Low rank tensor approximations with the emphasis on antisymmetric tensors

Erna Begović

University of Zagreb

ebegovic@fkit.hr

TU Braunschweig, 19th October 2016

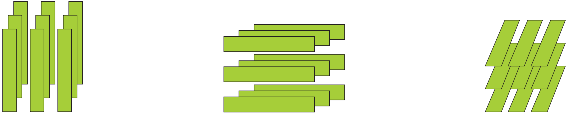
This work has been supported in part by Croatian Science Foundation under the project 3670.

OUTLINE

- Introduction
- Tensor decompositions and tensor rank
- Low multilinear rank approximation
- Jacobi method
- Numerical example
- Rank 1 approximation
- Multilinear rank d approximation

TENSORS - Basic concept and notation

- **Tensor** is multidimensional (finite) array
- **Order** of the tensor - dimension d
- Matrix $\mathbf{M}(i,j) \dashrightarrow$ Tensor $\mathcal{T}(i_1, i_2, \ldots, i_d)$
- **Fiber** - vector obtained by fixing all but one indices, i.e. $\mathcal{T}(:, i_2, \ldots, i_d)$
  Mode-k fibers



- **Slice** - matrix obtained by fixing all but two indices, i.e. $\mathcal{T}(:, :, i_3, \ldots, i_d)$

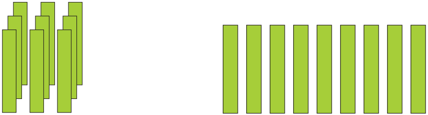
APPLICATIONS

- Physics (quantum physics, electromagnetism)
- Chemistry (quantum chemistry, computational chemistry, chemometrics)
- Engineering (signal processing, image and video processing)
- Social sciences (sociology, psychometrics)

MATRICIZATION

- **Unfolding** (matricization/flattening) - matrix representation of the tensor

$$\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \quad \dashrightarrow \quad \mathbf{T}_{(k)} \in \mathbb{R}^{n_k \times (n_1 \cdots n_{k-1} n_{k+1} \cdots n_d)}$$



MATRICIZATION - Example 1

$d = 3, n_1 = 3, n_2 = 4, n_3 = 2$

$$\mathcal{X}(:,:,1) = \begin{bmatrix} 9 & 4 & 5 & 4 \\ 2 & 8 & 0 & 6 \\ 6 & 7 & 8 & 1 \end{bmatrix}, \quad \mathcal{X}(:,:,2) = \begin{bmatrix} 2 & 0 & 1 & 5 \\ 3 & 3 & 9 & 1 \\ 7 & 6 & 5 & 0 \end{bmatrix}$$

$$\mathbf{X}_{(1)} = \left[\begin{array}{cccc|cccc} 9 & 4 & 5 & 4 & 2 & 0 & 1 & 5 \\ 2 & 8 & 0 & 6 & 3 & 3 & 9 & 1 \\ 6 & 7 & 8 & 1 & 7 & 6 & 5 & 0 \end{array}\right]$$

$$\mathbf{X}_{(2)} = \left[\begin{array}{ccc|ccc} 9 & 2 & 6 & 2 & 3 & 7 \\ 4 & 8 & 7 & 0 & 3 & 6 \\ 5 & 0 & 8 & 1 & 9 & 5 \\ 4 & 6 & 1 & 5 & 1 & 0 \end{array}\right]$$

$$\mathbf{X}_{(3)} = \left[\begin{array}{ccc|ccc|ccc|ccc} 9 & 2 & 6 & 4 & 8 & 7 & 5 & 0 & 8 & 4 & 6 & 1 \\ 2 & 3 & 7 & 0 & 3 & 6 & 1 & 9 & 5 & 5 & 1 & 0 \end{array}\right]$$

SQUARE MATRICIZATION

- **Square unfolding** - $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times n_4} \dashrightarrow \mathbf{T} \in \mathbb{R}^{n_1 n_3 \times n_2 n_4}$

$d = 4, n_1 = 3, n_2 = 2, n_3 = 2, n_4 = 2$

$$\mathcal{X}(:,:,1,1) = \begin{bmatrix} 2 & 3 \\ 6 & 5 \\ 9 & 1 \end{bmatrix}, \quad \mathcal{X}(:,:,1,2) = \begin{bmatrix} 8 & 3 \\ 6 & 9 \\ 2 & 2 \end{bmatrix},$$

$$\mathcal{X}(:,:,2,1) = \begin{bmatrix} 6 & 8 \\ 7 & 5 \\ 5 & 3 \end{bmatrix}, \quad \mathcal{X}(:,:,2,2) = \begin{bmatrix} 3 & 3 \\ 5 & 8 \\ 3 & 0 \end{bmatrix},$$

$$\mathbf{A}_{(12)} = \left[\begin{array}{cc|cc} 2 & 3 & 8 & 3 \\ 6 & 5 & 6 & 9 \\ 9 & 1 & 2 & 2 \\ \hline 6 & 8 & 3 & 3 \\ 7 & 5 & 5 & 8 \\ 5 & 3 & 3 & 0 \end{array}\right]$$

