Dekompozicije i aproksimacije tenzora

Erna Begović Kovač

Sveučilište u Zagrebu FKIT

ebegovic@fkit.unizg.hr

European Women in Mathematics

Znanstveni kolokvij Zdenka Blašković-Makanec

Zagreb, 11. siječnja 2024.

HRZZ Croatian Science Foundation

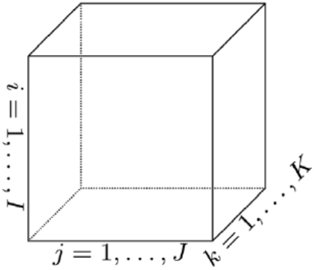
This work has been supported in part by Croatian Science Foundation under the project **UIP-2019-04-5200**.

OUTLINE

- Introduction
- Tucker decomposition
- Multilinear rank
- CP decomposition
- Tensor rank
- Tensor diagonalization
- CUR decomposition

INTRODUCTION

- vector a $\rightarrow$ matrix A $\rightarrow$ tensor $\mathcal{A}$
- A tensor is a multidimensional finite array.
- Computational challenges $\rightarrow$ curse of dimensionality



$I \times J \times K$ tensor $\mathcal{X}$

INTRODUCTION

The beginnings of tensor decompositions:

F. L. Hitchcock: *The expression of a tensor or a polyadic as a sum of products*. J. Math. Phys. 6 (1927) 164–189.

F. L. Hitchcock: *Multilple invariants and generalized rank of a p-way matrix or tensor*. J. Math. Phys. 7 (1927) 39–79.

L. R. Tucker: *Some mathematical notes on three-mode factor analysis*. Psychometrika 31 (1966) 279–311.

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.

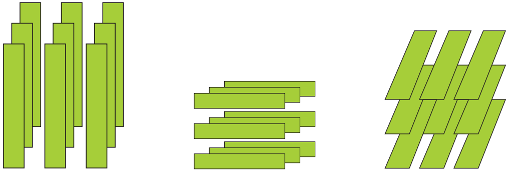
R. A. Harshman: *Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis*. UCLA Working Papers in Phonetics 16 (1970) 1–84.

Modern applications:

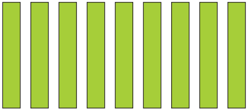
psychometrics, chemometrics, signal processing, computer vision, data mining, neuroscience, graph analysis, etc.

INTRODUCTION

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



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



INTRODUCTION

- The **mode-m product** of $\mathcal{X}$ and A is a tensor

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

- Properties:

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

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

- 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

- **Singular value decomposition (SVD)** of a matrix $A \in \mathbb{R}^{m \times n}$,

$$A = U \Sigma V^T,$$

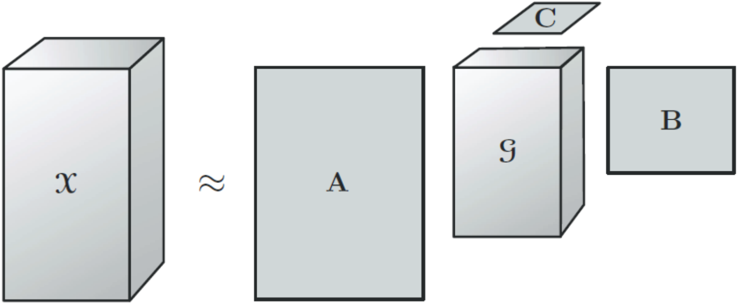
  where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal and $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$, $p = \min\{m, n\}$, such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$.

- The best low-rank application (**Eckhart-Young theorem**) of a matrix A is obtained by truncated SVD.

TUCKER DECOMPOSITION

Tucker decomposition of $\mathcal{X}$

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



HIGHER ORDER SVD

HOSVD is a special case of the Tucker decomposition

$$\mathcal{A} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 \cdots \times_d U_d,$$

where U_i, $1 \leq i \leq d$, are orthogonal matrices obtained from

$$A_{(i)} = U_i \Sigma_i V_i^T.$$

MULTILINEAR RANK

Multilinear rank of a tensor $\mathcal{A}$ is an d-tuple $(r_1, r_2, \ldots, r_d)$ where

$$r_i = \text{rank}(A_{(i)}), \quad 1 \leq i \leq d.$$

MULTILINEAR RANK

Multilinear rank of a tensor $\mathcal{A}$ is an d-tuple $(r_1, r_2, \ldots, r_d)$ where

$$r_i = \text{rank}(A_{(i)}), \quad 1 \leq i \leq d.$$

Low-rank multilinear rank approximations:

