Potential and applications of tensor-based algorithms

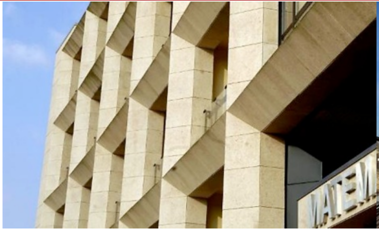
Martina Iannacito

Dipartimento di Matematica
Alma Mater Studiorum - Università di Bologna

Spring Semester Seminars in Numerical Linear Algebra and beyond

April 10, 2024

Table of Contents




Bachelor degree
UniPR
2014-2017

Ph.D.
INRIA Bordeaux
2019-2022

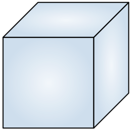
Postdoc
UniBO
2024

Master’s degree
UniTN
2017-2019

Postdoc
KU Leuven
2023-2024


From scalars to tensors

$\leftrightarrow$



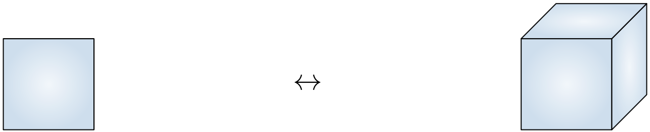
Matrix

- 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 $\mathbb{K}^{n_1}$

$\mathcal{Tensor}$

- object in $\mathbb{K}^{n_1 \times \cdots \times n_d}$
- set of $(n_{i_1} \cdots n_{i_k})$ 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 $k + \ell = d$

Where and why tensors?



Examples of tensor data

- Color images, video, ...
- Text mining: term $\times$ document $\times$ author
- (Social) networks: score $\times$ object $\times$ referee $\times$ criterion
- Ecological data: species $\times$ time $\times$ area $\times$ altitude $\times$ ...
- Face recognition: people $\times$ pose $\times$ illumination $\times$ angle

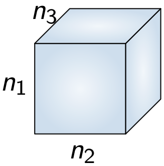
Tensor advantages

- better representation of intricate phenomena
- compression by factorization techniques
- uniqueness for some decomposition

Table of Contents

Basic definitions

Let $\mathcal{A}$ be an $(n_1 \times n_2 \times n_3)$ tensor



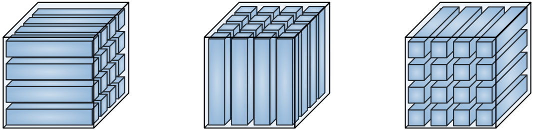
- $\{1, 2, 3\}$ are the **modes** of the tensors
- n_k is the **size** of the k-th mode
- $d = 3$ is the tensor **order**

Curse of dimensionality

The number of entries is $\mathcal{O}(n^d)$ with $n = \max n_i$

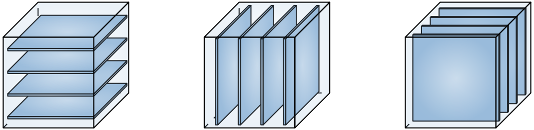
Fibers and slices

The tensor fibers are vectors extracted from the tensor fixing all indexes except one



$\mathcal{A}(\cdot, i_2, i_3)$ $\mathcal{A}(i_1, \cdot, i_3)$ $\mathcal{A}(i_1, i_2, \cdot)$

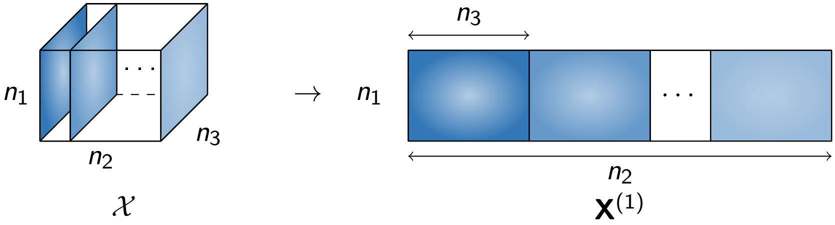
The tensor slices are matrices extracted from the tensor fixing all indexes except two



$\mathcal{A}(i_1, \cdot, \cdot)$ $\mathcal{A}(\cdot, i_2, \cdot)$ $\mathcal{A}(\cdot, \cdot, i_3)$

Unfolding

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



The 1-st mode matricization $\mathbf{X}^{(1)}$ is a $(n_1 \times n_2 n_3)$ matrix, obtained stacking the vectors

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

Products I

The **tensor product** of two vectors, $\mathbf{a}$ and $\mathbf{b}$, of length m and n results in a size $(m \times n)$ matrix $\mathbf{C} = \mathbf{a} \otimes \mathbf{b}$ s.t.

$$\mathbf{C}(i,j) = \mathbf{a}(i)\mathbf{b}(j).$$

The **Kronecker product**of two vectors, $\mathbf{a}$ and $\mathbf{b}$, of length m and n results in a length (mn) vector $\mathbf{c} = \mathbf{a} \otimes_{\mathrm{K}} \mathbf{b}$ such that

$$\mathbf{c}(h) = \mathbf{a}(i)\mathbf{b}(j)$$

where $h = (j-1)m + i$.

Remark

The vectorization of $\mathbf{a} \otimes \mathbf{b}$ is equal to $\mathbf{a} \otimes_{\mathrm{K}} \mathbf{b}$.

The **Kathri-Rao product** of two matrices $\mathbf{A}$ of size $(m_1 \times R)$ and $\mathbf{B}$ of size $(m_2 \times R)$ results in a size $(m_1 m_2 \times R)$ matrix $\mathbf{C} = \mathbf{A} \odot \mathbf{B}$ such that

$$\mathbf{C}(\cdot,j) = \mathbf{a}_j \otimes_{\mathrm{K}} \mathbf{b}_j.$$

Products II

The 1**st mode matrix-tensor product** of an $(n_1 \times n_2 \times n_3)$ tensor and a size $(n_1 \times m_1)$ matrix $\mathbf{G}$ results in an $(m_1 \times n_2 \times n_3)$ tensor $\mathcal{Y} = \mathcal{X} \times_1 \mathbf{G}$ s.t.