MULTIPLICATION AND NORM

- **Mode-k product**

$$\mathcal{X} \times_k \mathbf{M}$$

Properties:

(i) $\mathcal{Y} = \mathcal{X} \times_k \mathbf{M} \quad \Leftrightarrow \quad \mathbf{Y}_{(k)} = \mathbf{M}\mathbf{X}_{(k)},$

MULTIPLICATION AND NORM

- **Mode-k product**

$$\mathcal{X} \times_k \mathbf{M}$$

Properties:

(i) $\mathcal{Y} = \mathcal{X} \times_k \mathbf{M} \quad \Leftrightarrow \quad \mathbf{Y}_{(k)} = \mathbf{M}\mathbf{X}_{(k)}$,

(ii) $(\mathcal{X} \times_k \mathbf{M}_1) \times_k \mathbf{M}_2 = \mathcal{X} \times_k (\mathbf{M}_2\mathbf{M}_1)$,

(iii) $(\mathcal{X} \times_k \mathbf{M}_k) \times_j \mathbf{M}_j = (\mathcal{X} \times_j \mathbf{M}_j) \times_k \mathbf{M}_k$.

MULTIPLICATION AND NORM

- **Mode-k product**

$$\mathcal{X} \times_k \mathbf{M}$$

Properties:

(i) $\mathcal{Y} = \mathcal{X} \times_k \mathbf{M} \quad \Leftrightarrow \quad \mathbf{Y}_{(k)} = \mathbf{M}\mathbf{X}_{(k)}$,

(ii) $(\mathcal{X} \times_k \mathbf{M}_1) \times_k \mathbf{M}_2 = \mathcal{X} \times_k (\mathbf{M}_2\mathbf{M}_1)$,

(iii) $(\mathcal{X} \times_k \mathbf{M}_k) \times_j \mathbf{M}_j = (\mathcal{X} \times_j \mathbf{M}_j) \times_k \mathbf{M}_k$.

- $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, Frobenius norm

$$\|\mathcal{X}\|^2 = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} \mathcal{X}(i_1, i_2, \ldots, i_d)^2$$

SYMMETRIC AND ANTISYMMETRIC TENSORS

- Symmetric: $\mathcal{X}(i_1, i_2, \ldots, i_d) = \mathcal{X}(i_{\sigma(1)}, i_{\sigma(2)}, \ldots, i_{\sigma(d)})$
- Antisymmetric tensor

$$\mathcal{A}(i_1, i_2, \ldots, i_d) = (-1)^{|\sigma|}\mathcal{A}(i_{\sigma(1)}, i_{\sigma(2)}, \ldots, i_{\sigma(d)})$$

SYMMETRIC AND ANTISYMMETRIC TENSORS

- Symmetric: $\mathcal{X}(i_1, i_2, \ldots, i_d) = \mathcal{X}(i_{\sigma(1)}, i_{\sigma(2)}, \ldots, i_{\sigma(d)})$
- Antisymmetric tensor

$$\mathcal{A}(i_1, i_2, \ldots, i_d) = (-1)^{|\sigma|}\mathcal{A}(i_{\sigma(1)}, i_{\sigma(2)}, \ldots, i_{\sigma(d)})$$

Löwdin rules:

$$\begin{aligned}
(i)\ & \mathcal{A}(i,j,k) = 0, \quad \text{if } i = j \text{ or } i = k \text{ or } j = k,\\
(ii)\ & \mathcal{A}(i,j,k) = \mathcal{A}(j,k,i) = \mathcal{A}(k,i,j)\\
& \qquad = -\mathcal{A}(j,i,k) = -\mathcal{A}(k,j,i) = -\mathcal{A}(i,k,j), \ \text{otherwise.}
\end{aligned}$$

- **Antisymmetrizer** anti($\mathcal{X}$) - projection on the space of antisymmetric tensors

MATRICIZATION - Example 2

- Let $\mathcal{A} \in \mathbb{R}^{n \times n \times \cdots \times n}$ be an antisymmetric tensor of order d. Then
$$\mathbf{A}_{(k)} = (-1)^{|k-l|} \mathbf{A}_{(l)}, \quad 1 \leq k, l \leq d.$$

