top \mathbf{b}_K \\ \vdots & \ddots & \vdots \\ \mathbf{a}_K^\top \mathbf{a}_1 \otimes \mathbf{b}_K^\top \mathbf{b}_1 & \cdots & \mathbf{a}_K^\top \mathbf{a}_K \otimes \mathbf{b}_K^\top \mathbf{b}_K \end{bmatrix}$$

$$= (\mathbf{A}^\top \mathbf{A}) * (\mathbf{B}^\top \mathbf{B})$$

Definition (Hadamard product)

The **Hadamard product** of $\mathbf{A} \in \mathbb{R}^{I \times J}$ and $\mathbf{B} \in \mathbb{R}^{I \times J}$ is the matrix $\mathbf{A} * \mathbf{B} \in \mathbb{R}^{I \times J}$ with entries $(\mathbf{A} * \mathbf{B})_{ij} := a_{ij} b_{ij}$.

Introduction: The CP-AltLS algorithm

Returning to the least squares problem,

$$\begin{aligned}
\mathbf{A}^{(n)} &= \operatorname*{argmin}_{\mathbf{A}} \left\| \mathcal{X} - [\![\boldsymbol{\lambda}\,;\, \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(n-1)}, \mathbf{A}, \mathbf{A}^{(n+1)}, \dots, \mathbf{A}^{(N)}]\!] \right\| \\
&= \operatorname*{argmin}_{\mathbf{A}} \left\| \mathbf{X}_{(n)} - \mathbf{A}\boldsymbol{\Lambda}(\mathbf{K}^{(n)})^\top \right\|,
\end{aligned}$$

where $\mathbf{K}^{(n)} := \bigodot_{m \neq n} \mathbf{A}^{(m)}$. Hence

$$\begin{aligned}
\mathbf{A}^{(n)}\boldsymbol{\Lambda} &= \mathbf{X}_{(n)}((\mathbf{K}^{(n)})^\top)^+ \\
&= \mathbf{X}_{(n)}\mathbf{K}^{(n)}((\mathbf{K}^{(n)})^\top \mathbf{K}^{(n)})^+ \\
&= \mathbf{X}_{(n)}\mathbf{K}^{(n)}(\mathbf{H}^{(n)})^+,
\end{aligned}$$

where $\mathbf{H}^{(n)} := \ast_{m \neq n} \mathbf{G}^{(m)}$ and $\mathbf{G}^{(m)} := (\mathbf{A}^{(m)})^\top(\mathbf{A}^{(m)})$.

After solving for $\mathbf{A}^{(n)}\boldsymbol{\Lambda}$, we normalize its columns to obtain $\mathbf{A}^{(n)}$ and update $\boldsymbol{\lambda}$ accordingly.

Algorithm: CP-AltLS

Input: $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ $(n \in [N])$

for $n = 1$ **to** N

  $\mathbf{G}^{(n)} \leftarrow (\mathbf{A}^{(n)})^\top \mathbf{A}^{(n)}$

while *stopping condition has not been satisfied*

  for $n = 1$ **to** N

    $\mathbf{K}^{(n)} \leftarrow \bigodot_{m \neq n} \mathbf{A}^{(m)}$

    $\mathbf{H}^{(n)} \leftarrow \ast_{m \neq n} \mathbf{G}^{(m)}$

    $\mathbf{A}^{(n)} \leftarrow \mathbf{X}_{(n)} \mathbf{K}^{(n)} (\mathbf{H}^{(n)})^+$

    normalize columns of $\mathbf{A}^{(n)}$, updating $\boldsymbol{\lambda}$ accordingly

    $\mathbf{G}^{(n)} \leftarrow (\mathbf{A}^{(n)})^\top \mathbf{A}^{(n)}$

Output: $\boldsymbol{\lambda} \in \mathbb{R}^R$, $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ $(n \in [N])$