Minimization problem

$$\min \|\mathcal{A} - \tilde{\mathcal{A}}\|,$$

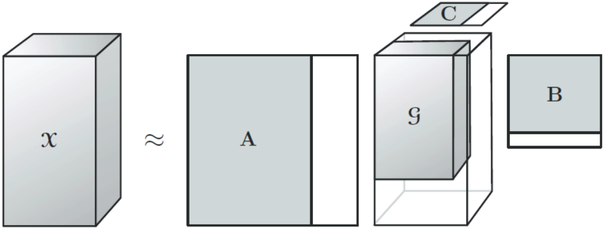
where

$$\tilde{\mathcal{A}} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 \cdots \times_d U_d$$

has multilinear rank $(r_1, r_2, \ldots, r_d)$.

LOW MULTILINEAR RANK APPROXIMATION

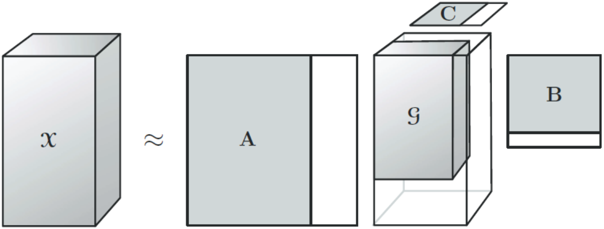
Truncated HOSVD



Eckhart-Young theorem does <u>not</u> hold for tensors.

LOW MULTILINEAR RANK APPROXIMATION

Truncated HOSVD



Eckhart-Young theorem does not hold for tensors.

- **Higher order orthogonal iterations** (HOOI)
- **Jacobi methods** for symmetric and antisymmetric tensors

LOW MULTILINEAR RANK APPROXIMATION

$$\min_{\mathcal{S},U_1,\ldots,U_d} \|\mathcal{A} - \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 \cdots \times_d U_d\|$$

It can be shown that

$$\|\mathcal{A}-\mathcal{S}\times_1 U_1 \times_2 U_2 \times_3 \cdots \times_d U_d\|^2 = \|\mathcal{A}\|^2 - \|\mathcal{A} \times_1 U_1^T \times_2 U_2^T \times_3 \cdots \times_d U_d^T\|^2$$

$\rightarrow$ Maximization problem:

$$\max_{U_1,\ldots,U_d} \|\mathcal{A} \times_1 U_1^T \times_2 U_2^T \times_3 \cdots \times_d U_d^T\|$$

HOOI

HOOI is an **alternating least squares** (ALS) algorithm:

One iteration is made of d microiterations
where matrix U_i is optimized, $1 \leq i \leq d$,
while the other matrices are fixed.

L. De Lathauwer, B. De Moor, J. Vandewalle: *On the best rank-*1 *and rank-*$(R_1, R_2, \ldots, R_N)$ *approximation of higher-order tensors*. SIAM J. Matrix Anal. Appl. 21(4) (2000) 1324–1342.

JACOBI ALGORITHM

- $\mathcal{A} \in \mathbb{R}^{n\times n\times n}$ symmetric:

$$a_{ijk} = a_{ikj} = a_{jki} = a_{jik} = a_{kij} = a_{kji}$$

$$\Rightarrow \quad A_{(1)} = A_{(2)} = A_{(3)}$$

- $\mathcal{A} \in \mathbb{R}^{n\times n\times n}$ antisymmetric:

$$a_{ijk} = -a_{ikj} = a_{jki} = -a_{jik} = a_{kij} = -a_{kji}$$

$$\Rightarrow \quad A_{(1)} = -A_{(2)} = A_{(3)}$$

In both cases $\mathrm{rank}(A_{(1)}) = \mathrm{rank}(A_{(2)}) = \mathrm{rank}(A_{(3)})$ and multilinear rank of $\mathcal{A}$ is (R, R, R).

M. Ishteva, P.-A. Absil, P. Van Dooren: *Jacobi algorithm for the best low multilinear rank approximation of symmmetric tensors*. SIAM J. Matrix Anal. Appl. 34(2) (2013) 651–672.

EBK, D. Kressner: *Structure-preserving low multilinear rank approximation of antisymmetric tensors*. SIAM. J. Matrix Anal. Appl. 38(3) (2017) 967–983.