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

The **tensor contraction** along the first mode of two tensors, $\mathcal{A}$ and $\mathcal{B}$ of size $(n_1 \times n_2 \times n_3)$ and size $(n_1 \times m_2 \times m_3)$ results in a size $(n_2 \times n_3 \times m_2 \times m_3)$ tensor $\mathcal{C} = \mathcal{A} \cdot_1 \mathcal{B}$ s.t.

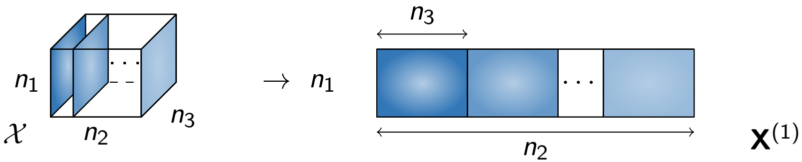
$$\mathcal{C}(i_2, i_3, j_2, j_3) = \sum_{i_1=1}^{n_1} \mathcal{A}(i_1, i_2, i_3)\mathcal{B}(i_1, j_2, j_3).$$

Remark

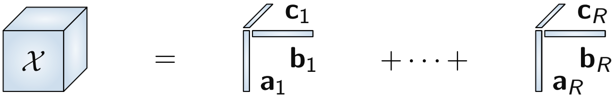
When more modes are contracted, the symbol is omitted!

Ranks

The **multilinear rank** of an $(n_1 \times n_2 \times n_3)$ tensor $\mathcal{X}$ is (r_1, r_2, r_3) where $r_h = \text{rank}(\mathbf{X}^{(h)})$.



The $(n_1 \times n_2 \times n_3)$ tensor $\mathcal{X}$ is a **rank-1 tensor** if it can be expressed as $\mathcal{X} = \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}$.



The **rank** of an $(n_1 \times n_2 \times n_3)$ tensor $\mathcal{X}$ is R the minimal number of rank-1 tensors that yield $\mathcal{X}$ in a linear combination

Table of Contents

Master's supervisors

Figure: Prof. A. Bernardi, University of Trento

- algebraic geometry
- algorithms for tensor decomposition

Figure: Prof. D. Rocchini, University of Bologna

- plant ecology
- algorithms to estimate biodiveristy

Background

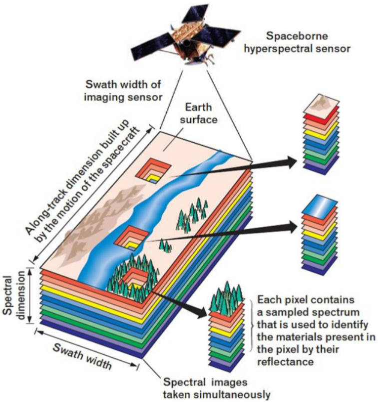


Figure: from [Bedini 2017].

Over a time series of Europe spectral images,

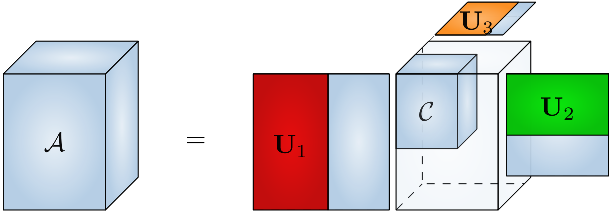
- get two images from two spectral bands ($\mathtt{RED}$ and $\mathtt{NIR}$);
- compute the normalized difference vegetation index per pixel, i.e.,

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

- compute a biodiversity index over the resulting $\mathtt{NDVI}$ image

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

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



If $\mathcal{A}$ is an $(n_1 \times n_2 \times n_3)$ tensor of multilinear rank (r_1, r_2, r_3), its Tucker decomposition is

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

where

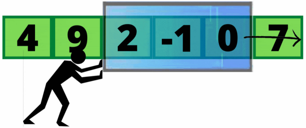
- the **core** tensor is $\mathcal{C}$ of size $(r_1 \times r_2 \times r_3)$;
- the i-th factor matrix is $\mathbf{U}_i$ an $(n_i \times r_i)$ orthogonal matrix.

Remark

The memory requirement is $\mathcal{O}(r^3 + nr)$ where $r = \max r_i$, $n = \max n_i$.

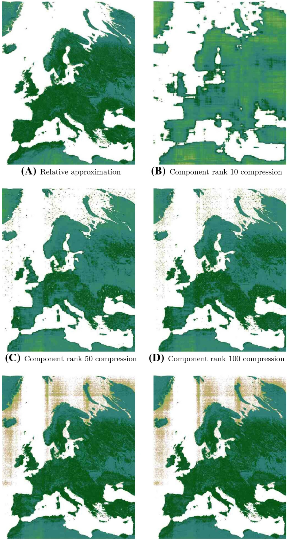
Experimental set-up

- organize the NIR and RED images into a tensor
- compress the images by approximating the tensor at different multilinear ranks
- construct the NDVI image from the compressed images
- estimate the biodiversity from the obtained image by moving window
- perform statistical analysis on the results



Moving window

Rényi index result [Bernardi et al. 2021]



(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

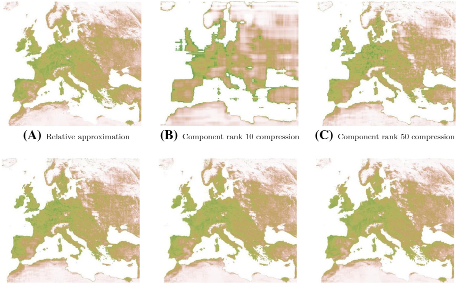
Rényi index

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. 2021]



(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

Rao index

Uses both 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%.

