Low-rank approximation tools for isogeometric analysis

Angelos Mantzaflaris

RICAM, Austrian Academy of Sciences

CIME-EMS Summer School on Splines and PDEs

Cetraro, July 4, 2017

Acknowledgements: B. Jüttler, B.N. Khoromskij, U. Langer and F. Scholz

Overview

1. Low-rank function approximation
   - Low-rank spline functions
   - Singular Value Decomposition
   - Adaptive Cross Approximation
   - Decoupling more than two variables
2. Complete low rank approximation
   - Tensor formats and tensor rank
   - Alternating least squares method
   - Higher-order SVD
3. Isogeometric analysis on tensor-product geometry mappings
   - Model problem and variational formulation
   - Separation rank and Kronecker rank
   - Computational complexity
4. Benchmarks

Overview

1. Low-rank function approximation
   - Low-rank spline functions
   - Singular Value Decomposition
   - Adaptive Cross Approximation
   - Decoupling more than two variables
2. Complete low rank approximation
   - Tensor formats and tensor rank
   - Alternating least squares method
   - Higher-order SVD
3. Isogeometric analysis on tensor-product geometry mappings
   - Model problem and variational formulation
   - Separation rank and Kronecker rank
   - Computational complexity
4. Benchmarks

Low-rank spline functions

- Tensor-product spline function:

$$f(\boldsymbol{x}) = \mathbf{A} : \mathbf{B}(\boldsymbol{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}\, \beta_i(x_1)\, \beta_i(x_2)$$

  - ❒ Coefficient matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{A} = [a_{ij}]$
  - ❒ Basis matrix $\mathbf{B}(\boldsymbol{x}) \in \mathbb{R}^{n \times n}$
    $\mathbf{B}(\boldsymbol{x}) = \beta(x_1) \otimes \beta(x_2) = \left[\beta_i(x_1)\beta_j(x_2)\right]$

- Rank$-R$ spline function:

  - ❒ The coefficient matrix $\mathbf{A} = \sum_{r=1}^{R} \boldsymbol{u}_r \otimes \boldsymbol{v}_r$, $R < n$

$$f(\boldsymbol{x}) = \mathbf{A} : \mathbf{B}(\boldsymbol{x}) = \sum_{r=1}^{R} (\boldsymbol{u}_r \otimes \boldsymbol{v}_r) : (\beta(x_1) \otimes \beta(x_2)) = \sum_{r=1}^{R} f_r^{(1)}(x_1)\, f_r^{(2)}(x_2)$$

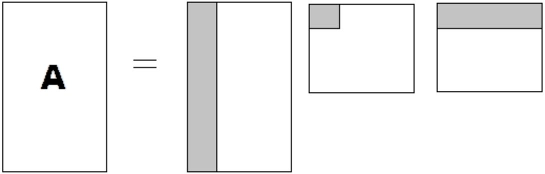
  - ❒ Evaluation at some $\boldsymbol{x}$: $\mathcal{O}(Rdp^2)$ vs $\mathcal{O}(p^{d+1})$

Low rank approximation by SVD

Singular value decomposition for $\mathbf{A} \in \mathbb{R}^{n \times m}$:

$$\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T$$

- $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$, $\sigma_k \geq 0$ and $\mathbf{U}, \mathbf{V}$ orthonormal matrices.
- Optimal approximation of $\mathbf{A}$ by $\mathbf{A}' \in \mathbb{R}^{n \times m}$ of rank R,
  Truncate $\Sigma' = \mathrm{diag}(\sigma_1, \ldots, \sigma_R, 0, \ldots, 0)$

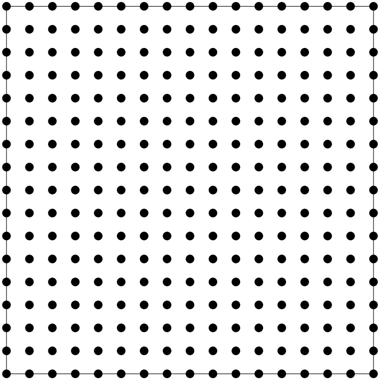


  Complexity $\mathcal{O}(\min(nm^2, mn^2))$, randomized $\mathcal{O}(R^2 \min(n, m))$
- Error estimate [Eckart-Young theorem] for $f(\boldsymbol{x}) = \mathbf{A} : \mathbf{B}(\boldsymbol{x})$:

$$\|f(\boldsymbol{x}) - \mathbf{A}' : \mathbf{B}(\boldsymbol{x})\|_{L^\infty}^2 \leq \|A - A'\|_F^2 = \sum_{r=R+1}^{n} \sigma_r^2$$

