Tensor-based algorithms: applications and challenges

Martina Iannacito

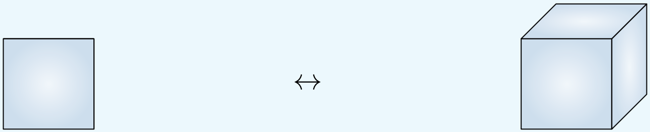
APO team seminar
Toulouse, December 19, 2023

KU LEUVEN kulak
STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics
European Research Council
Established by the European Commission

Overview

- Tensor preliminary
- Master's thesis
  - Biodiversity from spectral images
  - Results with Tucker model
- Doctoral thesis
  - Numerical linear algebra
  - Results with Tensor-train model
- Postdoctoral project
  - Canonical Polyadic Decomposition
  - New CPD algorithm
- Conclusion

From scalars to tensors



Matrix

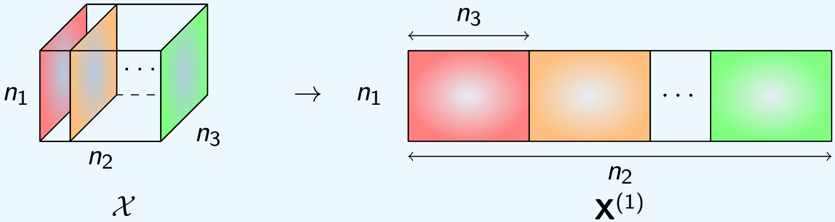
- object in $\mathbb{K}^{n_1 \times n_2}$
- set of n_2 elements in $\mathbb{K}^{n_1}$
- linear operator from $\mathbb{K}^{n_2}$ to $\mathbb{K}^{n_1}$

Tensor

- object in $\mathbb{K}^{n_1 \times \cdots \times n_d}$
- set of $(n_{i_1} \times n_{i_k})$ elements in $\mathbb{K}^{n_{j_1} \times \cdots \times n_{j_\ell}}$
- multilinear operator from $\mathbb{K}^{n_{j_1} \times \cdots \times n_{j_\ell}}$ to $\mathbb{K}^{n_{i_1} \times \cdots \times n_{i_k}}$ with $k + \ell = d$

Unfolding

Let $\mathcal{X}$ be a 3-order tensor of size $(n_1 \times n_2 \times n_3)$



Definition: the 1-st mode matricization $\mathbf{X}^{(1)}$ is a $(n_1 \times n_1 n_2)$ matrix, obtained stacking the vectors

$$\mathbf{x}_{i_1} = \text{vec}(\mathcal{X}(i_1, \cdot, \cdot)).$$

Matrix-tensor product

Let $\mathcal{X}$ be a $(n_1 \times n_2 \times n_3)$ tensor. If $\mathbf{G}$ is a $(n_1 \times m_1)$ matrix

Definition: the 1st mode matrix-tensor product

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{G}$$

a $(m_1 \times n_2 \times n_3)$ such that

$$\mathcal{Y}(j_1, i_2, i_3) = \sum_{i_1=1}^{n_1} \mathcal{X}(i_1, i_2, i_3)\mathbf{G}(i_1, j_1)$$

Computational: a straightforward way of getting $\mathcal{Y}$ the matrix-tensor product is computing

$$\mathbf{Y} = \mathbf{G}^T \mathbf{X}^{(1)}$$

and then tensorizing $\mathbf{Y}$ into $\mathcal{Y}$.

Overview

Tensor preliminary

Master's thesis

- Biodiversity from spectral images
- Results with Tucker model

Doctoral thesis

- Numerical linear algebra
- Results with Tensor-train model

Postdoctoral project

- Canonical Polyadic Decomposition
- New CPD algorithm

Conclusion

Master's people

Supervisors

Figure: Prof. A. Bernardi, University of Trento

Figure: Prof. D. Rocchini, University of Bologna

- algebraic geometry
- algorithms for tensor decomposition

- plant ecology
- algorithms for biodiveristy

Master's project

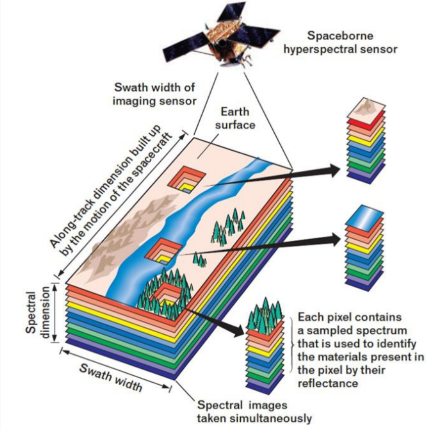


Figure: from [Bedini 2017].