Table of Contents

Ph.D. supervisors

Figure: Prof. O. Coulaud, Inria Bordeaux

Figure: Prof. L. Giraud, Inria Bordeaux

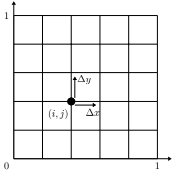
- tensor methods
- high-dimensional simulations

- numerical linear algebra
- finite precision arithmetic

Context

The problem

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



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

where $\boldsymbol{\mathcal{A}} : \mathbb{R}^{n_1 \times \cdots \times n_d} \to \mathbb{R}^{n_1 \times \cdots \times n_d}$ 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 linear algebra techniques

Tensor Train or Matrix Product States [Oseledets 2011]

Let $\mathcal{X}$ a tensor of order d and size $(n_1 \times \cdots \times n_d)$

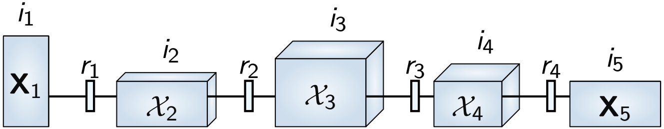


Figure: train of matrix - third-order tensors - matrix

then its TT-representation is $\mathcal{X} = \mathcal{X}_1 \mathcal{X}_2 \cdots \mathcal{X}_{d-1} \mathcal{X}_d$ s.t.

- the k-th **TT-core** is $\mathcal{X}_k$ an $(r_{k-1} \times n_k \times r_k)$ tensor
- the **TT-rank** is $(1, r_1, \ldots, r_{d-1}, 1)$
- $\mathcal{X}_1$ and $\mathcal{X}_d$ are two matrices

Remark

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

New variable accuracy approach

Which properties are maintained when objects are compressed by TT-format?

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

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

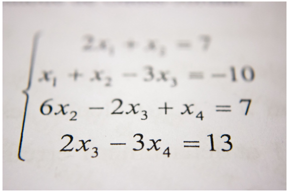
$$\delta\text{-storage}(\mathcal{Z}) = \overline{\mathcal{Z}} \qquad \text{s.t.} \qquad \frac{||\mathcal{Z} - \overline{\mathcal{Z}}||}{||\mathcal{Z}||} \leq \delta$$

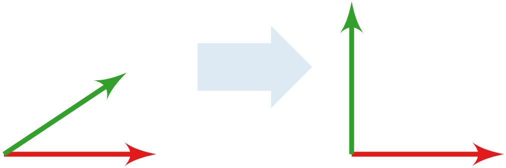
with fl is the classical floating point computational function dependent on u.

Numerical linear algebra methods

Iterative solver

- Generalized Minimal RESidual (GMRES)





Orthogonalization kernels

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

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*	$\|\|\mathbb{I}_k - \mathbf{Q}_k^\top \mathbf{Q}_k\|\|$
[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**	$\mathcal{O}(u)$

Classical and Modified Gram-Schmidt

Algorithm 1: $\boldsymbol{\mathcal{Q}}, \mathbf{R} = \texttt{TT-CGS}(\boldsymbol{\mathcal{A}}, \delta)$

Input: $\boldsymbol{\mathcal{A}} = [\mathcal{A}_1, \ldots, \mathcal{A}_m]$, $\delta \in \mathbb{R}_+$

for $i = 1, \ldots, m$ **do**
- $\mathcal{P} = \mathcal{A}_i$
- **for** $j = 1, \ldots, i-1$ **do**
  - $\mathbf{R}(i,j) = \langle \mathcal{A}_i, \mathcal{Q}_j \rangle$
  - $\mathcal{P} = \mathcal{P} - \mathbf{R}(i,j)\mathcal{Q}_j$
- $\mathcal{P} = \texttt{TT-rounding}(\mathcal{P}, \delta)$
- $\mathbf{R}(i,i) = ||\mathcal{P}||$
- $\mathcal{Q}_i = \mathcal{P}/\mathbf{R}(i,i)$

Output: $\boldsymbol{\mathcal{Q}} = \{\mathcal{Q}_1, \ldots, \mathcal{Q}_m\}$, $\mathbf{R}$

Algorithm 2: $\boldsymbol{\mathcal{Q}}, \mathbf{R} = \texttt{TT-MGS}(\boldsymbol{\mathcal{A}}, \delta)$

Input: $\boldsymbol{\mathcal{A}} = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$, $\delta \in \mathbb{R}_+$

for $i = 1, \ldots, m$ **do**
- $\mathcal{P} = \mathcal{A}_i$
- **for** $j = 1, \ldots, i-1$ **do**
  - $\mathbf{R}(i,j) = \langle \mathcal{P}, \mathcal{Q}_j \rangle$
  - $\mathcal{P} = \mathcal{P} - \mathbf{R}(i,j)\mathcal{Q}_j$
- $\mathcal{P} = \texttt{TT-rounding}(\mathcal{P}, \delta)$
- $\mathbf{R}(i,i) = ||\mathcal{P}||$
- $\mathcal{Q}_i = \mathcal{P}/\mathbf{R}(i,i)$

Output: $\boldsymbol{\mathcal{Q}} = \{\mathcal{Q}_1, \ldots, \mathcal{Q}_m\}$, $\mathbf{R}$

They readily write in TT-format.