Introduction: The CP-AltLS algorithm

- Existing results on CP-AltLS convergence are *limited in applicability* (e.g., to rank-one (Wang and Chu 2014) or orthogonal (Wang, Chu, and Yu 2015) decompositions), *qualitative*, *non-explicit*, and/or otherwise imprecise.
- Uschmajew proved local linear convergence to local minima of the (reconstruction) error satisfying a nondegeneracy condition, and established a quantitative sufficient condition for nondegeneracy in the rank-one case (Uschmajew 2012).
- Other methods for computing CP decompositions (e.g., the higher-order power method) have been studied, also under rank-one (Hu and Li 2018) or orthogonality assumptions (Hu and Ye 2023).

Introduction: The CP-AltLS algorithm

Our result (· , Iwen, Needell, and Wang 2025):

- arbitrary rank
- orthogonal and incoherent decompositions
- quantitative and explicit
- specific order of convergence
- more direct and less technical proof

Convergence of CP-AltLS

Convergence of CP-AltLS

We consider the "angular error" in the approximation of the exact factor matrices $\mathbf{A}^{(n)}$ by the approximate factor matrices $\mathbf{A}^{(n,k)}$ after k iterations:

$$\varepsilon_k := \max_{n\in[N],\, r\in[R]} \left|\sin \angle(\mathbf{a}_r^{(n,k)}, \mathbf{a}_r^{(n)})\right|.$$

Iteration numbers will be appended to the superscripts of vectors and matrices produced by the algorithm (e.g., $\boldsymbol{\lambda}^{(k)}$ will denote $\boldsymbol{\lambda}$ after k iterations, $\mathbf{A}^{(n,k)}$ will denote $\mathbf{A}^{(n)}$ after k iterations).

The weights $\boldsymbol{\lambda}^{(k)}$ will also converge *up to signs*, though we will not discuss this here.

Convergence of CP-AltLS

We will begin by analyzing a simplified version of CP-AltLS.

Algorithm: "Decoupled" CP-AltLS

Input: $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ $(n \in [N])$

while *stopping condition has not been satisfied*

- **for** $n = 1$ **to** N
  - $\mathbf{G}^{(n)} \leftarrow (\mathbf{A}^{(n)})^\top \mathbf{A}^{(n)}$
  - $\mathbf{K}^{(n)} \leftarrow \bigodot_{m \neq n} \mathbf{A}^{(m)}$
- **for** $n = 1$ **to** N
  - $\mathbf{H}^{(n)} \leftarrow \ast_{m \neq n} \mathbf{G}^{(m)}$
  - $\mathbf{A}^{(n)} \leftarrow \mathbf{X}_{(n)} \mathbf{K}^{(n)} (\mathbf{H}^{(n)})^+$
  - normalize columns of $\mathbf{A}^{(n)}$, updating $\boldsymbol{\lambda}$ accordingly

Output: $\boldsymbol{\lambda} \in \mathbb{R}^R$, $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ $(n \in [N])$

> $\mathbf{A}^{(n,k)}$ depends only on $\mathbf{A}^{(1,k-1)}, \ldots, \mathbf{A}^{(n-1,k-1)}, \mathbf{A}^{(n+1,k-1)}, \ldots, \mathbf{A}^{(N,k-1)}$
> instead of
> $\mathbf{A}^{(1,k)}, \ldots, \mathbf{A}^{(n-1,k)}, \mathbf{A}^{(n+1,k-1)}, \ldots, \mathbf{A}^{(N,k-1)}$

Convergence of CP-AltLS: Odeco tensors

Theorem (Local convergence of CP-AltLS for odeco tensors)

Suppose that $\mathcal{X} = [\![\boldsymbol{\lambda}\,;\, \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!]$ ($N \geq 3$) *for some* $\boldsymbol{\lambda} \in \mathbb{R}^R$ *and some* $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ *with* <u>*orthonormal columns*</u>*. In addition, let* $\kappa := \frac{\max_{r \in [R]} |\lambda_r|}{\min_{r \in [R]} |\lambda_r|}$.