Over a time series of Europe spectral images,

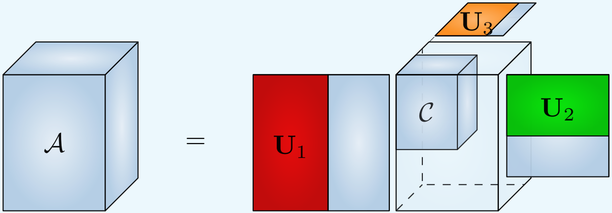
- get two images from two spectral bands (RED and NIR);
- compute the normalized difference vegetation index per pixel, i.e.,

$$\mathtt{NDVI}(i,j) = \frac{\mathtt{NIR}(i,j) - \mathtt{RED}(i,j)}{\mathtt{NIR}(i,j) + \mathtt{RED}(i,j)}$$

- compute a biodiversity index over the resulting NDVI image

What happens if the NDVI image is computed from the NIR and RED spectral images stored in a tensor and compressed?

Tucker's model [Tucker 1966; De Lathauwer et al. 2000]



If $\mathcal{A}$ is a $(n_1 \times n_2 \times n_3)$ tensor, its Tucker decomposition becomes

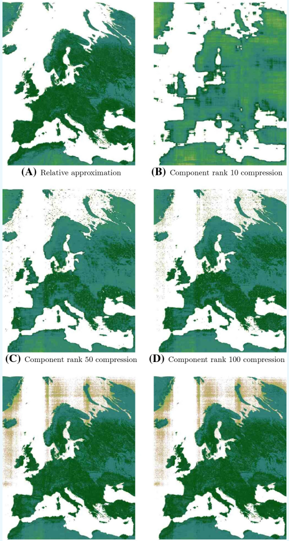
$$\mathcal{A} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$$

where

- $\mathcal{C}$ is a $(r_1 \times r_2 \times r_3)$ tensor;
- $\mathbf{U}_i$ is a $(n_i \times r_i)$ orthogonal matrix, called i-th factor matrix.

The memory requirement is $\mathcal{O}(r^d + nr)$ where $r = \max r_i$, $n = \max n_i$ and d is the tensor order.

Rényi index result [Bernardi et al. 2019]



(A) Relative approximation (B) Component rank 10 compression (C) Component rank 50 compression (D) Component rank 100 compression (E) Component rank 500 compression (F) Component rank 1000 compression

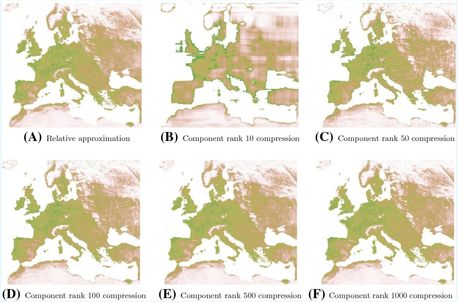
Rényi index

Uses only pixel value frequencies

Compression at multilinear rank $(i, i, 3)$ with $i \in \{10, 50, 100, 500, 1000\}$

- memory used ranges between 0.19% and 22%;
- average error per pixel ranges between 13% and 5%.

Rao index result [Bernardi et al. 2019]



Rao index

Uses only pixel values and their frequencies

Compression at multilinear rank $(i, i, 3)$ with $i \in \{10, 50, 100, 500, 1000\}$

- memory used ranges between 0.19% and 22%;
- average error per pixel ranges between 63% and 19%.