MATRICIZATION - Example 2

- Let $\mathcal{A} \in \mathbb{R}^{n \times n \times \cdots \times n}$ be an antisymmetric tensor of order d. Then

$$\mathbf{A}_{(k)} = (-1)^{|k-l|} \mathbf{A}_{(l)}, \quad 1 \leq k, l \leq d.$$

- $d = 3, n = 4$

$$\mathcal{A}(:,:,1) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & -1 & 0 & 3 \\ 0 & -2 & -3 & 0 \end{bmatrix}, \quad \mathcal{A}(:,:,2) = \begin{bmatrix} 0 & 0 & -1 & -2 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 4 \\ 2 & 0 & -4 & 0 \end{bmatrix}$$

$$\mathcal{A}(:,:,3) = \begin{bmatrix} 0 & 1 & 0 & -3 \\ -1 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 \end{bmatrix}, \quad \mathcal{A}(:,:,4) = \begin{bmatrix} 0 & 2 & 3 & 0 \\ -2 & 0 & 4 & 0 \\ -3 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{A}_{(1)} = \mathbf{A}_{(3)} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & -1 & -2 & 0 & 1 & 0 & -3 & 0 & 2 & 3 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -4 & -2 & 0 & 4 & 0 \\ 0 & -1 & 0 & 3 & 1 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & -3 & -4 & 0 & 0 \\ 0 & -2 & -3 & 0 & 2 & 0 & -4 & 0 & 3 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

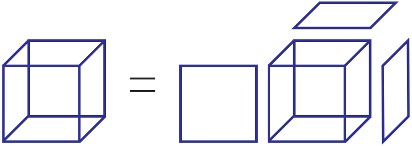
$$\mathbf{A}_{(2)} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 & -1 & 0 & 3 & 0 & -2 & -3 & 0 \\ 0 & 0 & -1 & -2 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 4 & 2 & 0 & -4 & 0 \\ 0 & 1 & 0 & -3 & -1 & 0 & 0 & -4 & 0 & 0 & 0 & 0 & 3 & 4 & 0 & 0 \\ 0 & 2 & 3 & 0 & -2 & 0 & 4 & 0 & -3 & -4 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Tensor decompositions and tensor rank

TUCKER DECOMPOSITION AND HOSVD

$\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$

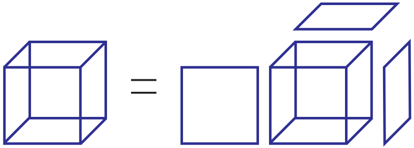
- Tucker (1966) $\mathcal{X} = \mathcal{T} \times_1 \mathbf{M}_1 \times_2 \mathbf{M}_2 \times_3 \cdots \times_d \mathbf{M}_d$



TUCKER DECOMPOSITION AND HOSVD

$\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$

- Tucker (1966) $\quad \mathcal{X} = \mathcal{T} \times_1 \mathbf{M}_1 \times_2 \mathbf{M}_2 \times_3 \cdots \times_d \mathbf{M}_d$



- **Higher order SVD** De Lathauwer et al. (2000)

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \cdots \times_d \mathbf{U}_d,$$

where $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is core tenzor, and $\mathbf{U}_k \in \mathbb{R}^{n_k \times n_k}$ are unitary matrices, $1 \le k \le d$.

For $d \ge 3$, $\mathcal{S}$ is not a diagonal tensor!

HOSVD Algorithm

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_d \mathbf{U}_d$$
$$\mathcal{S} = \mathcal{X} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \cdots \times_d \mathbf{U}_d^T$$

HOSVD algorithm, $\mathcal{X}$

FOR $k = 1, \ldots, d$
  Compute SVD of $\mathbf{X}_{(k)}$, $\mathbf{X}_{(k)} = \mathbf{U}_k \mathbf{S}_k \mathbf{V}_k^T$.
ENDFOR
$\mathcal{S} = \mathcal{X} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \cdots \times_d \mathbf{U}_d^T$

HOSVD of antisymmetric tensor

If $\mathcal{A}$ is antisymmetric

$$\mathbf{A}_{(1)} = -\mathbf{A}_{(2)} = \mathbf{A}_{(3)} = -\mathbf{A}_{(4)} = \cdots$$

and

$$\mathbf{U}_1 = \mathbf{U}_2 = \cdots = \mathbf{U}_d = \mathbf{U}$$

$$\Rightarrow \quad \mathcal{A} = \mathcal{S} \times_1 \mathbf{U} \times_2 \mathbf{U} \cdots \times_d \mathbf{U}$$