Then there exists an absolute constant $C > 0$ *such that if* $\varepsilon_0 < C\kappa^{-\frac{1}{N-2}} R^{-\frac{1}{N-1}}$, *then* $\varepsilon_k \to 0$ *with* <u>*order* $N-1$</u> (*more precisely,* $\varepsilon_k \leq \alpha^{(N-1)^k}$ *for some absolute constant* $0 < \alpha < 1$).

Convergence of CP-AltLS: Odeco tensors

Proof (outline)

Consider

$$\mathbf{A}^{(n,k)} = \mathbf{X}_{(n)}\mathbf{K}^{(n,k)}(\mathbf{H}^{(n,k)})^{+} = \mathbf{A}^{(n)}\mathbf{\Lambda}(\mathbf{K}^{(n)})^{\top}\mathbf{K}^{(n,k)}(\mathbf{H}^{(n,k)})^{+}.$$

WLOG assume that $\langle \mathbf{a}_r^{(n,k)}, \mathbf{a}_r^{(n)} \rangle \geq 0$ for all k, n, r by adjusting $\boldsymbol{\lambda}$.

- $\varepsilon_{k-1} \ll 1 \implies \mathbf{A}^{(m,k-1)} \approx \mathbf{A}^{(m)}\ (\forall m)$
- $\mathbf{G}^{(m,k)} = (\mathbf{A}^{(m,k-1)})^{\top}(\mathbf{A}^{(m,k-1)}) \approx (\mathbf{A}^{(m)})^{\top}\mathbf{A}^{(m)} = \mathbf{I}$
- $(\mathbf{K}^{(n)})^{\top}\mathbf{K}^{(n,k)} = \circledast_{m\neq n}(\mathbf{A}^{(m)})^{\top}\mathbf{A}^{(m,k-1)} \approx \circledast_{m\neq n}\mathbf{I} = \mathbf{I}$
- $\mathbf{H}^{(n,k)} = \circledast_{m\neq n}\mathbf{G}^{(m,k)} \approx \circledast_{m\neq n}\mathbf{I} = \mathbf{I}$
- $\mathbf{A}^{(n,k)} = \mathbf{A}^{(n)}\mathbf{\Lambda}(\mathbf{K}^{(n)})^{\top}\mathbf{K}^{(n,k)}(\mathbf{H}^{(n,k)})^{+} \approx \mathbf{A}^{(n)}\mathbf{\Lambda}\,\mathbf{I}\,\mathbf{I} = \mathbf{A}^{(n)}\mathbf{\Lambda}$
- $\mathbf{A}^{(n,k)} \approx \mathbf{A}^{(n)}\mathbf{\Lambda}\ (\forall n) \implies \varepsilon_k \ll 1$

Convergence of CP-AltLS: Odeco tensors

Proof (sketch)

Let us write $\mathbf{A} = [\mathbf{D}, \mathbf{D}']$ to indicate that $\mathbf{D}$ and $\mathbf{D}'$ are the diagonal and off-diagonal parts of $\mathbf{A}$.

If $\varepsilon_{k-1} = \max_{n,r} |\sin \angle(\mathbf{a}_r^{(n,k-1)}, \mathbf{a}_r^{(n)})| \ll 1$, then

$$\mathbf{a}_r^{(m,k-1)} = \mathbf{a}_r^{(m)} + \mathcal{O}(\varepsilon_{k-1}) \quad \text{for all } m, r,$$

so

$$(\mathbf{A}^{(m)})^\top \mathbf{A}^{(m,k-1)} = [\mathbf{I} + \mathcal{O}(\varepsilon_{k-1}), \mathcal{O}(\varepsilon_{k-1})] \quad \text{for all } m.$$

