Convergence of a Jacobi-type method for the approximate orthogonal tensor diagonalization

Erna Begović Kovač

For Short Communications Satellite 2022 (SCS 2022)
July 2022

OUTLINE

- Introduction
- Jacobi-type algorithm
- Convergence
- Numerical examples
- Modifications

Details can be found in

- E. Begović Kovač: Convergence of a Jacobi-type method for the approximate orthogonal tensor diagonalization.
  arXiv:2109.03722 [math.NA]

SVD-LIKE TENSOR DECOMPOSITION

- For the sake of simplicity, we work with real $n \times n \times n$ tensors. All results can be generalized to $n_1 \times n_2 \times \cdots \times n_d$ tensors.
- Two related problems:

♣ **SVD-like tensor decomposition**

$$\mathcal{A} = \mathcal{S} \times_1 U \times_2 V \times_3 W,$$

where $\mathcal{S}$ is a core tensor that mimics the diagonal matrix of singular values from the matrix SVD decomposition, and U, V, W are orthogonal matrices;

♣ **Tensor diagonalization problem** — find orthogonal transformations U, V, W that "diagonalize" $\mathcal{A}$.

INTRODUCTION

- A general $n \times n \times n$ tensor can not be diagonalized using orthogonal transformations.

  $\rightarrow$ We need $\mathcal{S}$ "as diagonal as possible".

- Off-diagonal norm of $\mathcal{S}$ is minimized,

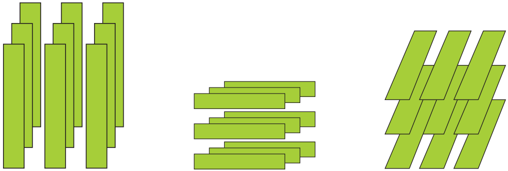
$$\mathrm{off}(\mathcal{S}) \rightarrow \min,$$

or dually, the Frobenius norm of the diagonal is maximized,

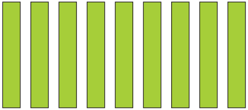
$$\|\mathrm{diag}(\mathcal{S})\|^2 = \sum_{i=1}^{n} \mathcal{S}_{iii}^2 \rightarrow \max .$$

PRELIMINARIES

- Tensor analogues of columns and rows are called **fibers**.



- mode-m **unfolding** $\leftrightarrow$ arranging mode-m fibers of $\mathcal{X}$ into columns of $X_{(m)}$.



PRELIMINARIES

- The **mode-m product** of $\mathcal{X}$ and $A \in \mathbb{R}^{c \times n}$ is a tensor

$$\mathcal{Y} = \mathcal{X} \times_m A \quad \text{such that} \quad Y_{(m)} = AX_{(m)}.$$

- Properties:

$$\mathcal{X} \times_p A \times_q B = \mathcal{X} \times_q B \times_p A, \quad p \neq q,$$

$$\mathcal{X} \times_m A \times_m B = \mathcal{X} \times_m (BA).$$

- **Tucker decomposition** of $\mathcal{X}$

$$\mathcal{X} = \mathcal{S} \times_1 A_1 \times_2 A_2 \times_3 \cdots \times_d A_d.$$

- The **norm** of $\mathcal{X}$

$$\|\mathcal{X}\| = \sqrt{\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} x_{i_1 i_2 \ldots i_d}^2}.$$

INTRODUCTION

$$\mathcal{A} = \mathcal{S} \times_1 U \times_2 V \times_3 W$$

- Core tensor $\mathcal{S}$ can be expressed as

$$\mathcal{S} = \mathcal{A} \times_1 U^T \times_2 V^T \times_3 W^T.$$

- For a given $n \times n \times n$ tensor $\mathcal{A}$ we need to find orthogonal matrices U, V, W that maximize the function

$$f(U, V, W) = \|\mathrm{diag}(\mathcal{A} \times_1 U^T \times_2 V^T \times_3 W^T)\|^2.$$