HOSVD algorithm

Compute SVD of $\mathbf{A}_{(1)}$, $\mathbf{A}_{(1)} = \mathbf{USV}^T$.
$\mathcal{S} = \mathcal{A} \times_1 \mathbf{U}^T \times_2 \mathbf{U}^T \times_3 \cdots \times_d \mathbf{U}^T$

HOSVD ⇢ SVD

$$\mathbf{M}_{(1)} = \mathbf{U}_1\Sigma_1\mathbf{V}_1^T, \qquad \mathbf{M}_{(1)} = \mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^T$$
$$\mathbf{M}_{(2)} = \mathbf{U}_2\Sigma_2\mathbf{V}_2^T, \qquad \mathbf{M}_{(2)} = \mathbf{M}^T = \mathbf{V}\Sigma\mathbf{U}^T$$

HOSVD ⇢ SVD

$$\mathbf{M}_{(1)} = \mathbf{U}_1 \Sigma_1 \mathbf{V}_1^T, \qquad \mathbf{M}_{(1)} = \mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^T$$
$$\mathbf{M}_{(2)} = \mathbf{U}_2 \Sigma_2 \mathbf{V}_2^T, \qquad \mathbf{M}_{(2)} = \mathbf{M}^T = \mathbf{V}\Sigma\mathbf{U}^T$$

$$\Rightarrow \mathbf{U} = \mathbf{U}_1 = \mathbf{V}_2, \quad \mathbf{V} = \mathbf{V}_1 = \mathbf{U}_2, \quad \Sigma = \Sigma_1 = \Sigma_2$$

HOSVD $\dashrightarrow$ SVD

$$\mathbf{M}_{(1)} = \mathbf{U}_1\Sigma_1\mathbf{V}_1^T, \qquad \mathbf{M}_{(1)} = \mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^T$$
$$\mathbf{M}_{(2)} = \mathbf{U}_2\Sigma_2\mathbf{V}_2^T, \qquad \mathbf{M}_{(2)} = \mathbf{M}^T = \mathbf{V}\Sigma\mathbf{U}^T$$

$$\Rightarrow \mathbf{U} = \mathbf{U}_1 = \mathbf{V}_2, \quad \mathbf{V} = \mathbf{V}_1 = \mathbf{U}_2, \quad \Sigma = \Sigma_1 = \Sigma_2$$

$$\mathbf{S} = \mathbf{M} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T = (\mathbf{U}_1^T\mathbf{M}) \times_2 \mathbf{U}_2^T = \mathbf{U}_2^T(\mathbf{U}_1^T\mathbf{M})^T = \mathbf{U}_2^T\mathbf{M}^T\mathbf{U}_1$$

HOSVD $\dashrightarrow$ SVD

$$\mathbf{M}_{(1)} = \mathbf{U}_1 \Sigma_1 \mathbf{V}_1^T, \qquad \mathbf{M}_{(1)} = \mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^T$$
$$\mathbf{M}_{(2)} = \mathbf{U}_2 \Sigma_2 \mathbf{V}_2^T, \qquad \mathbf{M}_{(2)} = \mathbf{M}^T = \mathbf{V}\Sigma\mathbf{U}^T$$

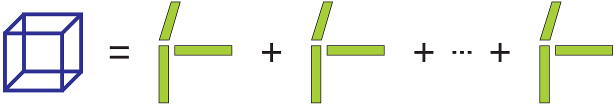
$$\Rightarrow \mathbf{U} = \mathbf{U}_1 = \mathbf{V}_2, \quad \mathbf{V} = \mathbf{V}_1 = \mathbf{U}_2, \quad \Sigma = \Sigma_1 = \Sigma_2$$

$$\mathbf{S} = \mathbf{M} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T = (\mathbf{U}_1^T \mathbf{M}) \times_2 \mathbf{U}_2^T = \mathbf{U}_2^T (\mathbf{U}_1^T \mathbf{M})^T = \mathbf{U}_2^T \mathbf{M}^T \mathbf{U}_1$$

$$\Rightarrow \mathbf{S} = \mathbf{V}^T \mathbf{M}^T \mathbf{U} = \Sigma$$

CP DECOMPOSITION

- Hitchcock (1927)
- CANDECOMP (canonical decomposition), Carroll and Chang (1970) / PARAFAC (parallel factors), Harshman (1970)



- **CP decomposition**, $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$

$$\mathcal{X} \approx \sum_{j=1}^{R} x_1^{(j)} \circ x_2^{(j)} \circ \ldots \circ x_d^{(j)},$$

with $R \in \mathbb{N}$, $x_k^{(j)} \in \mathbb{R}^{n_k}$, $1 \leq k \leq d$, $1 \leq j \leq R$.
$\circ$ stands for the outer product,

$$\mathcal{T} = x \circ y \circ z \quad \Leftrightarrow \quad \mathcal{T}(i, j, k) = x(i)y(j)z(k).$$

RANK AND MULTILINEAR RANK

- The smallest number R in the exact CP decomposition is called **tensor rank** (CP rank).
- There is no straightforward algorithm to determine the rank of a specific tensor. The problem is NP-hard.