Similarly,

$$\mathbf{G}^{(m,k)} = (\mathbf{A}^{(m,k-1)})^\top (\mathbf{A}^{(m,k-1)}) = [\mathbf{I}, \mathcal{O}(\varepsilon_{k-1})] \quad \text{for all } m.$$

Convergence of CP-AltLS: Odeco tensors

Proof (sketch)

Next,

$$
\begin{aligned}
(\mathbf{K}^{(n)})^\top \mathbf{K}^{(n,k)} &= \mathop{\Large\ast}_{m \neq n} (\mathbf{A}^{(m)})^\top \mathbf{A}^{(m,k-1)} \\
&= \mathop{\Large\ast}_{m \neq n} [\mathbf{I} + \mathcal{O}(\varepsilon_{k-1}), \mathcal{O}(\varepsilon_{k-1})] \\
&= [\mathbf{I} + \mathcal{O}(\varepsilon_{k-1}), \mathcal{O}(\varepsilon_{k-1}^{N-1})]
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{H}^{(n,k)} &= \mathop{\Large\ast}_{m \neq n} \mathbf{G}^{(m,k)} \\
&= \mathop{\Large\ast}_{m \neq n} [\mathbf{I}, \mathcal{O}(\varepsilon_{k-1})] \\
&= [\mathbf{I}, \mathcal{O}(\varepsilon_{k-1}^{N-1})].
\end{aligned}
$$

Convergence of CP-AltLS: Odeco tensors

Proof (sketch)

By a geometric series argument,

$$(\mathbf{H}^{(n,k)})^{+} = (\mathbf{I} + \mathcal{O}(\varepsilon_{k-1}^{N-1}))^{-1} = [\mathbf{I} + \mathcal{O}(\varepsilon_{k-1}^{N-1}), \mathcal{O}(\varepsilon_{k-1}^{N-1})].$$

Finally,

$$\begin{aligned}\mathbf{A}^{(n,k)} &= \mathbf{A}^{(n)}\mathbf{\Lambda}\,(\mathbf{K}^{(n)})^{\top}\mathbf{K}^{(n,k)}(\mathbf{H}^{(n,k)})^{+} \\ &= \mathbf{A}^{(n)}\mathbf{\Lambda}\,[\mathbf{I} + \mathcal{O}(\varepsilon_{k-1}), \mathcal{O}(\varepsilon_{k-1}^{N-1})]\,[\mathbf{I} + \mathcal{O}(\varepsilon_{k-1}^{N-1}), \mathcal{O}(\varepsilon_{k-1}^{N-1})] \\ &= \mathbf{A}^{(n)}\mathbf{\Lambda}\,[\mathbf{I} + \mathcal{O}(\varepsilon_{k-1}), \mathcal{O}(\varepsilon_{k-1}^{N-1})],\end{aligned}$$

which implies that $\varepsilon_k = \mathcal{O}(\varepsilon_{k-1}^{N-1})$.