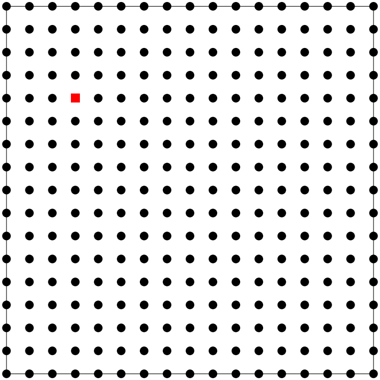
Low rank approximation by ACA

Adaptive Cross Approximation: $\mathbf{E}^{[0]} = \mathbf{A},\ \mathbf{E}^{[r]} = \mathbf{E}^{[r-1]} - \dfrac{\boldsymbol{u}_i^{[r-1]} \otimes \boldsymbol{v}_i^{[r-1]}}{e_{ij}^{[r-1]}}$



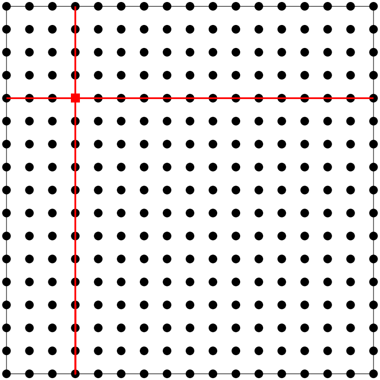
Low rank approximation by ACA

Adaptive Cross Approximation: $\mathbf{E}^{[0]} = \mathbf{A}, \ \mathbf{E}^{[r]} = \mathbf{E}^{[r-1]} - \dfrac{\boldsymbol{u}_i^{[r-1]} \otimes \boldsymbol{v}_i^{[r-1]}}{e_{ij}^{[r-1]}}$



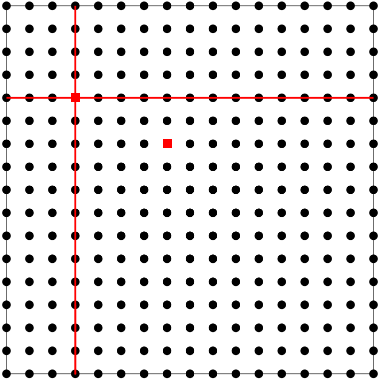
Low rank approximation by ACA

Adaptive Cross Approximation: $\mathbf{E}^{[0]} = \mathbf{A}, \ \mathbf{E}^{[r]} = \mathbf{E}^{[r-1]} - \dfrac{\boldsymbol{u}_i^{[r-1]} \otimes \boldsymbol{v}_i^{[r-1]}}{e_{ij}^{[r-1]}}$



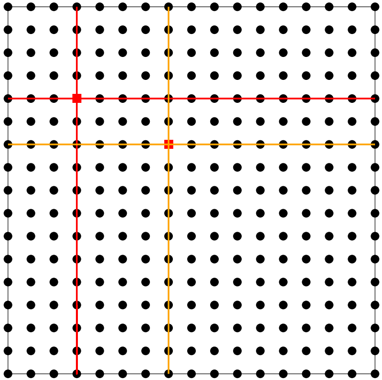
Low rank approximation by ACA

Adaptive Cross Approximation: $\mathbf{E}^{[0]} = \mathbf{A}, \ \mathbf{E}^{[r]} = \mathbf{E}^{[r-1]} - \dfrac{\boldsymbol{u}_i^{[r-1]} \otimes \boldsymbol{v}_i^{[r-1]}}{e_{ij}^{[r-1]}}$