Overview

- Tensor preliminary
- Master's thesis
  - Biodiversity from spectral images
  - Results with Tucker model
- Doctoral thesis
  - Numerical linear algebra
  - Results with Tensor-train model
- Postdoctoral project
  - Canonical Polyadic Decomposition
  - New CPD algorithm
- Conclusion

Ph.D. people

Supervisors

Figure: Prof. O. Coulaud, Inria Bordeaux

- tensor methods
- high-dimensional simulations

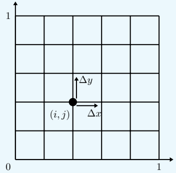
Figure: Prof. L. Giraud, Inria Bordeaux (usually in Toulouse)

- numerical linear algebra
- finite precision arithmetic

Context

The problem

$$\begin{cases} \mathcal{L}(u) & = f \quad \text{in} \quad \Omega \\ u & = f_0 \quad \text{in} \quad \partial\Omega \end{cases} \qquad \text{for} \qquad \Omega \subseteq \mathbb{R}^{n_1 \times \cdots \times n_d}.$$

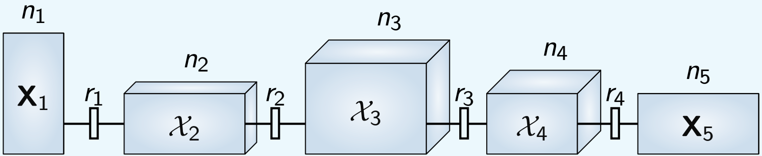


$$\mathcal{A}\mathcal{X} = \mathcal{B}$$

where $\mathcal{A} : \mathbb{R}^{n_1 \times \cdots \times n_d} \to \mathbb{R}^{n_1 \times \cdots \times n_d}$ is a multilinear operator and $\mathcal{B} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ a tensor.
For large scale-simulations we have to take into account

- memory costs $\mathcal{O}(n^d)$
- computational model
- numerical method

Tensor Train or Matrix Product States [Oseledets 2011]



Let $\mathcal{X}$ a tensor of order d and dimensions $(n_1 \times \cdots \times n_d)$, then its TT-representation is given by d TT-cores s.t.

- $\mathbf{X}_1$ a (n_1, r_1) matrix
- $\mathcal{X}_i$ is a $(r_{i-1} \times n_i \times r_i)$ tensor
- $\mathbf{X}_d$ is a $(r_{d-1} \times n_d)$ matrix

i.e., a *train of matrix - third-order tensors - matrix*.

The $(i_1, \ldots, i_d)$ element of $\mathcal{X}$ is

$$\mathcal{X}(i_1, \ldots, i_d) = \sum_{i=1}^{d} \sum_{s_i=1}^{r_i} \mathbf{X}(i_1, s_1)\mathcal{X}_1(s_1, i_2, s_2) \cdots \mathbf{X}_d(s_{d-1}, i_d).$$

The memory cost is $\mathcal{O}(dr^2 n)$ where $r = \max r_i$, $n = \max n_i$ and d is the tensor order.

New variable accuracy approach

What happens if objects are compressed by TT-format at computational level?

Assumptions

- compress **tensors** at accuracy δ with TT-format
- store matrices and vectors at accuracy u from standard IEEE model
- perform operation at accuracy u from standard IEEE model

new 'mixed'-precision framework

$$fl_\delta(\mathcal{X} \operatorname{op} \mathcal{Y}) = \delta\text{-storage}\big(fl(\mathcal{X} \operatorname{op} \mathcal{Y})\big)$$

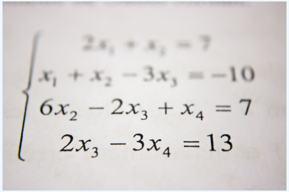
$$\delta\text{-storage}(\mathcal{Z}) = \overline{\mathcal{Z}} \qquad \text{s.t.} \qquad \frac{||\mathcal{Z} - \overline{\mathcal{Z}}||}{||\mathcal{Z}||} \leq \delta$$

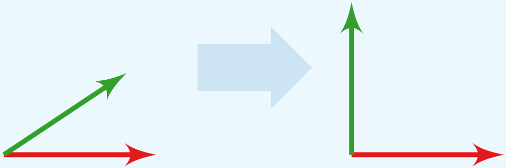
with *fl* is the classical floating point computational function dependent on u.