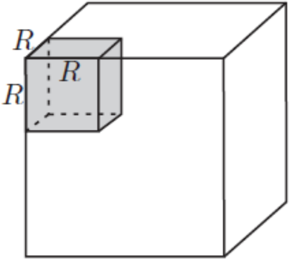
JACOBI ALGORITHM

Structure-preserving algorithm

Objective function:

$$f(U) = \|\mathcal{A} \times_1 U^T \times_2 U^T \times_3 U^T\| \to \max$$

Main idea: Apply plane rotations on the given tensor $\mathcal{A}$ in order to increase the norm of its $(R \times R \times \cdots \times R)$ subtensor with smallest indices.



JACOBI ALGORITHM

Iterative process

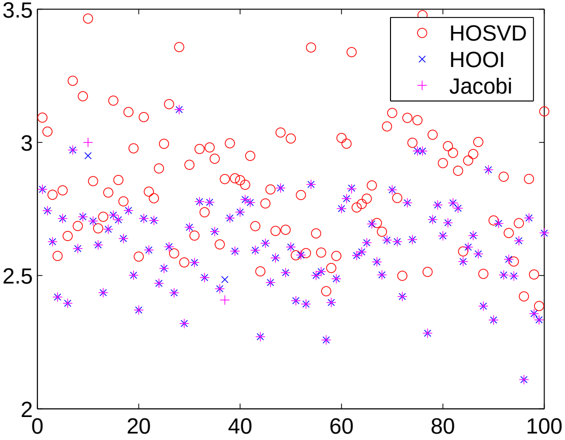
$$\mathcal{A}^{(k+1)} = \mathcal{A}^{(k)} \times_1 R_k \times_2 R_k \times_3 R_k, \quad \mathcal{A}^{(0)} = \mathcal{A},$$

$$R_k = R(i_k, j_k, \phi_k) = \begin{bmatrix} I & & & & \\ & \cos\phi & & -\sin\phi & \\ & & I & & \\ & \sin\phi & & \cos\phi & \\ & & & & I \end{bmatrix} \begin{matrix} \\ i \\ \\ j \\ \\ \end{matrix}$$

- Pivot strategy — choice of (i_k, j_k)
- Rotation angle ϕ_k
- Convergence theorems

NUMERICAL EXAMPLES 1

Approximation error: Low multilinear rank approximation of 100 random antisymmetric $10 \times 10 \times 10$ tensors.



TUCKER DECOMPOSITION – APPLICATIONS

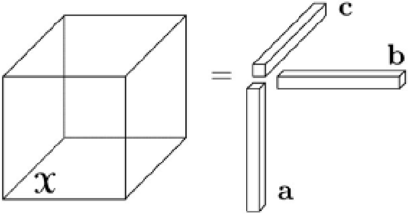
- Signal processing
- Principal component analysis (PCA)
- Computer vision (TensorFaces) — For example the data can be arranged into three modes: person, lighting conditions, and pixels. Additional modes such as expression, camera angle, and others can also be incorporated.

CP DECOMPOSITION

$\mathcal{X}$ is a **rank-1 tensor** if it can be written as the outer product of d vectors,

$$\mathcal{X} = v^{(1)} \circ v^{(2)} \circ \cdots \circ v^{(d)},$$

where $\circ$ stands for the outer product.

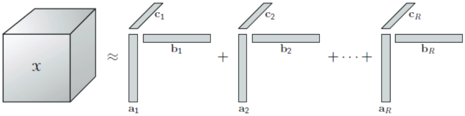


$$x_{ijk} = a_i b_j c_k$$

CP DECOMPOSITION

CANDECOMP/PARAFAC (CP) decomposition

$\rightarrow$ the sum of a finite number of rank-one tensors



$$\mathcal{X} \approx \sum_{i=1}^{R} a_i \circ b_i \circ c_i \equiv [[A, B, C]],$$

where $A = \begin{bmatrix} a_1 & a_2 & \cdots & a_R \end{bmatrix}$, etc.

H. A. L. Kiers: *Towards a standardized notation and terminology in multiway analysis*. J. Chemometrics 14 (2000) 105–122.

TENSOR RANK

The **rank** R of $\mathcal{X}$ is the smallest number of rank-1 components in an exact CP decomposition of $\mathcal{X}$.

The definition of tensor rank is an exact analogue to the definition of matrix rank, but the properties of matrix and tensor ranks are quite different.