Convergence of CP-AltLS: Odeco tensors

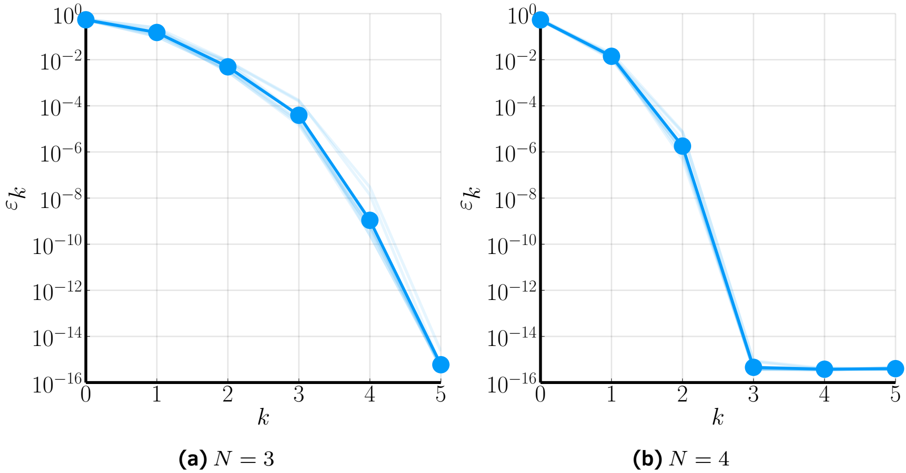


Figure: *Convergence of (decoupled) CP-AltLS for N^{th}-order odeco tensors.*

Convergence of CP-AltLS: Ideco tensors

Theorem (Local convergence of CP-AltLS for ideco tensors)

Suppose that $\mathcal{X} = [\![\boldsymbol{\lambda}\,;\, \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!]$ ($N \geq 3$) *for some* $\boldsymbol{\lambda} \in \mathbb{R}^R$ *and some* $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ *with* <u>*normalized columns*</u> *and* <u>*coherence at most* μ</u>*. In addition, let* $\kappa := \frac{\max_{r \in [R]} |\lambda_r|}{\min_{r \in [R]} |\lambda_r|}$*.*

Then there exists an absolute constant $C > 0$ *such that if* $\max\{\varepsilon_0, \mu\} < C\kappa^{-\frac{1}{N-2}} R^{-\frac{2}{N-2}}$*, then* $\varepsilon_k \to 0$ <u>*linearly*</u> *(more precisely,* $\varepsilon_k \leq \rho \varepsilon_{k-1}$ *for some absolute constant* $0 < \rho < 1$*).*

Convergence of CP-AltLS: Ideco tensors

Proof (outline)

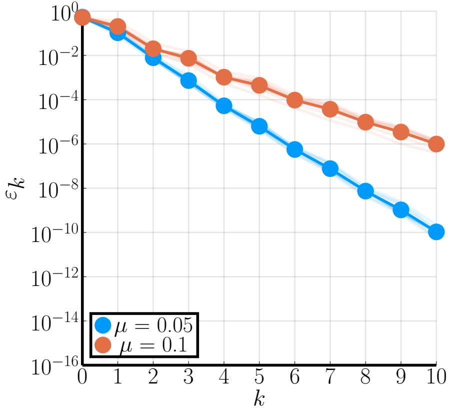
WLOG assume that $\langle \mathbf{a}_r^{(n,k)}, \mathbf{a}_r^{(n)} \rangle \geq 0$ for all k, n, r by adjusting $\boldsymbol{\lambda}$.

We no longer have $(\mathbf{A}^{(m)})^\top \mathbf{A}^{(m)} = \mathbf{I}$, so let us define

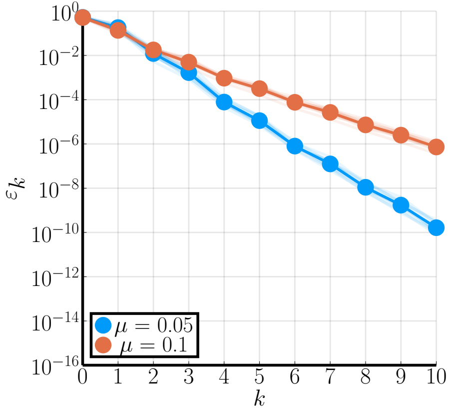
$$\mathbf{G}^{(m)} := (\mathbf{A}^{(m)})^\top \mathbf{A}^{(m)} \quad \text{and} \quad \mathbf{H}^{(n)} := \mathop{\ast}_{m \neq n} \mathbf{G}^{(m)}.$$