Numerical linear algebra methods

Iterative solver

- Generalized Minimal RESidual (GMRES)



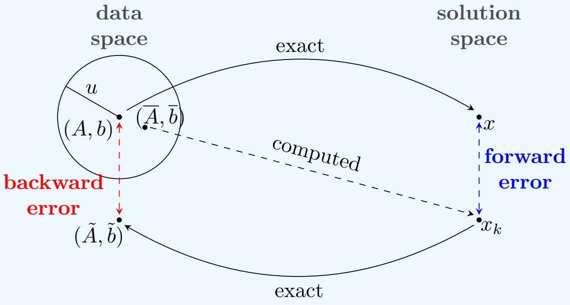


Orthogonalization kernels

- Classical and Modified Gram-Schmidt (CGS, MGS)
- Gram approach
- Householder transformation

GMRES property [Wilkinson 1963]

Given the linear system $\mathbf{Ax} = \mathbf{b}$ and a working precision u, then



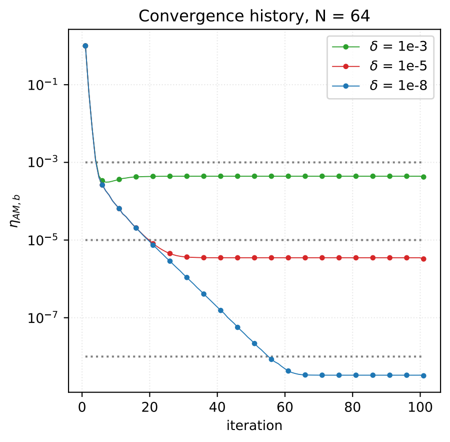
GMRES is backward stable, i.e.,

$$\eta_{A,b}(x_k) = \frac{||\mathbf{A}\mathbf{x}_k - \mathbf{b}||}{||\mathbf{A}||||\mathbf{x}_k|| + ||\mathbf{b}||} \sim \mathcal{O}(u)$$

TT-GMRES results [Dolgov 2013; Coulaud et al. 2022a]

Convection-Diffusion problem

$$\begin{cases} -\Delta \mathcal{U} & +\mathcal{V}\cdot\nabla\mathcal{U} = 0 \\ \mathcal{U}_{\{y=1\}} & = 1 \end{cases} \quad \text{in} \quad \Omega = [-1,1]^3$$



Orthogonalization schemes

Let $\mathbf{Q}_k = [\mathbf{q}_1, \ldots, \mathbf{q}_k]$ be the orthogonal basis produced by an orthogonalization kernel, then the **Loss Of Orthogonality** is

$$||\mathbb{I}_k - \mathbf{Q}_k^\top \mathbf{Q}_k||.$$

It measures the quality in terms of orthogonality of the computed basis. It is linked with the linearly dependency of the input vectors $\mathbf{A}_k = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$, estimated through $\kappa(\mathbf{A}_k)$.

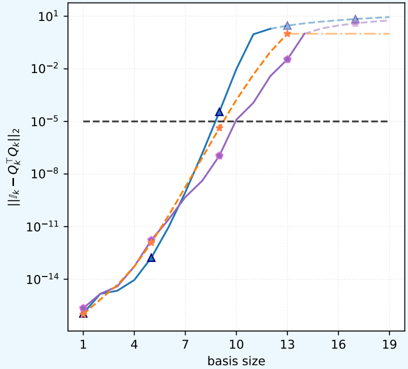
Matrix		
Source	*Algorithm*	$\left\|\left\|\mathbb{I}_k - \mathbf{Q}_k^\top \mathbf{Q}_k\right\|\right\|$
[Stathopoulos et al. 2002]	**Gram**	$\mathcal{O}(u\kappa^2(\mathbf{A}_k))$
[L. Giraud et al. 2005]	**CGS**	$\mathcal{O}(u\kappa^2(\mathbf{A}_k))$
[Björck 1967]	**MGS**	$\mathcal{O}(u\kappa(\mathbf{A}_k))$
[L. Giraud et al. 2005]	**CGS2**	$\mathcal{O}(u)$
[L. Giraud et al. 2005]	**MGS2**	$\mathcal{O}(u)$
[Wilkinson 1965]	**Householder**	$\mathcal{O}(u)$