Gram approach

Let $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_m]$, then we look for $\mathbf{A} = \mathbf{QR}$ with $\mathbf{Q}^\intercal \mathbf{Q} = \mathbb{I}_m$
compute the Gram matrix

$$\mathbf{A}^\intercal \mathbf{A} = (\mathbf{R}^\intercal \mathbf{Q}^\intercal)\mathbf{QR} = \mathbf{R}^\intercal \mathbf{R}$$

this is (almost) the **Cholesky** factorization of $\mathbf{A}^\intercal \mathbf{A}$ that can be written as

$$\mathbf{A}^\intercal \mathbf{A} = \mathbf{R}^\intercal \mathbf{R} = \mathbf{L}\mathbf{L}^\top$$

with the Cholesky factor $\mathbf{L} = \mathbf{R}^\intercal$ and then $\mathbf{Q}$ gets

$$\mathbf{Q} = \mathbf{A}\mathbf{R}^{-1} = \mathbf{A}(\mathbf{L}^\intercal)^{-1}$$

Gram approach

Algorithm 3: $\boldsymbol{\mathcal{Q}}, \mathbf{R} = \texttt{TT-Gram}(\boldsymbol{\mathcal{A}}, \delta)$

Input: $\boldsymbol{\mathcal{A}} = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$, $\delta \in \mathbb{R}_+$

$\mathbf{G}$ be the Gram matrix from $\boldsymbol{\mathcal{A}}$

$\mathbf{L} = \texttt{cholesky}(\mathbf{G})$

$\boldsymbol{\mathcal{Q}} = \{\mathcal{Q}_1, \ldots, \mathcal{Q}_d\}$ from $\boldsymbol{\mathcal{A}}$ and $(\mathbf{L}^\top)^{-1}$

for $i = 1, \ldots, m$ **do**

  $\mathcal{Q}_i = \delta\texttt{-storage}(\mathcal{Q}_i)$

Output: $\boldsymbol{\mathcal{Q}} = \{\mathcal{Q}_1, \ldots, \mathcal{Q}_m\}$, $\mathbf{R}$

In TT-format the following modifications occur

- $\mathbf{G}(i,j) = \langle \mathcal{A}_i, \mathcal{A}_j \rangle$
- $\mathbf{L}^\top$ inverse is explicitly computed
- $\mathcal{Q}_i$ is constructed as a linear combination of $\boldsymbol{\mathcal{A}}$ elements
- TT-rounding is used to compress at precision δ

Householder transformation

Given a vector $\mathbf{x} \in \mathbb{R}^n$ and a direction $\mathbf{y} \in \mathbb{R}^n$, the Householder reflector $\mathbf{H}$ reflects $\mathbf{x}$ along $\mathbf{y}$, i.e.,

$$\mathbf{Hx} = ||\mathbf{x}||\mathbf{y} \qquad \text{with} \qquad ||\mathbf{y}|| = 1.$$

Thanks to its properties, $\mathbf{H}$ writes as

$$\mathbf{H} = \mathbb{I}_n - \frac{2}{||\mathbf{u}||^2}\mathbf{u} \otimes \mathbf{u} \qquad \text{with} \qquad \mathbf{u} = (\mathbf{x} - ||\mathbf{x}||\mathbf{y}).$$

The practical implementation of the Householder transformation kernel uses the components of the input vectors.

Remark

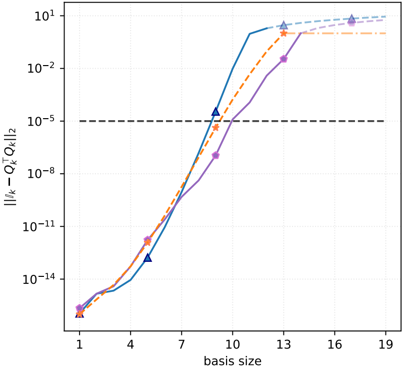
The Householder algorithm does **not** readily apply to tensor in TT-formats because of the compressed nature of this format.

TT-orthogonalization: numeric [Coulaud, Luc Giraud, 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: numeric [Coulaud, Luc Giraud, et al. 2022b]

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



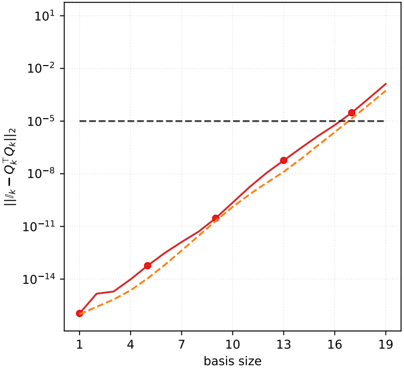
- **Gram approach**
- **CGS**
- $\kappa^2(\mathbf{A}_k)$

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

Loss of orthogonality for $\{\mathcal{A}_k\}_{k=1}^{20}$ TT-vectors of order 3, mode size 15, rounding accuracy $\delta = 10^{-5}$ compared with condition number of $\mathbf{A}_k = [\text{vec}(\mathcal{A}_1), \dots, \text{vec}(\mathcal{A}_k)]$

TT-orthogonalization: numeric [Coulaud, Luc Giraud, et al. 2022b]

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



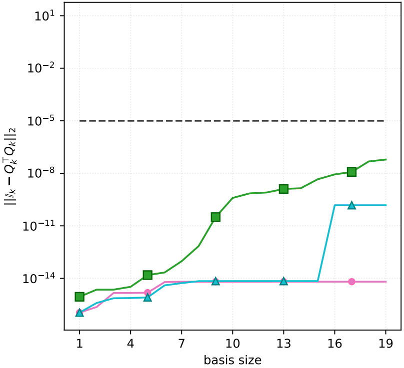
- **MGS**
- $\kappa(\mathbf{A}_k)$

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

Loss of orthogonality for $\{\mathcal{A}_k\}_{k=1}^{20}$ TT-vectors of order 3, mode size 15, rounding accuracy $\delta = 10^{-5}$ compared with condition number of $\mathbf{A}_k = [\text{vec}(\mathcal{A}_1), \dots, \text{vec}(\mathcal{A}_k)]$

TT-orthogonalization: numeric [Coulaud, Luc Giraud, et al. 2022b]

