Structure-preserving low multilinear rank approximation of antisymmetric tensors

Erna Begović Kovač

University of Zagreb
Faculty of Chemical Engineering and Technology
ebegovic@fkit.hr

Joint work with Daniel Kressner (EPF Lausanne)

6th Croatian Mathematical Congress
14.6.2016.

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

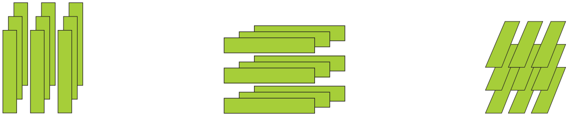
OUTLINE

- Introduction
- Tensor decompositions and tensor rank
- Low multilinear rank approximation
- Jacobi method
- Multilinear rank d approximation

E. Begović Kovač, D. Kressner: Structure-preserving low multilinear rank approximation of antisymmetric tensors. arXiv:1603.05010 [math.NA]

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. $\mathcal{T}(:, :, i_3, \ldots, i_d)$

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

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

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:

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

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

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:

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

- **Antisymmetrizer** anti($\mathcal{X}$) - projection on the space of antisymmetric tensors
- Applications: Quantum physics and quantum chemistry

MATRICIZATION

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

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

MATRICIZATION

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

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

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

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

MULTIPLICATION AND NORM

- **Mode-k product**

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

Property:

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

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

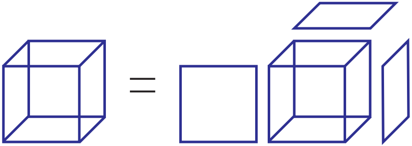
$$\|\mathcal{X}\|^2 = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} \mathcal{X}(i_1, i_2, \ldots, i_d)^2$$

TUCKER DECOMPOSITION AND HOSVD

TUCKER DECOMPOSITION AND HOSVD

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

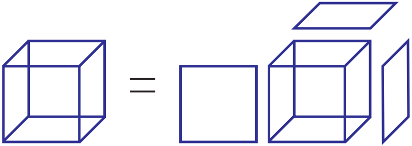
- Tucker (1966) $\mathcal{X} = \mathcal{T} \times_1 \mathbf{M}_1 \times_2 \mathbf{M}_2 \times_3 \cdots \times_d \mathbf{M}_d$



TUCKER DECOMPOSITION AND HOSVD

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

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



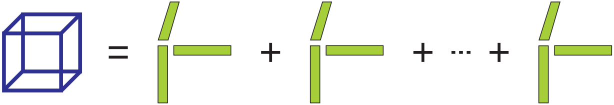
- **Higher order SVD** De Lathauwer et al. (2000)

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

where $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is core tenzor, and $U_k \in \mathbb{R}^{n_k \times n_k}$ are unitary matrices, $1 \leq k \leq d$.
For $d \geq 3$, $\mathcal{S}$ is not a diagonal tensor!

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

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

- 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)

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)

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 and HOOI

- **Truncated HOSVD** ($\sim$ 2000)
- Direct method
- Unlike T-SVD, does not give the best approximation
- Gives a good starting point for iterative algorithms

T-HOSVD and HOOI

- **Truncated HOSVD** ($\sim$ 2000)
- Direct method
- Unlike T-SVD, does not give the best approximation
- Gives a good starting point for iterative algorithms

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

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)

Jacobi rotations are $n \times n$ matrices

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

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 $M = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$, 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 $M = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$, let

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

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

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

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

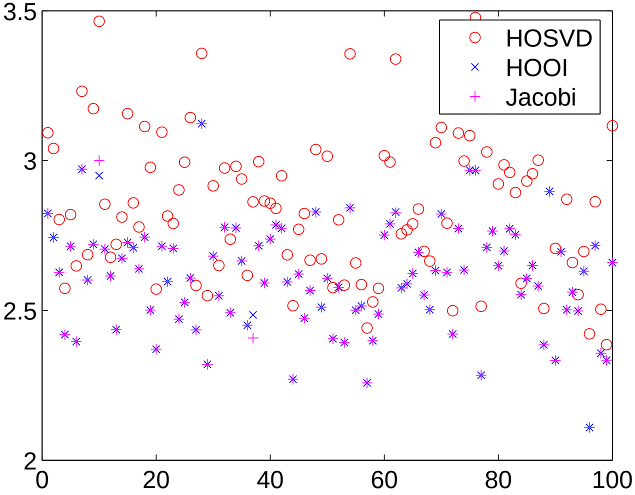


Figure: Multilinear rank 3 approximation of 100 random antisymmetric $10 \times 10 \times 10$ tensors.