TT-orthogonalization [Coulaud et al. 2022b]

$$\mathcal{X}_{k+1} = \texttt{TT-rounding}(\mathbf{\Delta}_d \mathcal{A}_k, \texttt{max\_rank} = 1) \quad \text{with} \quad \mathcal{A}_{k+1} = \frac{1}{||\mathcal{X}_{k+1}||}\mathcal{X}_{k+1}$$

TT-orthogonalization [Coulaud et al. 2022b]

$$\mathcal{X}_{k+1} = \texttt{TT-rounding}(\mathbf{\Delta}_d \mathcal{A}_k, \texttt{max\_rank} = 1) \quad \text{with} \quad \mathcal{A}_{k+1} = \frac{1}{||\mathcal{X}_{k+1}||}\mathcal{X}_{k+1}$$



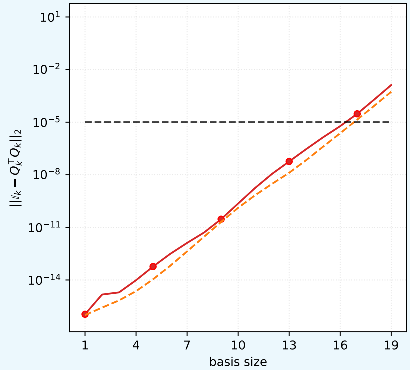
- **Gram approach**
- **CGS**
- $\kappa^2(\mathbf{A}_k)$

$\mathcal{O}(\delta \kappa^2(\mathbf{A}_k))$

Figure: Loss of orthogonality for $m = 20$ TT-vectors of order $d = 3$ and mode size $n = 15$, rounding precision $\delta = 10^{-5}$

TT-orthogonalization [Coulaud et al. 2022b]

$$\mathcal{X}_{k+1} = \texttt{TT-rounding}(\boldsymbol{\Delta}_d \mathcal{A}_k, \texttt{max\_rank} = 1) \quad \text{with} \quad \mathcal{A}_{k+1} = \frac{1}{||\mathcal{X}_{k+1}||}\mathcal{X}_{k+1}$$



- **MGS**
- $\kappa(\mathbf{A}_k)$

$\mathcal{O}(\delta\kappa(\mathbf{A}_k))$

Figure: Loss of orthogonality for $m = 20$ TT-vectors of order $d = 3$ and mode size $n = 15$, rounding precision $\delta = 10^{-5}$

TT-orthogonalization [Coulaud et al. 2022b]

$$\mathcal{X}_{k+1} = \texttt{TT-rounding}(\mathbf{\Delta}_d \mathcal{A}_k, \texttt{max\_rank} = 1) \quad \text{with} \quad \mathcal{A}_{k+1} = \frac{1}{||\mathcal{X}_{k+1}||}\mathcal{X}_{k+1}$$

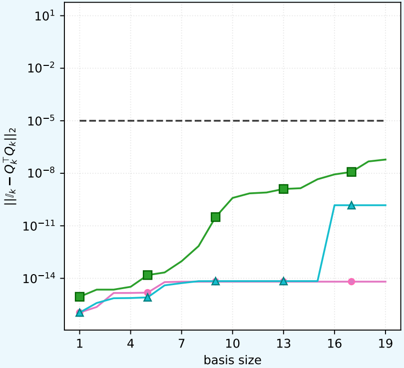


Figure: Loss of orthogonality for $m = 20$ TT-vectors of order $d = 3$ and mode size $n = 15$, rounding precision $\delta = 10^{-5}$

- **CGS2**
- **MGS2**
- **Householder transformation**

$\mathcal{O}(\delta)$

TT-orthogonalization [Coulaud et al. 2022b]

$$\mathcal{X}_{k+1} = \texttt{TT-rounding}(\boldsymbol{\Delta}_d \mathcal{A}_k, \texttt{max\_rank} = 1) \quad \text{with} \quad \mathcal{A}_{k+1} = \frac{1}{||\mathcal{X}_{k+1}||}\mathcal{X}_{k+1}$$