- J. Li, K. Usevich, P. Comon (SIMAX 2018, LAA 2019)
- C. D. Moravitz Martin, C. F. Van Loan (SIMAX 2008)

JACOBI-TYPE ALGORITHM

- The kth iteration:

$$\mathcal{A}^{(k+1)} = \mathcal{A}^{(k)} \times_1 R_{U,k}^T \times_2 R_{V,k}^T \times_3 R_{W,k}^T, \quad k \geq 0, \quad \mathcal{A}^{(0)} = \mathcal{A},$$

where $R_{U,k}, R_{V,k}, R_{W,k}$ are plane rotations of the form

$$R(i,j,\phi) = \begin{array}{cc} \left[\begin{array}{ccccccccccc}
1 & & & & & & & & & & \\
 & \ddots & & & & & & & & & \\
 & & 1 & & & & & & & & \\
 & & & \cos\phi & & & & -\sin\phi & & & \\
 & & & & 1 & & & & & & \\
 & & & & & \ddots & & & & & \\
 & & & & & & 1 & & & & \\
 & & & \sin\phi & & & & \cos\phi & & & \\
 & & & & & & & & 1 & & \\
 & & & & & & & & & \ddots & \\
 & & & & & & & & & & 1
\end{array}\right] & \begin{array}{c} \\ \\ \\ i \\ \\ \\ \\ j \\ \\ \\ \\ \end{array} \end{array}.$$

JACOBI-TYPE ALGORITHM

- Index pair (i, j) is called **pivot position**.
- Pivot positions are taken in any cyclic ordering.
- Matrices $R_{U,k}, R_{V,k}, R_{W,k}$ have the same pivot position (i_k, j_k), while the rotation angle ϕ_k is, in general, different for each rotation.
- Identity initialization: set $U_0 = V_0 = W_0 = I$. Then

$$\begin{aligned} U_{k+1} &= U_k R_{U,k}, \\ V_{k+1} &= V_k R_{V,k}, \\ W_{k+1} &= W_k R_{W,k}. \end{aligned}$$

MICROITERATIONS

- Each iteration consists of <u>three microiterations</u> where we hold two variables constant and vary the third one. We have

$$\mathcal{B}^{(k)} = \mathcal{A}^{(k)} \times_1 R_{U,k}^T,$$
$$\mathcal{C}^{(k)} = \mathcal{B}^{(k)} \times_2 R_{V,k}^T,$$
$$\mathcal{A}^{(k+1)} = \mathcal{C}^{(k)} \times_3 R_{W,k}^T,$$

where by $\mathcal{B}^{(k)}$ and $\mathcal{C}^{(k)}$ we denote intermediate steps.

ROTATION ANGLES

- To get the rotation angles for the pivot position (i, j) we observe a $2 \times 2 \times 2$ subproblem

$$\|\mathrm{diag}(\hat{S})\|^2 = \sigma_{iii}^2 + \sigma_{jjj}^2 \to \max,$$

where

$$\hat{\mathcal{S}} = \hat{\mathcal{A}} \times_1 \hat{U}^T \times_2 \hat{V}^T \times_3 \hat{W}^T,$$

$$\hat{\mathcal{S}}(:,:,1) = \begin{bmatrix} \sigma_{iii} & \sigma_{iji} \\ \sigma_{jii} & \sigma_{jji} \end{bmatrix}, \quad \hat{\mathcal{S}}(:,:,2) = \begin{bmatrix} \sigma_{iij} & \sigma_{ijj} \\ \sigma_{jij} & \sigma_{jjj} \end{bmatrix},$$

and $\hat{U}, \hat{V}, \hat{W}$ are 2×2 rotations.