TENSOR RANK

- The rank of a real-valued tensor may actually be different over $\mathbb{R}$ and $\mathbb{C}$.

Example:

$$X_{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad X_{(2)} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

Over $\mathbb{R}$: $\mathcal{X} = [[A, B, C]]$,

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}, \; B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \; C = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}$$

Over $\mathbb{C}$: $\mathcal{X} = [[A, B, C]]$,

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -\imath & \imath \end{bmatrix}, \; B = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ \imath & -\imath \end{bmatrix}, \; C = \begin{bmatrix} 1 & 1 \\ -\imath & \imath \end{bmatrix}$$

TENSOR RANK

- There is no straightforward algorithm to determine the rank of a specific given tensor.

An example from 1989 of a particular $9 \times 9 \times 9$ tensor whose rank is only known to be bounded between 18 and 23. Conjecture from 2009 that the rank is 19 or 20.

In practice, the rank of a tensor is determined numerically by fitting various rank-R CP models.

TENSOR RANK

The **maximum** rank is defined as the largest attainable rank.

The **typical** rank is any rank that occurs with probability greater than zero.

TENSOR RANK

The **maximum** rank is defined as the largest attainable rank.

The **typical** rank is any rank that occurs with probability greater than zero.

For the set of $m \times n$ matrices, the maximum and typical ranks are identical and equal to $\min\{m, n\}$.

For the set of $n_1 \times n_2 \times n_3$ tensor $\mathcal{X}$ there is only a weak upper bound on its maximum rank,

$$\operatorname{rank}(\mathcal{X}) \leq \min\{n_1 n_2, n_1 n_3, n_2 n_3\}.$$

For the typical rank of $2 \times 2 \times 2$ tensors Monte Carlo experiments reveal that the set of tensors of rank two fills about 79% of the space, while those of rank three fill 21%.

J. B. Kruskal: *Rank, decomposition, and uniqueness for 3-way and N-way arrays*. in Multiway Data Analysis, R. Coppi and S. Bolasco, eds., North-Holland, Amsterdam, 1989, 7–18.

COMPUTING THE CP DECOMPOSITION

For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ find

$$\min \|\mathcal{A} - \tilde{\mathcal{A}}\|, \quad \tilde{\mathcal{A}} = [[X, Y, Z]] = \sum_{i=1}^{R} x_i \circ y_i \circ z_i.$$

CP ALS algorithm: Optimization in one mode in each microiteration.

CP ALS

It can be shown that $\mathcal{A} = [[X, Y, Z]]$ implies

$$A_{(1)} = X(Z \odot Y)^T$$

and similar for other modes.

Here $\odot$ stands for the **Khatri-Rao product** of two matrices A and B,

$$A \odot B = \left[\begin{array}{cccc} a_1 \otimes b_1 & a_2 \otimes b_2 & \cdots & a_n \otimes b_n \end{array} \right],$$

where a_i and b_i are columns of A and B, respectively, and $\otimes$ is the Kronecker product.

Then, in the first mode we are looking for

$$\min_{\tilde{X}} \|A_{(1)} - \tilde{X}(Z \odot Y)^T\|, \quad \text{etc.}$$

LOW-RANK APPROXIMATION

CP ALS is one of the algorithms used for the low-rank approximation.

There are also structure-preserving variants of this approach.

For example, for $\mathcal{A}$ antisymmetric solve

$$\min \|\mathcal{A} - \tilde{\mathcal{A}}\|, \quad \text{rank}(\tilde{\mathcal{A}}) = r, \quad \tilde{\mathcal{A}} \text{ antisymmetric.}$$

EBK, L. Periša: *CP decomposition and low-rank approximation of antisymmetric tensors*. arXiv:2212.13389 [math.NA]

LOW-RANK APPROXIMATION — ANTISYM TENSORS

The simplest antisymmetric tensor is

$$\mathcal{E}(i_1, i_2, i_3) = \begin{cases} 1, & \text{if the indices make an even permutation of } (1,2,3), \\ -1, & \text{if the indices make an odd permutation of } (1,2,3), \\ 0, & \text{if two or more indices are equal.} \end{cases}$$

Tensor $\mathcal{E}$ is called the *Levi-Civita tensor*

$$E_{(1)} = \left[\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

LOW-RANK APPROXIMATION — ANTISYM TENSORS

For three given vectors x, y, z we define the antisymmetric tensor associated to these vectors as

$$\mathcal{A}_6(x,y,z) := \frac{1}{6}(x \circ y \circ z + y \circ z \circ x + z \circ x \circ y - x \circ z \circ y - y \circ x \circ z - z \circ y \circ x).$$

$\mathcal{E}$ is a special case of the antisymmetric tensor $\mathcal{A}_6(x,y,z)$. For $x = [6,0,0]^T$, $y = [0,1,0]^T$, $z = [0,0,1]^T$, we get $\mathcal{A}_6(x,y,z) = \mathcal{E}$.