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



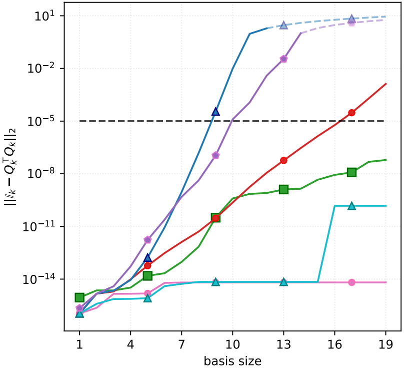
- **CGS2**
- **MGS2**
- **Householder transformation**

$\mathcal{O}(\delta)$

Loss of orthogonality for $\{\mathcal{A}_k\}_{k=1}^{20}$ TT-vectors of order 3, mode size 15, rounding accuracy $\delta = 10^{-5}$ compared with condition number of $\mathbf{A}_k = [\text{vec}(\mathcal{A}_1), \ldots, \text{vec}(\mathcal{A}_k)]$

TT-orthogonalization: numeric [Coulaud, Luc Giraud, 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** $\mathcal{O}(\delta\kappa^2(\mathbf{A}_k))$
- **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** $\mathcal{O}(\delta)$

Loss of orthogonality for $\{\mathcal{A}_k\}_{k=1}^{20}$ TT-vectors of order 3, mode size 15, rounding accuracy $\delta = 10^{-5}$ compared with condition number of $\mathbf{A}_k = [\text{vec}(\mathcal{A}_1), \ldots, \text{vec}(\mathcal{A}_k)]$

TT-orthogonalization: theory (work in progress)

Let $\mathbf{a}, \mathbf{b}, \mathbf{x}$ and $\mathbf{q} \in \mathbb{R}^n$, while $\delta \in (0,1)$ a precision, u a unit roundoff and $\gamma_n = \frac{nu}{1-nu}$.

Mixed inner product

In the mixed-precision system, the inner product between $\mathbf{a}$ and $\mathbf{q}$ is such that

$$|fl(\mathbf{a}^\top \tilde{\mathbf{q}}) - \mathbf{a}^\top \mathbf{q}| \leq ||\mathbf{a}||\, ||\mathbf{q}||\, (\gamma_n + \delta + \delta\gamma_n)$$

where $||\mathbf{q} - \tilde{\mathbf{q}}|| \leq \delta\, ||\mathbf{q}||$.

Mixed projection

Let $\tilde{\mathbf{P}} = \mathbb{I}_n - \tilde{\mathbf{a}}\tilde{\mathbf{b}}^\top$ be a normwise perturbed projector. If $\mathbf{y} = \tilde{\mathbf{P}}\mathbf{x} = \mathbf{x} - \tilde{\mathbf{a}}(\tilde{\mathbf{b}}^\top \mathbf{x})$ (exact arithmetic) and $\bar{\mathbf{y}} = fl(\tilde{\mathbf{P}}\mathbf{x})$ (finite arithmetic), then

$$||\mathbf{y} - \bar{\mathbf{y}}|| \leq (\gamma_m + \eta_\ell + \gamma_m\eta_\ell)(1 + ||\mathbf{a}||\, ||\mathbf{b}||\,)||\mathbf{x}||\,.$$

where $\eta_\ell = \ell\delta$ for $\ell \in \mathbb{N}$.

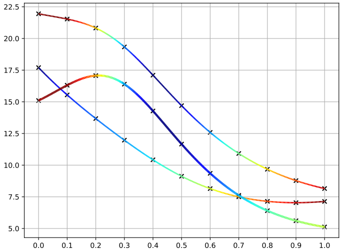
Table of Contents

Postdoctoral project

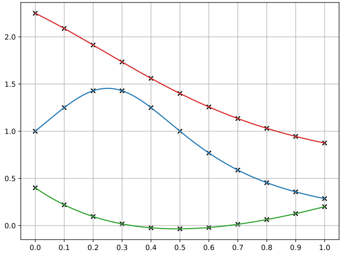
- Blind Source Separation (BSS)
- algebraic algorithm for Canonical Polyadic Decomposition
- improve the algorithm efficiency

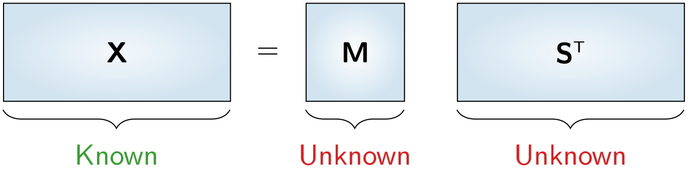
Figure: Prof. L. De Lathauwer, KU Leuven

Blind Source Separation problem



$$= \begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{bmatrix} \times$$

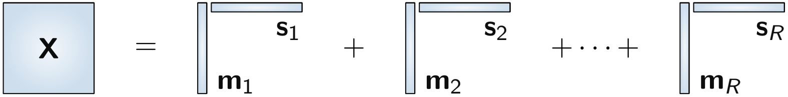




Factor Analysis and Blind Source Separation

- Decompose a data matrix in rank-1 terms that can be interpreted
  E.g. statistics, telecommunication, biomedical applications, chemometrics, data analysis, . . .

$$\mathbf{X} = \mathbf{M}\mathbf{S}^{\mathsf{T}}$$

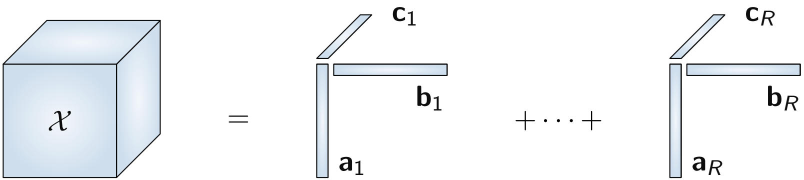


- $\mathbf{M}$: mixing matrix $\mathbf{S}$: source signals
- Matrix decomposition in rank-1 terms is not unique!

$$\mathbf{X} = (\mathbf{M}\mathbf{G})(\mathbf{G}^{-1}\mathbf{S}^{\mathsf{T}}) = \tilde{\mathbf{M}}\tilde{\mathbf{S}}^{\mathsf{T}}$$

What about tensor decomposition techniques?