RANK AND MULTILINEAR RANK

- The smallest number R in the exact CP decomposition is called **tensor rank** (CP rank).
- There is no straightforward algorithm to determine the rank of a specific tensor. The problem is NP-hard.
- **Multilinear rank** of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is d-tuple

$$(r_1, r_2, \ldots, r_d), \quad \text{where } r_k = \operatorname{rank}(\mathbf{X}_{(k)}),\ 1 \leq k \leq d.$$

RANK AND MULTILINEAR RANK

- The smallest number R in the exact CP decomposition is called **tensor rank** (CP rank).
- There is no straightforward algorithm to determine the rank of a specific tensor. The problem is NP-hard.
- **Multilinear rank** of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is d-tuple

$$(r_1, r_2, \ldots, r_d), \quad \text{where } r_k = \text{rank}(\mathbf{X}_{(k)}),\ 1 \leq k \leq d.$$

- If $\mathcal{A}$ is antisymmetric, then $r_1 = r_2 = \cdots = r_d = r$.
- We say that $\mathcal{A}$ has multilinear rank r and write $\mathcal{A} \in \mathcal{M}_r$.

MULTILINEAR RANK - antisymmetric tensor

Theorem (B., Kressner, 2016)

Let $\mathcal{A} \in \mathbb{R}^{n \times n \times \cdots \times n}$ be an antisymmetric tensor of order $d \geq 3$. Then the multilinear rank r of $\mathcal{A}$ satisfies

(i) $r = 0$, for $n < d$;

(ii) $r \leq d$, for $n = d$ or $n = d + 1$;

(iii) $r \leq n$, for $n \geq d + 2$.

There exist tensors $\mathcal{A}$ for which equality is attained in (i)–(iii).

Corrolary (B., Kressner, 2016)

Let $\mathcal{A} \in \mathbb{R}^{n \times n \times \cdots \times n}$ be an antisymmetric tensor of order $d \geq 3$. Then the multilinear rank r of $\mathcal{A}$ attains the values from the set

$$\{0, d, d + 2, \ldots, n\}.$$

Low multilinear rank approximation

LOW MULTILINEAR RANK APPROXIMATION

- **Minimization problem**: For a given antisymmetric tensor $\mathcal{A}$, find an antisymmetric tensor $\hat{\mathcal{A}} \in \mathcal{M}_r$, such that

$$\|\mathcal{A} - \hat{\mathcal{A}}\|^2 \to \min .$$

LOW MULTILINEAR RANK APPROXIMATION

- **Minimization problem**: For a given antisymmetric tensor $\mathcal{A}$, find an antisymmetric tensor $\hat{\mathcal{A}} \in \mathcal{M}_r$, such that

$$\|\mathcal{A} - \hat{\mathcal{A}}\|^2 \to \min .$$

- Dual **maximization problem** (De Lathauwer, 2000): find matrices $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_d$ with orthonormal coulmns, such that

$$\|\mathcal{A} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \cdots \times_d \mathbf{U}_d^T\|^2 \to \max .$$

Then,

$$\mathcal{S} = \mathcal{A} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \cdots \times_d \mathbf{U}_d^T ,$$
$$\hat{\mathcal{A}} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \cdots \times_d \mathbf{U}_d .$$

$\mathcal{A}$ antisymmetric $\dashrightarrow$ $\mathbf{U}_1 = \cdots = \mathbf{U}_d$

T-HOSVD

- **Truncated HOSVD** ($\sim$ 2000)

FOR $k = 1, \ldots, d$
    Compute SVD of $\mathbf{A}_{(k)}$, $\mathbf{A}_{(k)} = \mathbf{U}_k \mathbf{S}_k \mathbf{V}_k^T$.
    $\mathbf{W}_k = \mathbf{U}_k(:, 1 : r_k)$
ENDFOR
$\mathcal{S} = \mathcal{A} \times_1 \mathbf{W}_1^T \times_2 \mathbf{W}_2^T \times_3 \cdots \times_d \mathbf{W}_d^T$