Then we are looking for $\tilde{\mathcal{A}}$ such that

$$\tilde{\mathcal{A}} = \mathcal{A}_6(x,y,z).$$

CP DECOMPOSITION – APPLICATIONS

- CP originates from psychometrics — analyzing multiple similarities from a variety of subjects
- Chemometrics — modeling of fluorescence excitation-emission data
- Neuroscience — brain imaging, analysis of MRI data and EEG data
- Image processing — compression and classification

TENSOR DIAGONALIZATION

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

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

Diagonalization problem: find orthogonal transformations U, V, W that “diagonalize” $\mathcal{A}$.

TENSOR DIAGONALIZATION

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

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

Diagonalization problem: find orthogonal transformations U, V, W that "diagonalize" $\mathcal{A}$.

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

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

TENSOR DIAGONALIZATION

Jacobi-type algorithms

- C. D. Moravitz Martin, C. F. Van Loan: *A Jacobi-type method for computing orthogonal tensor decompositions*. SIAM J. Matrix Anal. Appl. 30(3) (2008) 1219–1232.
- J. Li, K. Usevich, P. Comon: *Globally Convergent Jacobi-Type Algorithms for Simultaneous Orthogonal Symmetric Tensor Diagonalization*. SIAM J. Matrix Anal. Appl. 39(1) (2018) 1–22.
- J. Li, K. Usevich, P. Comon: *On approximate diagonalization of third order symmetric tensors by orthogonal transformations*. Linear Algebra Appl. 576(1) (2019) 324–351.
- K. Usevich, J. Li, P. Comon: *Approximate matrix and tensor diagonalization by unitary transformations: Convergence of Jacobi-type algorithms*. SIAM J. Optim. 30(4) (2020) 2998–3028.
- EBK: *Convergence of a Jacobi-type method for the approximate orthogonal tensor diagonalization*. Calcolo 60, 3 (2023)
- EBK, A. Perković: *Trace maximization algorithm for the approximate tensor diagonalization*. arXiv:2111.01466 [math.NA] Aaccepted for publication in Linear and Multilinear Algebra

TENSOR DIAGONALIZATION

Two approaches:

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

$$\text{off}(\mathcal{S}) \to \min,$$

where

$$\text{off}^2(\mathcal{X}) = \|\mathcal{X}\|^2 - \|\text{diag}(\mathcal{X})\|^2.$$

Dually, the Frobenius norm of the diagonal is maximized,

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

- ♠ Tensor-trace of $\mathcal{S}$ is maximized

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

TENSOR DIAGONALIZATION

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

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

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

$$g(U, V, W) = \mathrm{tr}(\mathcal{A} \times_1 U^T \times_2 V^T \times_3 W^T).$$

Main idea: Apply plane rotations on the given tensor $\mathcal{A}$ in order to increase its diagonal.

JACOBI ALGORITHM

Iterative process

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

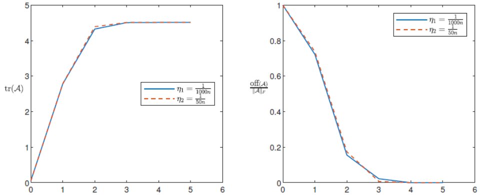
where $R_{U,k}, R_{V,k}, R_{W,k}$ are plane rotations of the form

$$R(i_k, j_k, \phi_k) = \begin{bmatrix} I & & & & \\ & \cos\phi & & -\sin\phi & \\ & & I & & \\ & \sin\phi & & \cos\phi & \\ & & & & I \end{bmatrix} \begin{matrix} \\ i \\ \\ j \\ \\ \end{matrix}$$