- $\varepsilon_{k-1} \ll 1 \implies \mathbf{A}^{(m,k-1)} \approx \mathbf{A}^{(m)} \ (\forall m)$
- $\mathbf{G}^{(m,k)} = (\mathbf{A}^{(m,k-1)})^\top (\mathbf{A}^{(m,k-1)}) \approx (\mathbf{A}^{(m)})^\top \mathbf{A}^{(m)} = \mathbf{G}^{(m)}$
- $(\mathbf{K}^{(n)})^\top \mathbf{K}^{(n,k)} = \ast_{m \neq n} (\mathbf{A}^{(m)})^\top \mathbf{A}^{(m,k-1)} \approx \ast_{m \neq n} \mathbf{G}^{(m)} = \mathbf{H}^{(n)}$
- $\mathbf{H}^{(n,k)} = \ast_{m \neq n} \mathbf{G}^{(m,k)} \approx \ast_{m \neq n} \mathbf{G}^{(m)} = \mathbf{H}^{(n)}$
- $\mathbf{A}^{(n,k)} = \mathbf{A}^{(n)} \boldsymbol{\Lambda} (\mathbf{K}^{(n)})^\top \mathbf{K}^{(n,k)} (\mathbf{H}^{(n,k)})^+ \approx \mathbf{A}^{(n)} \boldsymbol{\Lambda} \mathbf{H}^{(n)} (\mathbf{H}^{(n)})^+ = \mathbf{A}^{(n)} \boldsymbol{\Lambda}$
- $\mathbf{A}^{(n,k)} \approx \mathbf{A}^{(n)} \boldsymbol{\Lambda} \ (\forall n) \implies \varepsilon_k \ll 1$

Convergence of CP-AltLS: Ideco tensors



(a) $N = 3$



(b) $N = 4$

Figure: *Convergence of (decoupled) CP-AltLS for N^{th}-order ideco tensors.*

Algorithm: CP-AltLS

Input: $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ $(n \in [N])$

for $n = 1$ **to** N

  $\mathbf{G}^{(n)} \leftarrow (\mathbf{A}^{(n)})^\top \mathbf{A}^{(n)}$

while *stopping condition has not been satisfied*

  for $n = 1$ **to** N

    $\mathbf{K}^{(n)} \leftarrow \bigodot_{m \neq n} \mathbf{A}^{(m)}$

    $\mathbf{H}^{(n)} \leftarrow \mathop{\ast}_{m \neq n} \mathbf{G}^{(m)}$

    $\mathbf{A}^{(n)} \leftarrow \mathbf{X}_{(n)} \mathbf{K}^{(n)} (\mathbf{H}^{(n)})^+$

    normalize columns of $\mathbf{A}^{(n)}$, updating $\boldsymbol{\lambda}$ accordingly

    $\mathbf{G}^{(n)} \leftarrow (\mathbf{A}^{(n)})^\top \mathbf{A}^{(n)}$

Output: $\boldsymbol{\lambda} \in \mathbb{R}^R$, $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ $(n \in [N])$

In the standard CP-AltLS algorithm, we have

$$\mathbf{K}^{(n,k)} = \left(\bigodot_{m>n} \mathbf{A}^{(m,k-1)} \right) \odot \left(\bigodot_{m<n} \mathbf{A}^{(m,k)} \right) \text{ instead of } \bigodot_{m \neq n} \mathbf{A}^{(m,k-1)}.$$

Similarly,

$$\mathbf{H}^{(n,k)} = \left(\mathop{\Large\ast}_{m>n} \mathbf{G}^{(m,k-1)} \right) \ast \left(\mathop{\Large\ast}_{m<n} \mathbf{G}^{(m,k)} \right) \text{ instead of } \mathop{\Large\ast}_{m \neq n} \mathbf{G}^{(m,k-1)},$$

where

$$\mathbf{G}^{(m,k)} = (\mathbf{A}^{(m,k)})^\top (\mathbf{A}^{(m,k)}) \text{ instead of } (\mathbf{A}^{(m,k-1)})^\top (\mathbf{A}^{(m,k-1)}).$$

Convergence of CP-AltLS: The standard algorithm

To account for this, we consider the angular error of each individual mode $m \in [N]$,

$$\varepsilon_{m,k} := \max_{r \in [R]} |\sin \angle(\mathbf{a}_r^{(m,k)}, \mathbf{a}_r^{(m)})|,$$