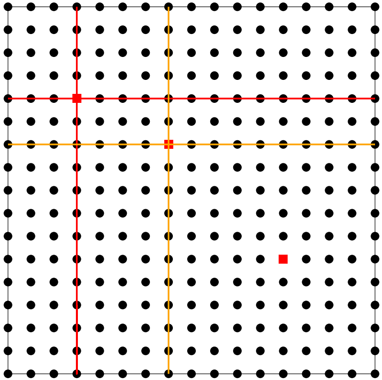
Low rank approximation by ACA

Adaptive Cross Approximation: $\mathbf{E}^{[0]} = \mathbf{A}, \; \mathbf{E}^{[r]} = \mathbf{E}^{[r-1]} - \dfrac{\boldsymbol{u}_i^{[r-1]} \otimes \boldsymbol{v}_i^{[r-1]}}{e_{ij}^{[r-1]}}$



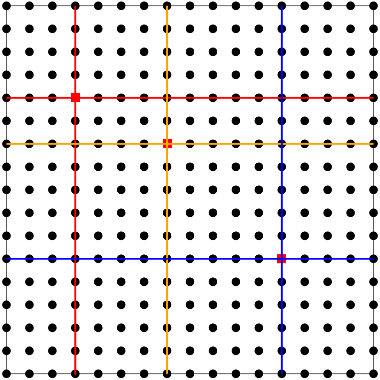
Low rank approximation by ACA

Adaptive Cross Approximation: $\mathbf{E}^{[0]} = \mathbf{A}, \ \mathbf{E}^{[r]} = \mathbf{E}^{[r-1]} - \dfrac{\boldsymbol{u}_i^{[r-1]} \otimes \boldsymbol{v}_i^{[r-1]}}{e_{ij}^{[r-1]}}$



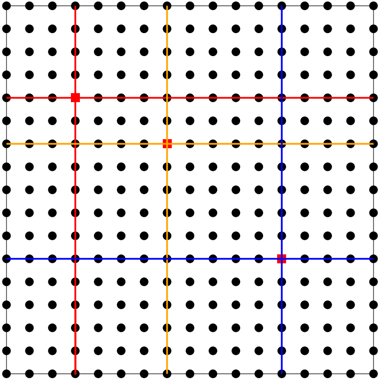
Low rank approximation by ACA

Adaptive Cross Approximation: $\mathbf{E}^{[0]} = \mathbf{A},\ \mathbf{E}^{[r]} = \mathbf{E}^{[r-1]} - \dfrac{\boldsymbol{u}_i^{[r-1]} \otimes \boldsymbol{v}_i^{[r-1]}}{e_{ij}^{[r-1]}}$



Low rank approximation by ACA

Adaptive Cross Approximation: $\mathbf{E}^{[0]} = \mathbf{A}, \; \mathbf{E}^{[r]} = \mathbf{E}^{[r-1]} - \dfrac{\boldsymbol{u}_i^{[r-1]} \otimes \boldsymbol{v}_i^{[r-1]}}{e_{ij}^{[r-1]}}$



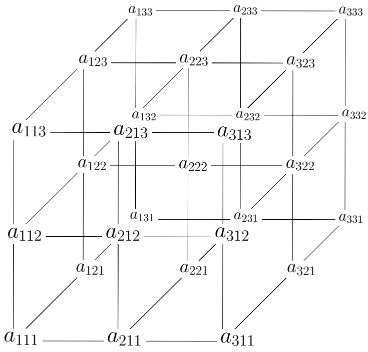
- $\mathbf{A} \approx U C^{-1} V$
- Choose pivots iteratively until $\|E^{[r]}\| \approx 0$
  - Full pivoting $\mathcal{O}(nm)$
  - Partial pivoting $\mathcal{O}(R \min(n, m))$
- Maximum-volume choice is optimal:

$$C^* = \underset{C \text{ submat. } \mathbf{A}}{\operatorname{argmax}} \left(|\det C|\right)$$

Partial decoupling for more then two variables

- Volumetric spline
  $f(\boldsymbol{x}) = \mathbf{A} : \mathbf{B}(\boldsymbol{x})$
- Tensor of order 3
  $\mathbf{A} = [a_{ijk}]$

  $\mathbf{B} = \beta(x_1)\otimes\beta(x_2)\otimes\beta(x_3)$



Unfoldings:

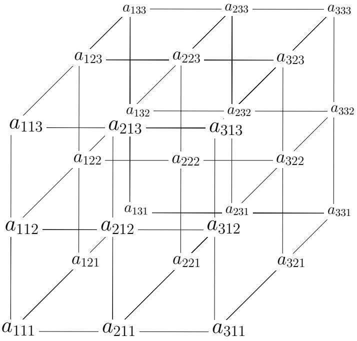
$$A^{(1)} = \begin{bmatrix} a_{111} & \cdots & a_{133} \\ a_{211} & \cdots & a_{233} \\ a_{311} & \cdots & a_{333} \end{bmatrix}, A^{(2)} = \begin{bmatrix} a_{111} & \cdots & a_{313} \\ a_{121} & \cdots & a_{323} \\ a_{131} & \cdots & a_{333} \end{bmatrix}, A^{(3)} = \begin{bmatrix} a_{111} & \cdots & a_{331} \\ a_{112} & \cdots & a_{332} \\ a_{113} & \cdots & a_{333} \end{bmatrix}$$

$$\sum_{r=1}^{R_1} f_r^{(1)}(x_1) f_r^{(23)}(x_2,x_3), \quad \sum_{r=1}^{R_2} f_r^{(2)}(x_2) f_r^{(13)}(x_1,x_3), \quad \sum_{r=1}^{R_3} f_r^{(3)}(x_3) f_r^{(12)}(x_1,x_2)$$

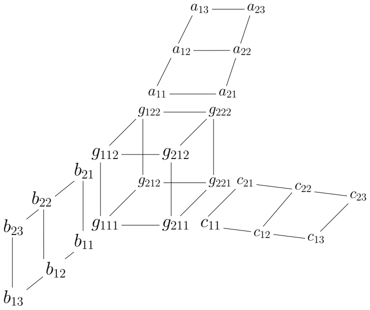
Overview

Low rank approximation : 3D, 4D, ...

Full tensor:



Tucker: generalizes matrix SVD (core+side matrices)

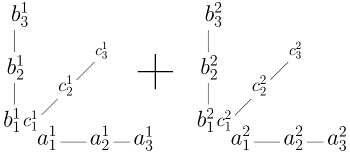


Algorithms for
low rank tensor approximation:

- Higher-order SVD
- Alternating least squares (ALS):

$$\min \|\mathbf{A} - \mathbf{A}'\|_F$$

Canonical: sum of rank-1 tensors (skeleton vectors)



Formal definition

Given $\boldsymbol{n} = (n_1, \ldots, n_d) \in \mathbb{Z}_+^n$, a(real) $\boldsymbol{n}$-tensor $\mathbf{A} \in \mathbb{W}_{\boldsymbol{n}}$ generalizes matrix structure to more than two indices, i.e. its elements are $t_{i_1,\ldots,i_d}$.

- The *canonical format of rank R*

$$\mathbf{A} = \sum_{r=1}^{R} \bigotimes_{k=1}^{d} \mathbf{v}^{(k,r)}, \quad i.e., \quad a_{\boldsymbol{i}} = a_{i_1,\ldots,i_d} = \sum_{r=1}^{R} \prod_{k=1}^{d} v_{i_k}^{(k,r)},$$

- *The* Tucker format of rank $(\rho_1, \ldots, \rho_d)$

$$\mathbf{A} = \sum_{r_1=1}^{\rho_1} \cdots \sum_{r_d=1}^{\rho_d} c_{r_1,\ldots,r_d} \bigotimes_{k=1}^{d} \mathbf{v}^{(k,r_k)}, \text{ i.e., } a_{\boldsymbol{i}} = \sum_{r_1=1}^{\rho_1} \cdots \sum_{r_d=1}^{\rho_d} c_{r_1,\ldots,r_d} \prod_{k=1}^{d} v_{i_k}^{(k,r_k)}.$$

- Tensors of fixed rank form a smooth manifold $\mathbb{S} \subset \mathbb{W}_{\boldsymbol{n}}$. Truncation (or projection) operator:

$$T_{\mathbb{S}} : \mathbb{W}_{\boldsymbol{n}} \to \mathbb{S} : \qquad T_{\mathbb{S}}\mathbf{A} = \operatorname*{argmin}_{\mathbf{U} \in \mathbb{S}} \|\mathbf{A} - \mathbf{U}\|,$$