PIVOT CONDITION

- To ensure the convergence of the algorithm, we only take pivot pair (i,j) such that (at least) one of the following inequalities is true:

$$|\langle \nabla_U f, U\dot{R}(i,j,0)\rangle| \geq \eta \|\nabla_U f\|_2,$$
$$|\langle \nabla_V f, V\dot{R}(i,j,0)\rangle| \geq \eta \|\nabla_V f\|_2,$$
$$|\langle \nabla_W f, W\dot{R}(i,j,0)\rangle| \geq \eta \|\nabla_W f\|_2,$$

  where $0 < \eta \leq \frac{2}{n}$ and $\dot{R}(i,j,0)$ denotes $\frac{\partial}{\partial\phi} R(i,j,\phi)\Big|_{\phi=0}$.
- If (i,j) does not satisfy any of these conditions, then it will be skipped and we move to the next pair in the cycle.
- It is always possible to find an appropriate pivot pair.

ALGORITHM

Jacobi-type algorithm

Input: $\mathcal{A} \in \mathbb{R}^{n\times n\times n}$
Output: orthogonal matrices U, V, W
$\mathcal{A}^{(0)} = \mathcal{A}$; $U_0 = I$; $V_0 = I$; $W_0 = I$;
REPEAT
  Select (i_k, j_k).
  Find ϕ_k for $R_{U,k}$.
  $\mathcal{B}^{(k)} = \mathcal{A}^{(k)} \times_1 R_{U,k}$
  $U_{k+1} = U_k R_{U,k}$
  Find ϕ_k for $R_{V,k}$.
  $\mathcal{C}^{(k)} = \mathcal{B}^{(k)} \times_2 R_{V,k}$
  $V_{k+1} = V_k R_{V,k}$
  Find ϕ_k for $R_{W,k}$.
  $\mathcal{A}^{(k+1)} = \mathcal{C}^{(k)} \times_3 R_{W,k}$
  $W_{k+1} = W_k R_{W,k}$
UNTIL convergence

CONVERGENCE

Lemma

For any differentiable function $f : O_n \times O_n \times O_n \to \mathbb{R}$, $U, V, W \in O_n$, *and* $0 < \eta \leq \frac{2}{n}$ *it is always possible to find index pairs* (i_U, j_U), (i_V, j_V), (i_W, j_W) *such that*

(i) $|\langle \nabla_U f(U, V, W), U\dot{R}(i_U, j_U, 0)\rangle| \geq \eta \|\nabla_U f(U, V, W)\|_2$,

(ii) $|\langle \nabla_V f(U, V, W), V\dot{R}(i_V, j_V, 0)\rangle| \geq \eta \|\nabla_V f(U, V, W)\|_2$,

(iii) $|\langle \nabla_W f(U, V, W), W\dot{R}(i_W, j_W, 0)\rangle| \geq \eta \|\nabla_W f(U, V, W)\|_2$,

where $\dot{R}(i, j, 0) = \frac{\partial}{\partial \phi} R(i, j, \phi)\Big|_{\phi=0}$.

Lemma

Let U_k, V_k, W_k, $k \geq 0$ *be the sequences generated by the algorithm. Let* $\overline{U}, \overline{V}, \overline{W}$ *be a triplet of orthogonal matrices satisfying* $\nabla f(\overline{U}, \overline{V}, \overline{W}) \neq 0$. *Then there exist* $\epsilon > 0$ *and* $\delta > 0$ *such that*

$$\|U_k - \overline{U}\|_2 < \epsilon, \quad \|V_k - \overline{V}\|_2 < \epsilon, \quad \|W_k - \overline{W}\|_2 < \epsilon$$

imply

$$f(U_{k+1}, V_{k+1}, W_{k+1}) - f(U_k, V_k, W_k) \geq \delta.$$

CONVERGENCE

Theorem

Every accumulation point (U, V, W) obtained by the Jacobi-type algorithm is a stationary point of the function