Numerical examples - Approximation error

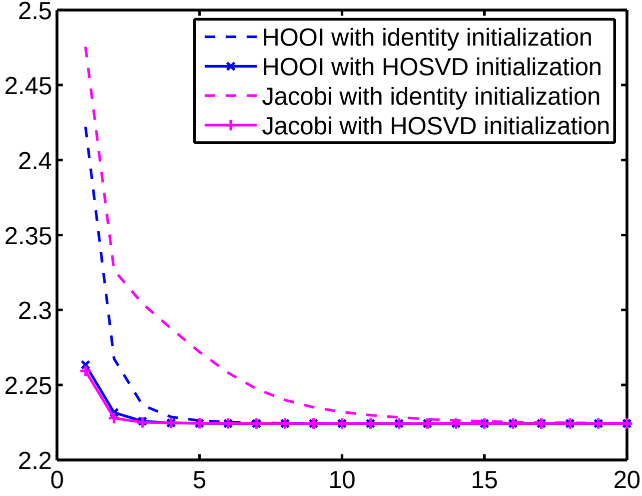


Figure: Multilinear rank 6 approximation of a random $10 \times 10 \times 10$ tensor.

MULTILINEAR RANK d APPROXIMATION

MULTILINEAR RANK d APPROXIMATION

- $\mathcal{A}$ antisymmetric
  $\mathcal{A} \dashrightarrow \mathcal{B} = x_1 \circ x_2 \circ \cdots \circ x_d$ rank 1, unstructured
  $\mathcal{B} \dashrightarrow \hat{\mathcal{A}}$ multilinear rank d, antisymmetric

MULTILINEAR RANK d APPROXIMATION

- $\mathcal{A}$ antisymmetric
  $\mathcal{A} \dashrightarrow \mathcal{B} = x_1 \circ x_2 \circ \cdots \circ x_d$ rank 1, unstructured
  $\mathcal{B} \dashrightarrow \hat{\mathcal{A}}$ multilinear rank d, antisymmetric

Theorem (B., Kressner)

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

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

- 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

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

is the best multilinear rank d approximation.

HOPM AND NEW INITIALIZATION

- **Minimization problem**: $\|\mathcal{A} - u_1 \circ u_2 \circ \ldots \circ u_d\|^2 \to \min$
- **Higher order power method** ($\sim$ 2002)
- ALS algorithm,
  in each microiteration one vector u_k, $1 \leq k \leq d$, is updated
- Converges to local minima

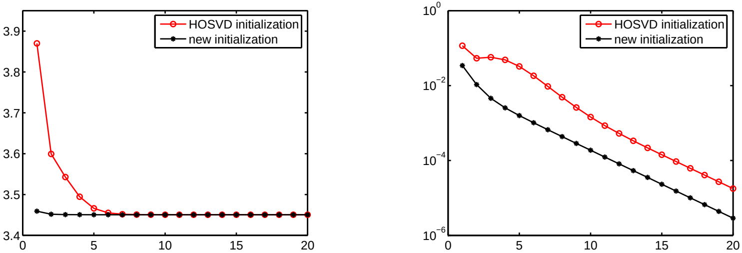
HOPM AND NEW INITIALIZATION

- **Minimization problem**: $\|\mathcal{A} - u_1 \circ u_2 \circ \ldots \circ u_d\|^2 \rightarrow \min$
- **Higher order power method** ($\sim$ 2002)
- ALS algorithm,
  in each microiteration one vector u_k, $1 \leq k \leq d$, is updated
- Converges to local minima

- Standard initialization by HOSVD: $u_k = \mathbf{U}_k(:,1),\ 1 \leq k \leq d$.
- New initialization for $d = 4$

Form square matricization $\mathbf{A}_{(1,2)}$.
Compute eigenvector $v \in \mathbb{R}^{n^2}$ belonging to largest eigenvalue of $\mathbf{A}_{(1,2)}$.
Form $\bar{\mathbf{V}} \in \mathbb{R}^{n \times n}$.
Compute $\bar{\mathbf{V}} = \mathbf{U}\Sigma\mathbf{V}^T$.
$u_i = \mathbf{U}(:,i),\ 1 \leq i \leq 4$.

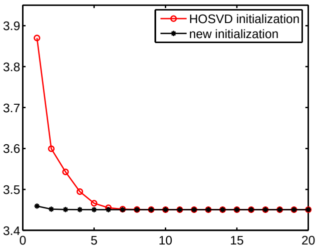
$d = 4$: New initialization - Convergence



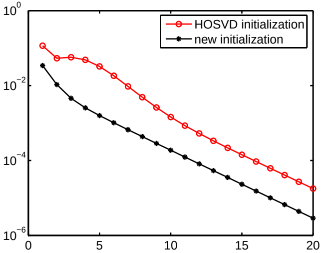
(a) Approximation error (b) Norm of gradient

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

$d = 4$: New initialization - Convergence



(a) Approximation error



(b) Norm of gradient

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

THANK YOU!