Low-rank approximation in Canonical format

- Alternating least squares (ALS) algorithm:

$$\min \mathrm{E}\left(V^{(1)}, \dots, V^{(d)}\right) := \min \left\| \mathbf{A} - \sum_{r=1}^{R} \bigotimes_{k=1}^{d} \mathbf{v}^{(k,r)} \right\|_F^2 .$$

Iterate (until stopping criterion);

$$V^{(1)'} = \operatorname*{argmin}_{U \in \mathbb{R}^{n_1 \times R}} \mathrm{E}\left(U, V^{(2)}, \dots, V^{(d)}\right)$$

$$\vdots$$

$$V^{(d)'} = \operatorname*{argmin}_{U \in \mathbb{R}^{n_d \times R}} \mathrm{E}\left(V^{(1)'}, \dots, V^{(d-1)'}, U\right) .$$

Each iteration costs $\mathcal{O}(Rdn^d)$

Low-rank approximation in Tucker format

- HOSVD algorithm (higher-order SVD)

1. Compute the SVD of all unfoldings, i.e.,

$$A^{(k)} = \tilde{U}^{(k)} \Sigma^{(k)} \tilde{V}^{(k)^T} = U^{(k)} V^{(k)^T} \quad k = 1, \ldots, d$$

   and truncate so that $U^{(k)} \in \mathbb{R}^{n_k \times \rho_k}$.

2. Compute the core tensor

$$\mathbf{C} = U^{(1)} \cdots U^{(d)} \cdot \mathbf{A}$$

3. Return $U^{(1)}, \ldots, U^{(d)}$ and $\mathbf{C}$, so that $\tilde{\mathbf{A}} = U^{(1)} \cdots U^{(d)} \cdot \mathbf{C}$

- Analogously to the 2D case:

$$\|\mathbf{A} - \tilde{\mathbf{A}}\|_F^2 \leq \sum_{r_1=\rho_1+1}^{n} \sigma_{1,r_1}^2 + \cdots + \sum_{r_d=\rho_d+1}^{n} \sigma_{d,r_d}^2$$

  where $\Sigma^{(k)} = \mathrm{diag}(\sigma_{k,r_k})$.

- Complexity $\mathcal{O}(dn^{d+1})$, randomized $\mathcal{O}(dR^2 n)$

Overview

Model problem and variational formulation

- Differential operator

$$Lu = -\nabla \cdot (A(x)\nabla u) + c(x)u.$$

- Boundary value problem $G : \hat{\Omega} \to \Omega$,

$$\begin{cases} Lu &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{cases} \tag{1}$$

- Pull-back to $\hat{\Omega} = [0,1]^d$:

$$\begin{cases} \hat{L}\hat{u} &= \hat{f} \quad \text{in } \hat{\Omega}, \\ \hat{u} &= 0 \quad \text{on } \partial\hat{\Omega}. \end{cases}$$

$$\hat{L}\hat{u} = |\det J_G| L\left(\hat{u} \circ G^{-1}\right) = -\hat{\nabla} \cdot (K\hat{\nabla}\hat{u}) + \omega\hat{u},$$

$$\text{where} \begin{cases} K &= |\det(J_G)|\, J_G^{-1} A J_G^{-T}, \\ \omega &= |\det(J_G)|\, c, \\ \hat{f} &= |\det J_G|\, f \circ G \end{cases}$$

Separation rank and Kronecker rank

Consider the bilinear form associated to $\hat{L}$:

$$a(\hat{u}, \hat{v}) = \int_{\hat{\Omega}} \nabla \hat{u}^{\intercal} \cdot \boldsymbol{K} \cdot \nabla \hat{v} + \int_{\hat{\Omega}} \hat{u}\, \hat{v}\, \omega$$

- we need to evaluate the matrix $A_{ij} = a(B_i, B_j)$.
- tensor-product basis: $B_i = \beta_{i_1} \cdots \beta_{i_d}$.

$$a(B_i, B_j) \approx \sum_{i=1}^{R} a_r^{(1)}(\beta_{i_1}, \beta_{j_1}) \cdots a_r^{(d)}(\beta_{i_d}, \beta_{j_d})$$