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

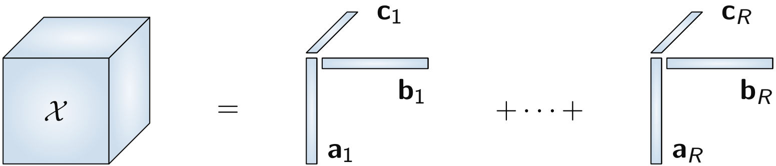
$$\mathcal{A} = \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r$$

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

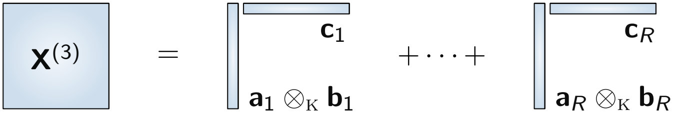
- unique under mild assumptions
- memory cost $\mathcal{O}(dnR)$
- NP-hard problem
- algorithms affected by numerical instabilities

CPD reformulation

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



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



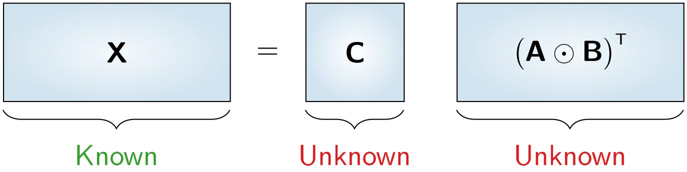
$$(\mathbf{a}_r \otimes_{\mathrm{K}} \mathbf{b}_r) \in \mathcal{V} = \left\{ \operatorname{vec}(\mathbf{Z}) : \begin{vmatrix} z_{i_1 j_1} & z_{i_1 j_2} \\ z_{i_2 j_1} & z_{i_2 j_2} \end{vmatrix} = 0 \right\} \text{ algebraic variety}$$

Algebraic algorithm: high view

Let $\mathcal{X}$ be a $(n_1 \times n_2 \times n_3)$ tensor, then

$$\mathbf{X}^{(3)} = \sum_{r=1}^{R} (\mathbf{a}_r \otimes_{\mathrm{K}} \mathbf{b}_r) \otimes \mathbf{c}_r^{\mathsf{T}} = \begin{bmatrix} \mathbf{a}_1 \otimes_{\mathrm{K}} \mathbf{b}_1 & \cdots & \mathbf{a}_R \otimes_{\mathrm{K}} \mathbf{b}_R \end{bmatrix} \mathbf{C}^{\mathsf{T}}.$$

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



1. compute $\mathbf{C}^{-1}$ from $\mathbf{X}$ using algebraic geometry properties;
2. decompose at rank 1 each column of $(\mathbf{C}^{-1}\mathbf{X})^{\mathsf{T}} = \begin{bmatrix} \mathbf{a}_1 \otimes_{\mathrm{K}} \mathbf{b}_1 & \cdots & \mathbf{a}_R \otimes_{\mathrm{K}} \mathbf{b}_R \end{bmatrix}$;
3. compute $\mathbf{C}$ by solving $(\mathbf{A} \odot \mathbf{B})\mathbf{C}^{\mathsf{T}} = \mathbf{X}$.

Retrieving C - I

The vector $\mathbf{e}$ is a column of $\mathbf{C}^{-1}$ if and only if $\mathbf{eX}$ is equal to $\mathbf{a}_\ell \otimes_{\mathrm{K}} \mathbf{b}_\ell$.
We try to characterize the vector $\mathbf{e}$, defining the matrix-valued function

$$\mathbf{W}(\mathbf{e})(i_1, i_2) = \sum_{i_3=1}^{n_3} \mathbf{e}(i_3)\mathcal{X}(i_1, i_2, i_3) = \mathcal{X} \times_3 \mathbf{e}$$

$$\Downarrow$$

$\mathbf{e}$ is a column of $\mathbf{C}^{-1}$ if and only if $\mathbf{W}(\mathbf{e})$ is equal to a $\mathbf{a}_\ell \otimes \mathbf{b}_\ell$

$$\Updownarrow$$

$$\text{rank}(\mathbf{W}(\mathbf{e})) = 1$$

$$\Updownarrow$$

The determinant of each (2×2) minor of $\mathbf{W}(\mathbf{e})$ is equal to 0

Retrieving C - II

The determinant of each (2×2) minor of $\mathbf{W}(\mathbf{e})$ is equal to 0

⇕

There exists a system of $C_I^2 C_J^2$ homogeneous degree 2 polynomial equations in $\mathbf{e}$

$$P_{h_1 h_2}(\mathbf{e}) = \sum_{\substack{k_1,k_2=1 \\ k_1 \leq k_2}}^{n_3} \mathcal{Q}(h_1, h_2, k_1, k_2)\, \mathbf{e}(k_1)\mathbf{e}(k_2)$$

⇕

searching elements in the kernel of $\mathbf{Q}^{(1,2)}$ which can be mapped to symmetric rank-1 matrices.