- Direct method
- Unlike truncated SVD, does not give the best approximation
- Gives a good starting point for iterative algorithms

HOOI

- **Higher order orthogonal iterations** ($\sim$ 2006)
- ALS algorithm,
  in each microiteration one matrix $\mathbf{U}_k$, $1 \leq k \leq d$, is updated
- In practice, converges to the stationary point

HOOI

- **Higher order orthogonal iterations** ($\sim$ 2006)
- ALS algorithm,
  in each microiteration one matrix $\mathbf{U}_k$, $1 \leq k \leq d$, is updated
- In practice, converges to the stationary point

HOOI algorithm, $d = 3$

Take initial U_1, U_2, U_3.
REPEAT
  $\mathcal{X} = \mathcal{A} \times_2 U_2^T \times_3 U_3^T$
  Compute SVD of $\mathbf{X}_{(1)}$, $\mathbf{X}_{(1)} = \mathbf{USV}^T$, $U_1 = U(:, 1 : r_1)$.
  $\mathcal{X} = \mathcal{A} \times_1 U_1^T \times_3 U_3^T$
  Compute SVD of $\mathbf{X}_{(2)}$, $\mathbf{X}_{(2)} = \mathbf{USV}^T$, $U_2 = U(:, 1 : r_2)$.
  $\mathcal{X} = \mathcal{A} \times_1 U_1^T \times_2 U_2^T$
  Compute SVD of $\mathbf{X}_{(3)}$, $\mathbf{X}_{(3)} = \mathbf{USV}^T$, $U_3 = U(:, 1 : r_3)$.
UNTIL convergence
$\mathcal{B} = \mathcal{A} \times_3 U_3^T$
$\hat{\mathcal{A}} = \mathcal{B} \times_1 U_1 \times_2 U_2 \times_3 U_3$

Jacobi method

JACOBI METHOD

Idea: Apply Jacobi rotations in order to maximize the norm of $(r, r, \ldots, r)$-subtensor with smallest indices.
(Ishteva et al., 2013, symmetric case)

Jacobi rotations are $n \times n$ matrices

$$\mathbf{R}(i,j,\phi) = \begin{bmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & & \cos\phi & & & & -\sin\phi & & & \\ & & & & 1 & & & & & & \\ & & & & & \ddots & & & & & \\ & & & & & & 1 & & & & \\ & & & \sin\phi & & & & \cos\phi & & & \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ i \\ \\ \\ \\ j \\ \\ \\ \\ \end{matrix},$$

where $(i,j) = (i(k), j(k))$ is called the k-th pivot pair.

JACOBI METHOD - Algorithm

$$\|\mathcal{A} \times_1 \mathbf{U}^T \times_2 \mathbf{U}^T \times_3 \cdots \times_d \mathbf{U}^T\|^2 \to \max .$$

For the sake of simplicity, assume $d = 3$, $\mathcal{A} \in \mathbb{R}^{n \times n \times n}$.
For $\mathbf{Q}$ orthogonal and $\mathbf{M} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$, let

$$f(\mathbf{Q}) = \|\mathcal{A} \times_1 \mathbf{MQ}^T \times_2 \mathbf{MQ}^T \times_3 \mathbf{MQ}^T\|^2.$$

Maximization problem: $f(\mathbf{Q}) \to \max .$

JACOBI METHOD - Algorithm

$$\|\mathcal{A} \times_1 \mathbf{U}^T \times_2 \mathbf{U}^T \times_3 \cdots \times_d \mathbf{U}^T\|^2 \to \max .$$

For the sake of simplicity, assume $d = 3$, $\mathcal{A} \in \mathbb{R}^{n\times n\times n}$.

For $\mathbf{Q}$ orthogonal and $\mathbf{M} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$, let

$$f(\mathbf{Q}) = \|\mathcal{A} \times_1 \mathbf{M}\mathbf{Q}^T \times_2 \mathbf{M}\mathbf{Q}^T \times_3 \mathbf{M}\mathbf{Q}^T\|^2.$$

Maximization problem: $f(\mathbf{Q}) \to \max .$

REPEAT
- Choose $(i(k), j(k))$.
- Find ϕ.
- $\mathbf{R}_k = \mathbf{R}(i(k), j(k), \phi(k))$
- $\mathbf{Q}_{k+1} = \mathbf{Q}_k \mathbf{R}_k$
- $\mathcal{A}_{k+1} = \mathcal{A}_k \times_1 \mathbf{R}_k^T \times_2 \mathbf{R}_k^T \times_3 \mathbf{R}_k^T$