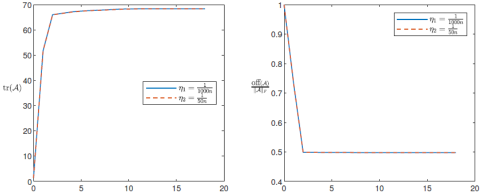
- Pivot strategy — choice of (i_k, j_k)
- Rotation angle ϕ_k
- Convergence theorems

NUMERICAL EXAMPLES 2

Change in the trace and the relative off-norm



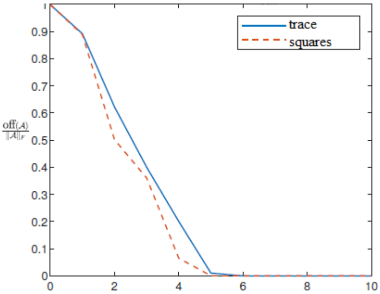
Diagonalizable $10 \times 10 \times 10 \times 10$ tensor



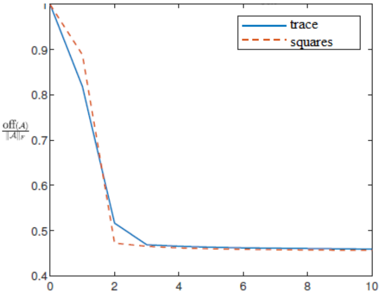
Non-diagonalizable $5 \times 5 \times 5 \times 5 \times 5 \times 5$ tensor

NUMERICAL EXAMPLES 3

Trace maximization compared to the maximization of the squares of the diagonal elements for two random $20 \times 20 \times 20$tensors.



Tensor 1



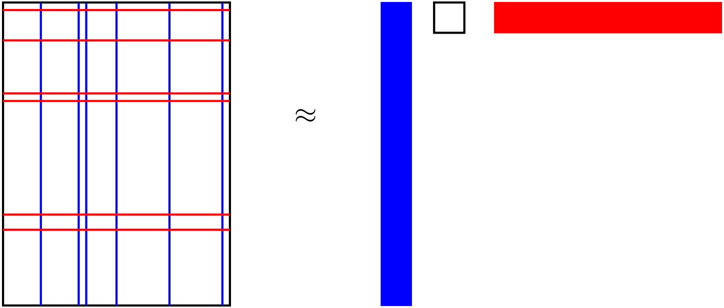
Tensor 2

CUR DECOMPOSITION

A matrix **CUR factorization** of $A \in \mathbb{C}^{m \times n}$ is decomposition of the form

$$A = CUR,$$

where $C \in \mathbb{C}^{m \times k}$ contains k columns of A, $R \in \mathbb{C}^{k \times n}$ contains k rows of A, and $U \in \mathbb{C}^{k \times k}$.



CUR DECOMPOSITION

A **CUR decomposition** of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times \cdots \times n_d}$ is given by

$$\mathcal{A} \approx \mathcal{S} \times_1 C_1 \times_2 C_2 \times_3 \cdots \times_d C_d,$$

where $\mathcal{S} \in \mathbb{C}^{r_1 \times r_2 \times \cdots \times r_d}$ is a core tensor and matrices $C_j \in \mathbb{C}^{n_j \times r_j}$, $1 \leq j \leq d$, contain r_j mode-j fibers of $\mathcal{A}$.

CUR DECOMPOSITION

Problem: choice of the preserved fibers.

- Deterministic algorithms
- Probabilistic algorithms
- Hybrid approach between CUR and T-HOSVD

M. W. Mahoney, M. Maggioni, P. Drineas: *Tensor-CUR decompositions for tensor-based data.* SIAM J. Matrix Anal. Appl. 30(3) (2008) 957–987.

A. K. Saibaba: HOID: Higher order interpolatory decomposition for tensors based on Tucker representation. SIAM J. Matrix Anal. Appl. 37(3) (2016) 1223–1249.

EBK: *Hybrid CUR-type decomposition of tensors in the Tucker format*. BIT Numer. Math. 62(1) (2022) 125–138.

SUMMARY

- SVD: $A = U\Sigma V^T$
- HOSVD: $\mathcal{A} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 \cdots \times_d U_d$
- CP: $\mathcal{X} \approx \sum_{i=1}^{R} a_i \circ b_i \circ c_i$
- Tensor diagonalisation: maximization of the diagonal (Frobenius norm or trace)
- CUR decomposition: preserving original fibers

T. G. Kolda, B. W. Bader: *Tensor decompositions and applications*. SIAM Rev. 51(3) (2009) 455–500.

Hvala na pažnji!