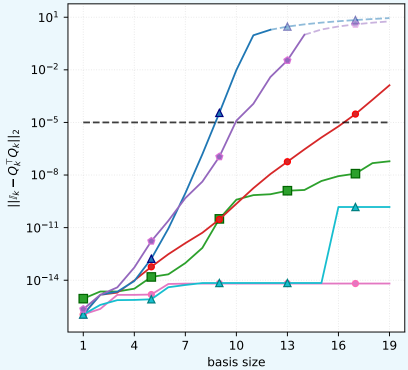


Figure: Loss of orthogonality for $m = 20$ TT-vectors of order $d = 3$ and mode size $n = 15$, rounding precision $\delta = 10^{-5}$

- **Gram approach** $\mathcal{O}(\delta\kappa^2(\mathbf{A}_k))$
- **CGS** $\mathcal{O}(\delta\kappa^2(\mathbf{A}_k))$
- **MGS** $\mathcal{O}(\delta\kappa(\mathbf{A}_k))$
- **CGS2** $\mathcal{O}(\delta)$
- **MGS2** $\mathcal{O}(\delta)$
- **Householder transformation** $\mathcal{O}(\delta)$

Overview

- Tensor preliminary
- Master's thesis
  - Biodiversity from spectral images
  - Results with Tucker model
- Doctoral thesis
  - Numerical linear algebra
  - Results with Tensor-train model
- Postdoctoral project
  - Canonical Polyadic Decomposition
  - New CPD algorithm
- Conclusion

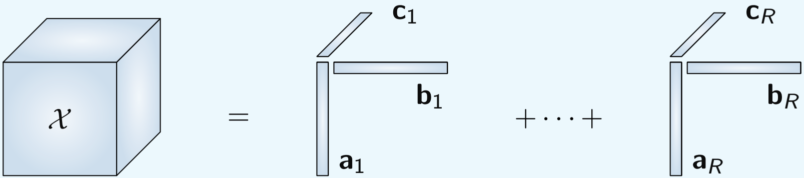
Postdoctoral project

New algorithm for Canonical Polyadic Decomposition

- formalize previous results from I. Domanov;
- improve the algorithm efficiency;
- evaluate its quality;
- test in signal processing cases.

Figure: Prof. L. De Lathauwer, KU Leuven

Canonical Polyadic Decomposition [Hitchcock 1927; Harshman 1970; Carroll et al. 1970]



If $\mathcal{A}$ is a $(n_1 \times n_2 \times n_3)$ tensor of rank R, its CPD decomposition is

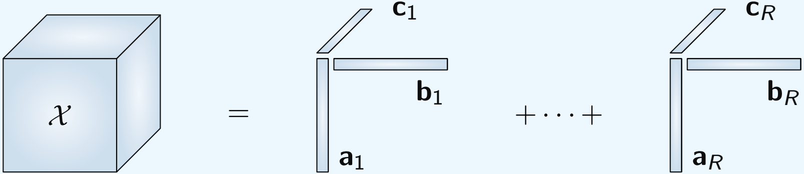
$$\mathcal{A} = \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r$$

where $\mathbf{a}_r \in \mathbb{K}^{n_1}$, $\mathbf{b}_r \in \mathbb{K}^{n_2}$ and $\mathbf{c}_r \in \mathbb{K}^{n_3}$ with $i = 1, \ldots, R$. Its properties are

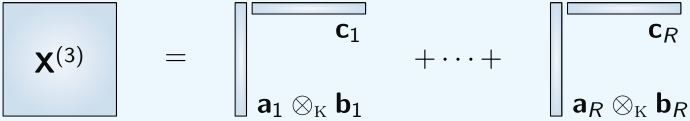
- unique under mild assumption
- memory cost $\mathcal{O}(dNR)$
- NP-hard problem
- algorithms affected by numerical instabilities

Problem reformulation

$$\text{if } \mathcal{X} = \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 + \ldots + \mathbf{a}_R \otimes \mathbf{b}_R \otimes \mathbf{c}_R$$



$$\text{then } \mathbf{X}^{(3)} = (\mathbf{a}_1 \otimes_{\mathrm{K}} \mathbf{b}_1) \otimes \mathbf{c}_1^T + \ldots + (\mathbf{a}_R \otimes_{\mathrm{K}} \mathbf{b}_R) \otimes \mathbf{c}_R^T$$