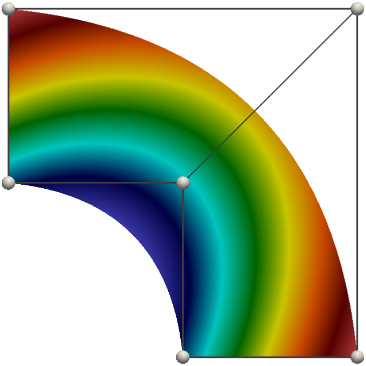
This representation implies that the matrix A can be written in Kronecker product format:

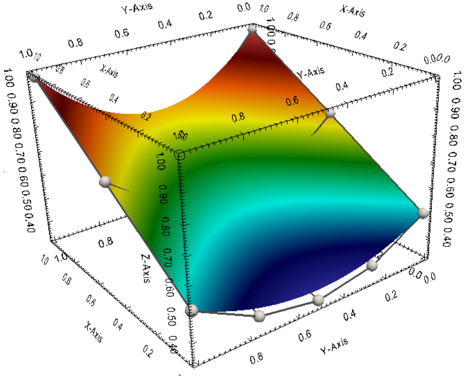
$$A \approx \sum_{r=1}^{R} A_r^{(1)} \otimes \cdots \otimes A_r^{(d)}$$

We call the integer R the $\varepsilon-$Kronecker rank of A.

Example in 2D: Quarter annulus

$$a(\hat{u}, \hat{v}) = \int_{\hat{\Omega}} \nabla \hat{u}^{\mathsf{T}} \cdot \boldsymbol{K} \cdot \nabla \hat{v} + \int_{\hat{\Omega}} \hat{u}\, \hat{v}\, \textcolor{blue}{\omega}$$





$$\omega(\hat{\boldsymbol{x}}) \; = \; \det J(\hat{\boldsymbol{x}}) \; = \; \textcolor{blue}{\omega^{(1)}(\hat{x}_1)}\, \textcolor{red}{\omega^{(2)}(\hat{x}_2)}$$

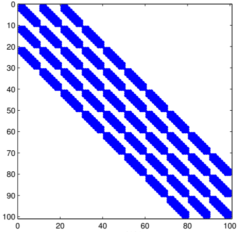
The Kronecker format of Galerkin matrices

- d-D $\rightsquigarrow$ 1-D integrals: $\displaystyle\int_{\hat{\Omega}} \prod_{k=1}^{d} \omega^{(k)}(x_k)\,\mathrm{d}\boldsymbol{x} = \prod_{k=1}^{d} \int_0^1 \omega^{(k)}(x_k)\mathrm{d}x_k$
- Mass matrix:

$$M = \sum_{r=1}^{R} M_r^{(1)} \otimes M_r^{(2)}$$

- Stiffness matrix:

$$S = \sum_{r=1}^{R} S_r^{(1)} \otimes M_r^{(2)} + M_r^{(1)} \otimes S_r^{(2)} + U_r^{(1)} \otimes U_r^{(2)\intercal} + U_r^{(1)\intercal} \otimes U_r^{(2)}$$

$\displaystyle\sum_{r=1}^{R}$  $\otimes$  $=$ 

Computational complexity

- Quadrature with $\mathcal{O}(p)$ nodes per element/direction: $\mathcal{O}(n^d p^{3d})$.
- Lower bound for element-wise strategies: $\mathcal{O}(n^d p^{2d})$
- Optimal complexity with respect to output size $\mathcal{O}(n^d p^d)$.