UNTIL convergence

JACOBI METHOD - Convergence

- **Pivot pairs** (i, j) are taken from the set

$$(1, r+1), (1, r+2), \ldots, (1, n), (2, r+1), \ldots, (r, n).$$

- **Rotation angle** ϕ maximizes

$$\psi(\phi) = \sum_{p,q=1}^{r} \left(\mathcal{A}(i, p, q) \cos\phi + \mathcal{A}(j, p, q) \sin\phi\right)^2.$$

JACOBI METHOD - Convergence

- **Pivot pairs** (i, j) are taken from the set

$$(1, r+1), (1, r+2), \ldots, (1, n), (2, r+1), \ldots, (r, n).$$

- **Rotation angle** ϕ maximizes

$$\psi(\phi) = \sum_{p,q=1}^{r} (\mathcal{A}(i, p, q)\cos\phi + \mathcal{A}(j, p, q)\sin\phi)^2.$$

Theorem (B., Kressner, 2016)

Let $(\mathbf{Q}_k)_k$ be a sequence of orthogonal matrices obtained by the Jacobi algorithm with $\mathcal{A} \in \mathbb{R}^{n\times n\times\cdots\times n}$ antisymmetric. Every accumulation point of $(\mathbf{Q}_k)_k$ is a stationary point of function $f(\mathbf{Q}) = \|\mathcal{A} \times_1 \mathbf{MQ}^T \times_2 \mathbf{MQ}^T \times_3 \mathbf{MQ}^T\|^2$.

Numerical examples - Approximation error

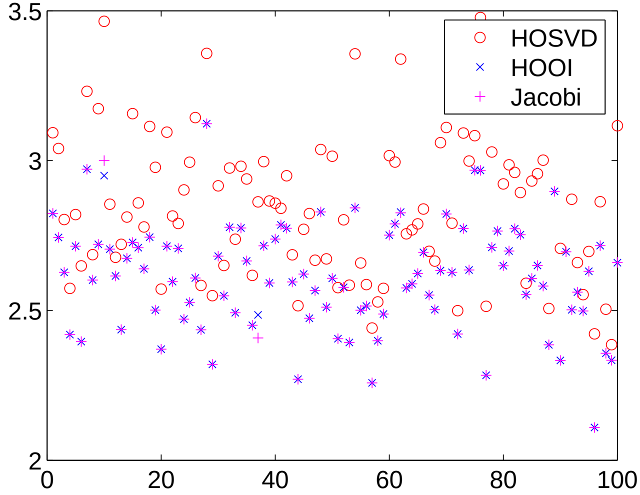


Figure: Multilinear rank 3 approximation of 100 random antisymmetric $10 \times 10 \times 10$ tensors.

Numerical examples - Convergence

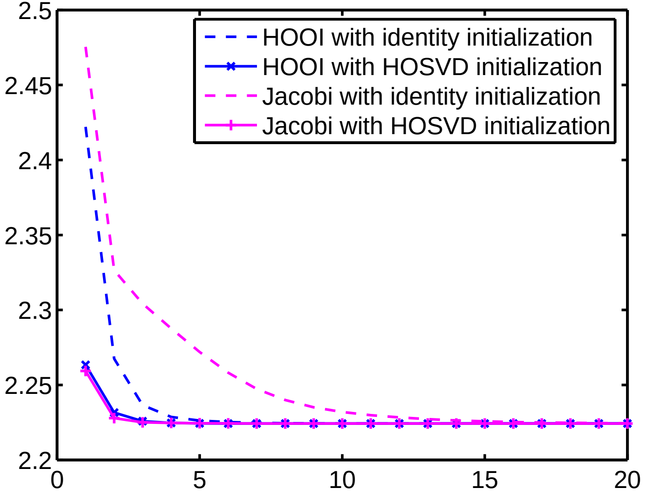


Figure: Multilinear rank 6 approximation of a random $10 \times 10 \times 10$ tensor.

Rank 1 approximation

RANK 1 APPROXIMATION

- **Minimization problem**:

$$\|\mathcal{A} - u_1 \circ u_2 \circ \ldots \circ u_d\|^2 \rightarrow \min,$$

- Dual **maximization problem** (Zhang, Golub, 2001.):

$$\sum_{i_1,\ldots,i_d} (\mathcal{A}(i_1,\ldots,i_d) u_1(i_1) \cdots u_d(i_d)) \rightarrow \max,$$

with $\|u_1\|_2 = \cdots = \|u_d\|_2 = 1.$

HOPM

- **Higher order power method** ($\sim$ 2002)
- ALS algorithm,
  in each microiteration one of the vectors $u_1, u_2, \ldots, u_d$ is updated.
- Converges to local min/max.