Retrieving C - III

The number of solution to the homogeneous degree 2 polynomial system can be predicted using the algebraic geometry theorem [Conca et al. 1994] as

$$\max\left\{R, \binom{n_1 - 1}{2}\binom{n_2 - 1}{2}\right\}$$

if $n_3 = r = (n_1 - 1)(n_2 - 1)$

- if R solutions are found, then
  - organize them into a $(n_3 \times n_3 \times R)$ tensor and compute its CPD to retrieve $\mathbf{C}^{-1}$;
  - perform a column-wise SVD of $(\mathbf{X}\mathbf{C}^{-1})$ to retreive $\mathbf{A}$ and $\mathbf{B}$
  - solve for $\mathbf{C}$ the equation $(\mathbf{A} \odot \mathbf{B})\mathbf{C}^{\top} = \mathbf{X}$
- if more than R solutions are found, we need to discard some solutions

⇩

Embedding procedure

Embedding procedure

If the kernel of $\mathbf{Q}^{(1,2)}$ has more than R elements which can be mapped to rank-1 symmetric matrices, then we can define new polynomial equations as

$$(\mathbf{e}(k_1))^{d_1} \cdots (\mathbf{e}(k_\ell))^{d_\ell} P_{h_1 h_2}(\mathbf{e}) = 0$$

where ℓ is the number of times we need to repeat the embedding.

⇓

Search for the elements in the kernel of $(\mathbb{I}_{R^\ell} \otimes_{\mathrm{K}} \mathbf{Q}^{(1,2)})$ which can be mapped to symmetric rank-1 tensors of order $\ell + 2$.

⇓

From [Conca et al. 1994], the value ℓ is known **a priori** as the minimum integer such that

$$\binom{n_1 - 1}{2 + \ell} \binom{n_2 - 1}{2 + \ell} \leq R \qquad \text{if} \qquad n_3 = R = (n_1 - 1)(n_2 - 1)$$

Numerical drawbacks

- the construction of $\mathbf{Q}^{(1,2)}$ has complexity $C^2_{n_1} C^2_{n_2} C^2_{R+1}$, how can we improve it? Could we benefit from the matrix $\mathbf{Q}^{(1,2)}$ structure?
- the matrix $\mathbb{I}_{R^\ell} \otimes \mathbf{Q}^{(1,2)}$ has dimension $(R^\ell C^2_{n_1} C^2_{n_2} \times C^2_{R+1} R^\ell)$, thus its SVD becomes quickly expensive
- the auxiliary tensor formed by the R symmetric order $\ell + 2$ tensors has to be decomposed by CPD. How to compute this CPD reliably and effectively?
- the results of [Conca et al. 1994] holds under the assumption of generically independence of the system polynomials, how can we guarantee it from the input tensor $\mathcal{X}$?
- could the number of symmetric elements in the kernel of $\mathbb{I}_{R^\ell} \otimes \mathbf{Q}^{(1,2)}$ be predicted a priori if $R \neq (n_1 - 1)(n_2 - 1)$?
- could we assess the robustness to numerical errors of this algorithm?

Table of Contents

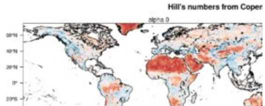
Wrap up

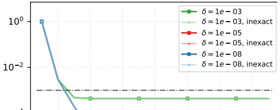
Tensor methods used in

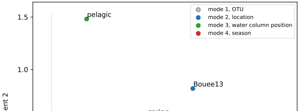
- data analysis problem as compression methods
  - e.g., the Tucker's decomposition
- scientific computing as new policy for computational methods
  - e.g., the Tensor-Train decomposition
- signal processing
  - e.g., the Canonical Polyadic Decomposition

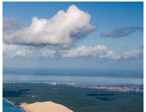
Other projects

- `rasterdiv` an R package to compute biodiversity indexes [Rocchini et al. 2021]
- Generalized Minimal RESidual in variable accuracy [Agullo et al. 2022]
- Inexact TT-GMRES features for parametric operators [Coulaud, Luc Giraud, et al. 2022a]
- High Order Correspondance Analysis applied to ecological datasets [Coulaud, Franc, et al. 2021]
- Bind Inferference Suppression algorithm









Thank you for the attention!
Questions? Advice?

References I

- Agullo, Emmanuel et al. (Sept. 2022). *The backward stable variants of GMRES in variable accuracy*. Research Report RR-9483. Inria, pp. 1–77.
- 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 (Dec. 2021). "High order singular value decomposition for plant diversity estimation". In: *Bollettino dell'Unione Matematica Italiana* 14.4, 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.

References II

- Conca, Aldo and Jürgen Herzog (1994). "On the Hilbert Function of Determinantal Rings and Their Canonical Module". In: *Proceedings of the American Mathematical Society* 122.3, pp. 677–681.
- Coulaud, Olivier, Alain Franc, and Martina Iannacito (Nov. 2021). *Extension of Correspondence Analysis to multiway data-sets through High Order SVD: a geometric framework*. Research Report RR-9429. Inria center at the University of Bordeaux; INRAE.
- Coulaud, Olivier, Luc Giraud, and Martina Iannacito (2022a). *A note on GMRES in TT-format*. Research Report RR-9484. Inria Bordeaux Sud-Ouest.
- — (Nov. 2022b). *On some orthogonalization schemes in Tensor Train format*. Research Report RR-9491. Inria center at the University of Bordeaux.
- 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.

References III

- 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.
- 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.

- Rocchini, Duccio et al. (Feb. 2021). “rasterdiv—An Information Theory tailored R package for measuring ecosystem heterogeneity from space: To the origin and back”. In: *Methods in Ecology and Evolution* 12.6, pp. 1093–1102.
- Saad, Youcef and Martin H. Schultz (1986). “GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems”. In: *SIAM Journal on Scientific and Statistical Computing* 7.3, pp. 856–869.
- 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.

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.