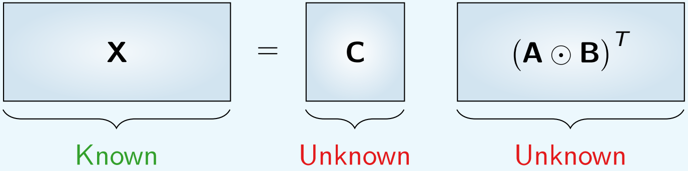
$$(\mathbf{a}_r \otimes_{\mathrm{K}} \mathbf{b}_r) \in \mathcal{V} = \left\{ \mathrm{vec}(\mathbf{Z}) : \begin{vmatrix} z_{i_1 j_1} & z_{i_1 j_2} \\ z_{i_2 j_1} & z_{i_2 j_2} \end{vmatrix} = 0 \right\} \text{ algebraic variety}$$

Algebraic algorithm: high view

Let $\mathcal{X}$ be a $(n_1 \times n_2 \times R)$ tensor, then

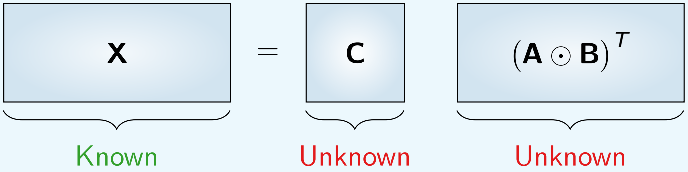
$$\mathbf{X}^{(3)} = \sum_{r=1}^{R} (\mathbf{a}_r \otimes_{\mathrm{K}} \mathbf{b}_r) \otimes \mathbf{c}_r^T = (\mathbf{A} \odot \mathbf{B})\mathbf{C}^T.$$

If $\mathbf{X} = \left(\mathbf{X}^{(3)}\right)^T$, then



1. compute $\mathbf{C}^{-1}$ columns from $\mathbf{X}$ using algebraic geometry properties;
2. compute $(\mathbf{A} \odot \mathbf{B})$ as the transposed product of $\mathbf{C}^{-1}\mathbf{X}$;
3. factorize $(\mathbf{A} \odot \mathbf{B}) = [\mathbf{a}_1 \otimes_{\mathrm{K}} \mathbf{b}_1, \ldots, \mathbf{a}_R \otimes_{\mathrm{K}} \mathbf{b}_R]$ to recover $\mathbf{A}$ and $\mathbf{B}$;
4. compute $\mathbf{C}$ by solving $(\mathbf{A} \odot \mathbf{B})\mathbf{C} = \mathbf{X}$.

Algebraic algorithm outline



1. compute the factor matrix $\mathbf{C}^{-1}$ from $\mathbf{X}$;
   - 1.1 compute $\mathbf{Q}$ describing the algebraic variety;
   - 1.2 compute the space $\mathcal{E}_0 = \text{null}(\mathbf{Q}) \cap \text{vec}(\text{Sym}_R^d)$
     - 1.2.1 if dim $\mathcal{E}_0 = R$, then compute $\mathbf{C}^{-1}$ by a CPD of $\{\mathbf{e}_1^{\otimes d}, \dots, \mathbf{e}_R^{\otimes d}\}$ basis of $\mathcal{E}_0$;
     - 1.2.2 if dim $\mathcal{E}_0 > R$, then compute $\mathcal{E}_{h+1}$ such that

$$\mathcal{E}_{h+1} = (\mathbb{K} \otimes \mathcal{E}_h) \cap \text{vec}(\text{Sym}_R^{d+h})$$

       until dim $\mathcal{E}_{h+1} = R^{h+1}$ and go to step 1.2.1;
2. compute $(\mathbf{A} \odot \mathbf{B})$ as $\mathbf{C}^{-1}\mathbf{X}$ transposed;
3. factorize each column of $(\mathbf{A} \odot \mathbf{B})$ at rank-1 to retrieve $\mathbf{A}$ and $\mathbf{B}$ by SVD;
4. compute $\mathbf{C}$ solving $(\mathbf{A} \odot \mathbf{B})^T \mathbf{C} = \mathbf{X}$.