HOPM

- **Higher order power method** ($\sim$ 2002)
- ALS algorithm,
  in each microiteration one of the vectors $u_1, u_2, \ldots, u_d$ is updated.
- Converges to local min/max.

HOPM algorithm

Take initial $u_1, u_2, \ldots, u_d$.
REPEAT
$u_1 = \mathcal{A} \times_2 u_2^T \times_3 u_3^T \cdots \times_d u_d^T$, $u_1 = u_1 / \|u_1\|$.
$\ldots$
$u_d = \mathcal{A} \times_1 u_1^T \times_2 u_2^T \cdots \times_{d-1} u_{d-1}^T$, $u_d = u_d / \|u_d\|$.
UNTIL convergence
$\alpha = \mathcal{A} \times_1 u_1^T \times_2 u_2^T \cdots \times_d u_d^T$
Return approximation $\alpha u_1 \circ u_2 \circ \cdots \circ u_d$

MULTILINEAR RANK d APPROXIMATION

- $\mathcal{A}$ antisymmetric
  $\mathcal{A} \dashrightarrow \mathcal{B} = u_1 \circ u_2 \circ \cdots \circ u_d$ rank 1, unstructured
  $\mathcal{B} \dashrightarrow \hat{\mathcal{A}}$ multilinear rank d, antisymmetric

MULTILINEAR RANK d APPROXIMATION

- $\mathcal{A}$ antisymmetric
  $\mathcal{A} \dashrightarrow \mathcal{B} = u_1 \circ u_2 \circ \cdots \circ u_d$ rank 1, unstructured
  $\mathcal{B} \dashrightarrow \hat{\mathcal{A}}$ multilinear rank d, antisymmetric

Theorem (B., Kressner, 2016)

Let $\mathcal{A} \in \mathbb{R}^{n \times \cdots \times n}$ be an antisymmetric tensor of order d. It holds

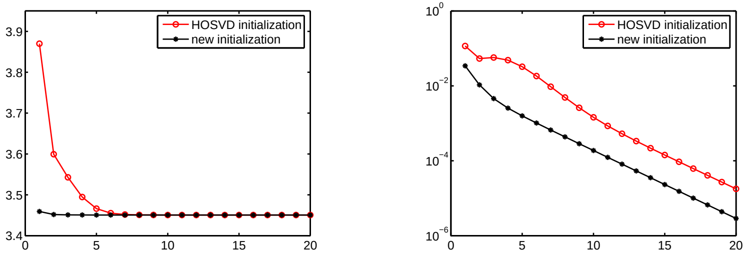
$$\begin{aligned} &\max \left\{ \|\mathcal{A} \times_1 U^T \cdots \times_d U^T\| \;\middle|\; U \in \mathbb{R}^{n \times d}, U^T U = I_d \right\} \\ =&d! \max \left\{ |\mathcal{A} \times_1 v_1^T \cdots \times_d v_d^T| \;\middle|\; \|v_1\|_2 = \cdots = \|v_d\|_2 = 1 \right\}. \end{aligned}$$

- If $\mathcal{B} = \alpha u_1 \circ u_2 \circ \cdots \circ u_d$, is the best rank-1 approximation and u_k are mutually orthonormal, $1 \leq k \leq d$, then

$$\hat{\mathcal{A}} = d! \text{ anti}(\mathcal{B})$$

  is the best multilinear rank d approximation.

$d = 4$: New initialization for rank 1 approximation - Convergence



(a) Approximation error (b) Norm of gradient

Figure: Multilinear rank 4 approximation of a random $10 \times 10 \times 10 \times 10$ tensor.

REFERENCES

- E. Begović, D. Kressner: *Structure-preserving low multilinear rank approximation of antisymmetric tensors.* arXiv:1603.05010 [math.NA]
- T. G. Kolda, B. W. Bader: *Tensor Decompositions and Applications.* SIAM Rev. 51 (3) (2009) 455-500.
- M. Ishteva, P.-A. Absil, P. Van Dooren: *Jacobi algorithm for the best low multilinear rank approximation of symmetric tensors.* SIAM J. Matrix Anal. Appl. 34 (2) (2013) 651–672.
- L. De Lathauwer, B. De Moor, J. Vandewalle: *A multilinear singular value decomposition.* SIAM J. Matrix Anal. Appl. 21 (4) (2000) 1253–1278.
- J. D. Carroll, J. J. Chang: *Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition.* Psychometrika 35 (1970) 283–319.
- R. A. Harshman: *Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis.* UCLA working papers in phonetics 16 (1970) 1–84.

THANK YOU!