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

Easter egg

Happy birthday Daniele!

spontaneously from J. C. F. Gauss.AI

Background on GMRES I [Saad et al. 1986]

To solve $\mathbf{Ax} = \mathbf{b}$ with initial guess $\mathbf{x}_0 = 0$, at the k-th iteration GMRES minimizes the norm of residual

$$||\mathbf{r}_k|| = \min_{\mathbf{x} \in \mathcal{K}_k(\mathbf{A},\mathbf{b})} ||\mathbf{Ax} - \mathbf{b}||$$

in the Krylov space $\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \text{span}\{\mathbf{b}, \mathbf{Ab}, \ldots, \mathbf{A}^{k-1}\mathbf{b}\}$

Background on GMRES I [Saad et al. 1986]

To solve $\mathbf{Ax} = \mathbf{b}$ with initial guess $\mathbf{x}_0 = 0$, at the k-th iteration GMRES minimizes the norm of residual

$$||\mathbf{r}_k|| = \min_{\mathbf{x} \in \mathcal{K}_k(\mathbf{A},\mathbf{b})} ||\mathbf{Ax} - \mathbf{b}||$$

in the Krylov space $\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \text{span}\{\mathbf{b}, \mathbf{Ab}, \dots, \mathbf{A}^{k-1}\mathbf{b}\}$
Practically, let $\mathbf{V}_k = [\mathbf{v}_1, \dots, \mathbf{v}_k]$ such that

$$\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$$

thanks to the **Arnoldi relation**

$$\mathbf{AV}_k = \mathbf{V}_{k+1}\overline{\mathbf{H}}_k \qquad \text{with} \qquad \mathbf{V}_{k+1}^T \mathbf{V}_{k+1} = \mathbb{I}_{k+1}$$

Background on GMRES I [Saad et al. 1986]

To solve $\mathbf{Ax} = \mathbf{b}$ with initial guess $\mathbf{x}_0 = 0$, at the k-th iteration GMRES minimizes the norm of residual

$$||\mathbf{r}_k|| = \min_{\mathbf{x} \in \mathcal{K}_k(\mathbf{A},\mathbf{b})} ||\mathbf{Ax} - \mathbf{b}||$$

in the Krylov space $\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \text{span}\{\mathbf{b}, \mathbf{Ab}, \dots, \mathbf{A}^{k-1}\mathbf{b}\}$
Practically, let $\mathbf{V}_k = [\mathbf{v}_1, \dots, \mathbf{v}_k]$ such that

$$\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$$

thanks to the **Arnoldi relation**

$$\mathbf{AV}_k = \mathbf{V}_{k+1}\overline{\mathbf{H}}_k \qquad \text{with} \qquad \mathbf{V}_{k+1}^T \mathbf{V}_{k+1} = \mathbb{I}_{k+1}$$

Commonly Householder or Modified Gram-Schmidt algorithms are used to construct $\mathbf{V}_k$

Background on GMRES II [Saad et al. 1986]

Thanks to the Arnoldi relation, in exact arithmetic the residual can be written as

$$||\mathbf{r}_k|| = \min_{\mathbf{x}\in\mathcal{K}_k(\mathbf{A},\mathbf{b})} ||\mathbf{A}\mathbf{x} - \mathbf{b}|| = \min_{\mathbf{y}} \left|\left|\beta\mathbf{e}_1 - \overline{\mathbf{H}}_k\mathbf{y}\right|\right| = ||\tilde{\mathbf{r}}_k||.$$

Remark

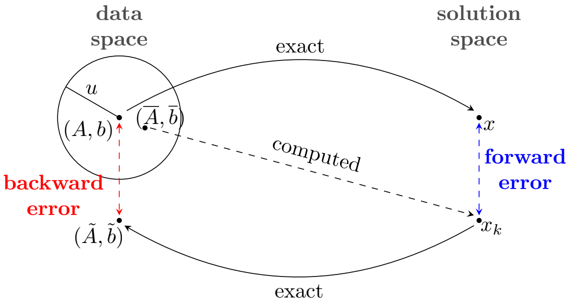
In finite arithmetic, the exact residual $\mathbf{r}_k$ and the LS-residual $\tilde{\mathbf{r}}_k$ differ!

If the smaller minimization problem is solved by $\mathbf{y}_k$, the updated iterative solution is

$$\mathbf{x}_k = \mathbf{V}_k\mathbf{y}_k$$

GMRES property [Wilkinson 1963]

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



GMRES is backward stable, i.e.,

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

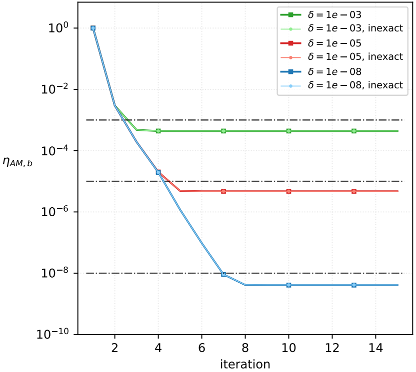
TT-GMRES results [Dolgov 2013; Coulaud, Luc Giraud, et al. 2022a]

Convection-Diffusion problem

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

TT-GMRES modifications

- Arnoldi basis compressed at accuracy δ
- Iterative solution compressed at accuracy δ



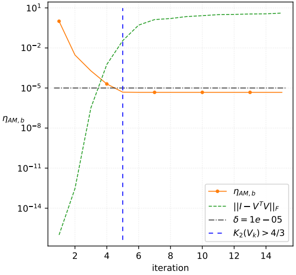
TT-GMRES: inexact variant

TT-GMRES

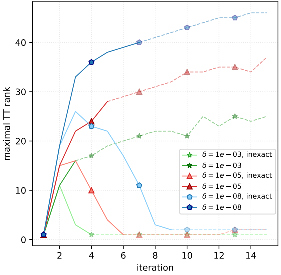
- constant rounding accuracy δ
- iterative solution rounded at precision δ

inexact TT-GMRES

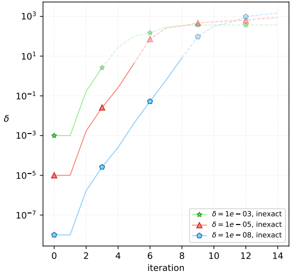
- increasing rounding accuracy $\delta / ||\tilde{\mathbf{r}}||$
- iterative solution at IEEE precision



Convergence history vs LOO



Maximal TT-rank comparison



δ values