Challenges

- efficiently construct $\mathbf{Q}$ and its kernel
- estimate the dimension of the intersection with Sym_R^{d+h}
- efficiently construct a basis for $\mathfrak{E}_h$
- compute the CPD of $\{\mathbf{e}_1^{\otimes(h+d)}, \ldots, \mathbf{e}_d^{\otimes(h+d)}\}$
- estimate the quality of the algorithm and its robustness

Overview

Tensor preliminary

Master's thesis

- Biodiversity from spectral images
- Results with Tucker model

Doctoral thesis

- Numerical linear algebra
- Results with Tensor-train model

Postdoctoral project

- Canonical Polyadic Decomposition
- New CPD algorithm

Conclusion

Wrap up

Tensor methods used in

- data analysis problem as compression methods
  - by the Tucker's decomposition
- scientific computing as new policy for computational methods
  - by the Tensor-Train decomposition
- signal processing
  - by the Canonical Polyadic Decomposition

Thank you for the attention!
Questions? Advice?

References I

- Bedini, Enton (2017). "The use of hyperspectral remote sensing for mineral exploration: a review". In: *Journal of Hyperspectral Remote Sensing*.
- Bernardi, Alessandra, Martina Iannacito, and Duccio Rocchini (2019). "High order singular value decomposition for plant diversity estimation". In: *Bollettino dell'Unione Matematica Italiana* 14, pp. 557–591.
- Björck, Åke (Mar. 1967). "Solving linear least squares problems by Gram-Schmidt orthogonalization". In: *BIT Numerical Mathematics* 7.1, pp. 1–21.
- Carroll, J. Douglas and Jih-Jie Chang (Sept. 1970). "Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition". In: *Psychometrika* 35.3, pp. 283–319.
- Coulaud, Olivier, Luc Giraud, and Martina Iannacito (2022a). *A note on GMRES in TT-format*. Research Report RR-9384. Inria Bordeaux Sud-Ouest.

References II

Coulaud, Olivier, Luc Giraud, and Martina Iannacito (Nov. 2022b). *On some orthogonalization schemes in Tensor Train format*. Tech. rep. RR-9491. Inria Bordeaux - Sud-Ouest.

De Lathauwer, Lieven, Bart De Moor, and Joos Vandewalle (2000). "A Multilinear Singular Value Decomposition". In: *SIAM Journal on Matrix Analysis and Applications* 21.4, pp. 1253–1278.

Dolgov, S. V. (2013). "TT-GMRES: solution to a linear system in the structured tensor format". In: *Russian Journal of Numerical Analysis and Mathematical Modelling* 28.2, pp. 149–172.

Giraud, L., J. Langou, and M. Rozloznik (2005). "The loss of orthogonality in the Gram-Schmidt orthogonalization process". In: *Computers & Mathematics with Applications* 50.7. Numerical Methods and Computational Mechanics, pp. 1069–1075.

References III

Harshman, Richard A. (1970). "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-model factor analysis". In: vol. 16. UCLA Working Papers in Phonetics. University Microfilms, Ann Arbor, Michigan, pp. 1–84.

Hitchcock, Frank L. (1927). "The Expression of a Tensor or a Polyadic as a Sum of Products". In: *Journal of Mathematics and Physics* 6.1-4, pp. 164–189.

Oseledets, I. V. (2011). "Tensor-Train Decomposition". In: *SIAM Journal on Scientific Computing* 33.5, pp. 2295–2317.

Stathopoulos, Andreas and Kesheng Wu (2002). "A Block Orthogonalization Procedure with Constant Synchronization Requirements". In: *SIAM Journal on Scientific Computing* 23.6, pp. 2165–2182.

Tucker, Ledyard R (1966). "Some mathematical notes on three-modes factor analysis". In: *Psychometrika* 31.3, pp. 279–311.

References IV

Wilkinson, J. H. (1963). *Rounding Errors in Algebraic Processes*. Vol. 32. Notes on Applied Science. Also published by Prentice-Hall, Englewood Cliffs, NJ, USA, 1964, translated into Polish as *Bledy Zaokragleń w Procesach Algebraicznych* by PWW, Warsaw, Poland, 1967 and translated into German as *Rundungsfehler* by Springer-Verlag, Berlin, Germany, 1969. Reprinted by Dover Publications, New York, 1994. London, UK: HMSO, pp. vi + 161.

— (1965). *The algebraic eigenvalue problem*. en. Numerical Mathematics and Scientific Computation. Oxford, England: Clarendon Press.