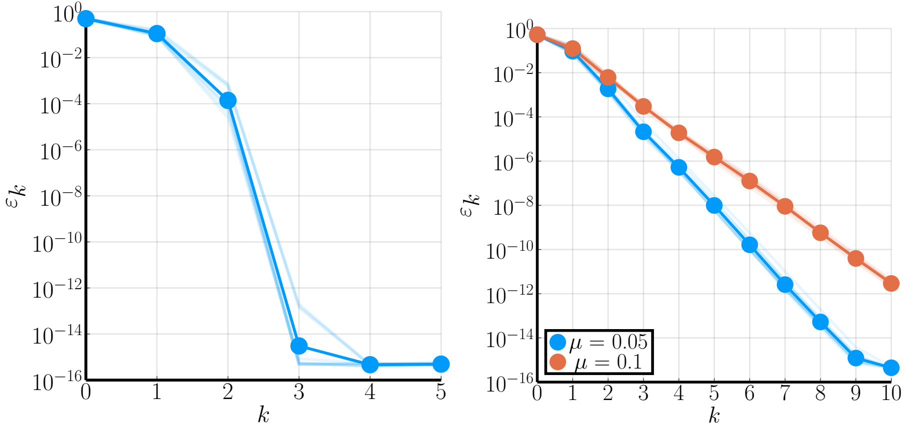
and *assume inductively* that

$$\varepsilon_{m,k} \leq \varepsilon_{k-1} = \max_{m \in [N]} \varepsilon_{m,k-1} \quad \text{for all } m < n.$$

That is, in the computation of $\mathbf{A}^{(n,k)}$, we assume that the factor matrices $\mathbf{A}^{(m,k)}$ computed in the *current* iteration for $m < n$ are *no less accurate than* the factor matrices $\mathbf{A}^{(m,k-1)}$ computed in the *previous* iteration.

Thus, the results for the decoupled algorithm also hold for the standard algorithm.

Convergence of CP-AltLS: The standard algorithm



(a) *Odeco tensor* **(b)** *Ideco tensors*

Figure: *Convergence of CP-AltLS for 3^{rd}-order tensors.*

In fact, the standard algorithm applied to odeco tensors appears to converge with an order *greater than* $N - 1$... why?

Examining the convergence proof, we find that instead of $\varepsilon_k = \mathcal{O}(\varepsilon_{k-1}^{N-1})$, we have

$$\varepsilon_{n,k} = \mathcal{O}(\varepsilon_{N,k-1} \cdots \varepsilon_{n+1,k-1} \cdot \varepsilon_{n-1,k} \cdots \varepsilon_{1,k}).$$

For example, if $N = 3$ and $\varepsilon_0 = \max\{\varepsilon_{1,0}, \varepsilon_{2,0}, \varepsilon_{3,0}\} = \mathcal{O}(\varepsilon)$, then

$$\begin{aligned}
\varepsilon_{1,1} &= \mathcal{O}(\varepsilon \cdot \varepsilon) &&= \mathcal{O}(\varepsilon^2) \\
\varepsilon_{2,1} &= \mathcal{O}(\varepsilon \cdot \varepsilon^2) &&= \mathcal{O}(\varepsilon^3) \\
\varepsilon_{3,1} &= \mathcal{O}(\varepsilon^3 \cdot \varepsilon^2) &&= \mathcal{O}(\varepsilon^5) \\
\varepsilon_{1,2} &= \mathcal{O}(\varepsilon^5 \cdot \varepsilon^3) &&= \mathcal{O}(\varepsilon^8) \\
&\vdots && \vdots
\end{aligned}$$