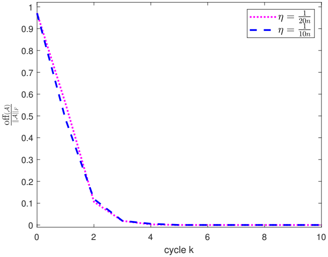
$$f(U, V, W) = \|diag(\mathcal{A} \times_1 U^T \times_2 V^T \times_3 W^T)\|^2.$$

NUMERICAL EXAMPLES 1

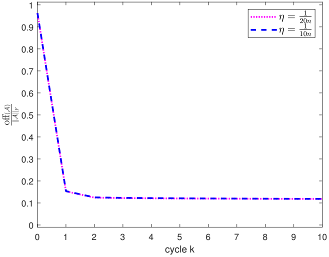
We observe the relative off-norm of a tensor which is given as

$$\frac{\mathrm{off}(\mathcal{A})}{\|\mathcal{A}\|}.$$

Change in the relative off-norm for two $30 \times 30 \times 30$ tensors with different values of η.



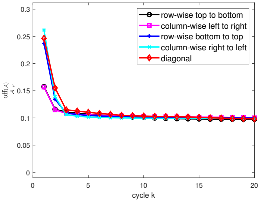
Diagonalizable tensor



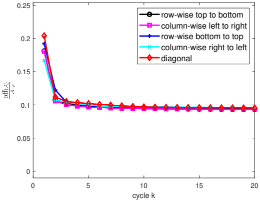
Non-diagonalizable tensor

NUMERICAL EXAMPLES 2

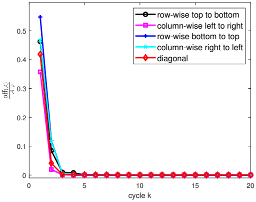
Different cyclic pivot strategies on four $10 \times 10 \times 10$ tensors.



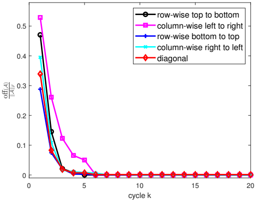
Tensor 1



Tensor 2



Tensor 3



Tensor 4

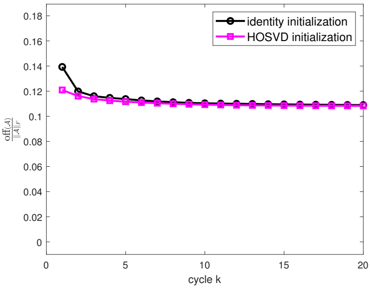
HOSVD INITIALIZATION

- Take

$$\widetilde{\mathcal{S}} = \mathcal{A} \times_1 \widetilde{U}^T \times_2 \widetilde{V}^T \times_3 \widetilde{W}^T$$

where $\widetilde{U}$, $\widetilde{V}$, and $\widetilde{W}$ are matrices of left singular vectors of $A_{(1)}$, $A_{(2)}$, and $A_{(3)}$, respectively. We set

$$\mathcal{A}^{(0)} = \widetilde{\mathcal{S}}, \quad U_0 = \widetilde{U}, \quad V_0 = \widetilde{V}, \quad W_0 = \widetilde{W}.$$



Random $20 \times 20 \times 20$ tensor

TRACE MAXIMIZATION

- Instead of maximizing

$$\|\mathrm{diag}(\mathcal{S})\|^2 = \sum_{i=1}^{n} \mathcal{S}_{iii}^2,$$

  we can opt for

$$\mathrm{tr}(\mathcal{S}) = \sum_{i=1}^{n} \mathcal{S}_{iii} \to \max .$$

More can be found in

- E. Begović Kovač, A. Bokšić: Trace maximization algorithm for the approximate tensor diagonalization.
  arXiv:2111.01466 [math.NA]

THANK YOU!

Erna Begović Kovač

University of Zagreb

ebegovic@fkit.hr

HRZZ Croatian Science Foundation

This work has been supported in part by Croatian Science Foundation under the project **UIP-2019-04-5200**.