Convergence of CP-AltLS: The standard algorithm

Hence

$$\begin{aligned} \varepsilon_1 &= \max\{\mathcal{O}(\varepsilon^2), \mathcal{O}(\varepsilon^3), \mathcal{O}(\varepsilon^5)\} &&= \mathcal{O}(\varepsilon^2) \\ \varepsilon_2 &= \max\{\mathcal{O}(\varepsilon^8), \mathcal{O}(\varepsilon^{13}), \mathcal{O}(\varepsilon^{21})\} &&= \mathcal{O}(\varepsilon^8) \\ \vdots & \qquad\qquad\qquad \vdots && \quad \vdots \end{aligned}$$

The order of convergence is $q^3 \approx 4.2$, where $q = \frac{1+\sqrt{5}}{2} \approx 1.6$ is the positive solution of $1 + q = q^2$.

Theorem (Local convergence of CP-AltLS for odeco tensors)

Suppose that $\mathcal{X} = [\![\boldsymbol{\lambda}; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!]$ ($N \geq 3$) …

… *if* $\varepsilon_0 < C\kappa^{-\frac{1}{N-2}} R^{-\frac{1}{N-1}}$*, then* $\varepsilon_k \to 0$ *with order* q^N*, where* q *is the (unique) positive solution of* $1 + \cdots + q^{N-3} + q^{N-2} = q^{N-1}$.

Convergence of CP-AltLS: The standard algorithm

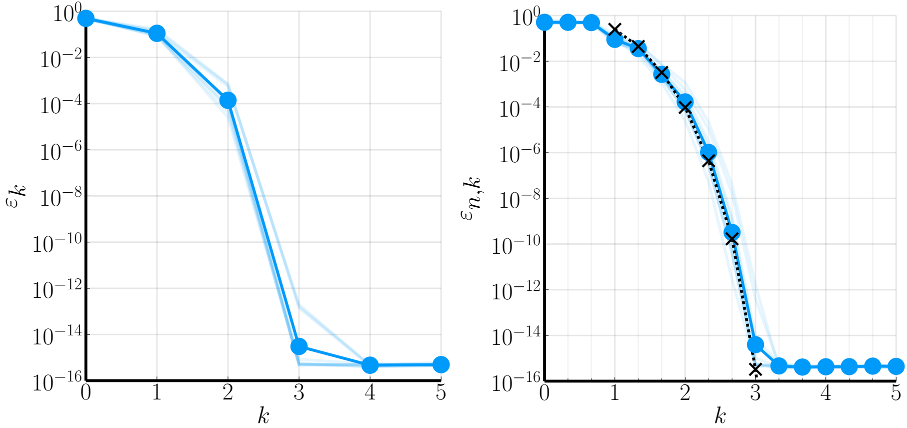


Figure: *Convergence of CP-AltLS for a 3^{rd}-order odeco tensor. The $\times$'s are $\varepsilon_{3,k-1} \cdots \varepsilon_{n+1,k-1} \cdot \varepsilon_{n-1,k} \cdots \varepsilon_{1,k}$ (that is, the products of the angular errors in the two preceding subiterations).*

Conclusion

Conclusion – summary

Our result (· , Iwen, Needell, and Wang 2025):

- arbitrary rank: not limited to rank-one tensors
- orthogonal and incoherent decompositions: not limited to odeco tensors
- quantitative and explicit: inequalities with computable constants
- specific order of convergence: order q^N (where $1 + \cdots + q^{N-2} = q^{N-1}$) for N^{th}-order odeco tensors, order 1 for ideco tensors
- more direct and less technical proof: matrix analysis

Conclusion – future work

- Convergence acceleration through factor matrix orthogonalization (Sharan and Valiant 2017) or coherence reduction
- Tensor recovery from "noisy measurements" (e.g., recovering $\mathcal{X}$ from $\mathcal{A}(\mathcal{X}) + \mathcal{E}$ for some linear operator $\mathcal{A}$ and some unknown/random tensor $\mathcal{E}$)
- Convergence of AltLS for other tensor decompositions (e.g., Tucker, DEDICOM)

Thank you for your attention!