Using low-rank functions and Kronecker format:

- Tensor decomposition time:
  proportional to the (frequently coarse) geometry mesh
  and the rank parameter. Worst case: $\mathcal{O}(Rdn^d)$.
- "Univariate matrices" (mass, stiffness, advection,..): $\mathcal{O}(Rd\,n\,p^3)$.
- Sparse Kronecker product time $\mathcal{O}(Rn^d p^d)$ $\rightsquigarrow$ dominant cost
- Storage cost: $O(Rdnp)$ for KF vs $O(Rdn^d p^d)$
- Matrix-vector product cost: $O(Rdpn^d)$ for KF vs $O(p^d n^d)$

Error considerations

consistency error = error caused by numerical integration
Total (consistency) error = approximation error + separation error

- Spline projection error in $\Pi(f) = \mathbf{C} : \mathbf{B}$

$$\|f - \Pi(f)\|_\infty \leq \varepsilon_{f,\Pi}$$

- Overall (consistency) error

$$\|f - T_{\mathbb{S}}(\mathbf{C}) : \mathbf{B}\|_\infty \leq \varepsilon_{f,\Pi} + \tau_{\mathbf{C}}$$

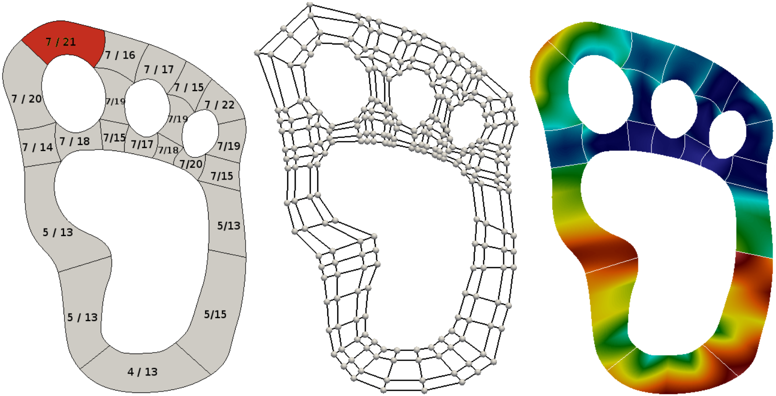
eg. in 2D $\tau_{\mathbf{C}} \leq \sqrt{\sum_{r>R} \sigma_r^2}$ where σ_r are the singular values of **C**.

- Both $\varepsilon_{f,\Pi}$ and $\tau_{\mathbf{C}}$ can be set to required tolerance,
- Knowledge of properties of f allows for a wise choice of Π, eg. $\boldsymbol{K}$, ω have known degree/continuity.

Overview

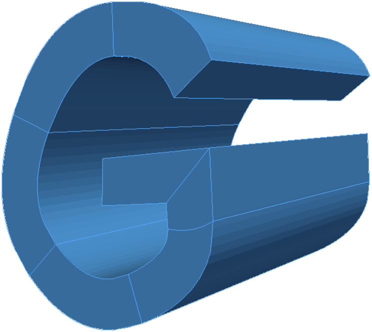
1. Tensor decomposition and rank profiles 2D

mass/stiffness rank | control grids | $|\det J|$

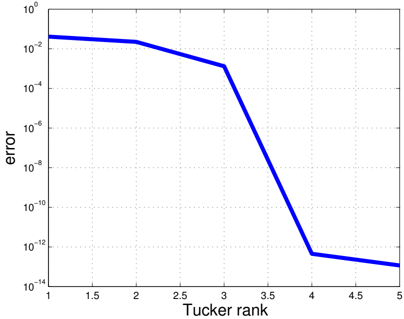


- The rank values lead to significant reduced computation costs
- Rank values can be optimized even further

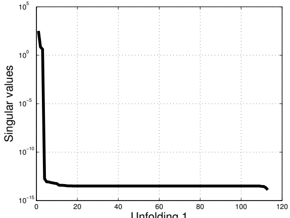
1. Tensor decomposition and rank profiles 3D

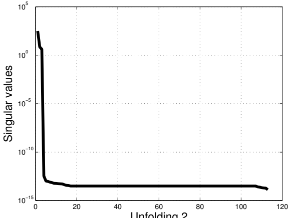


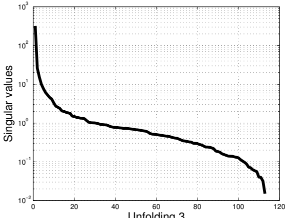
- Tucker rank of $\omega(\hat{\boldsymbol{x}}) = \det G(\hat{\boldsymbol{x}})$



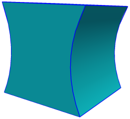
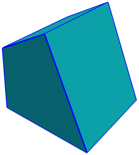
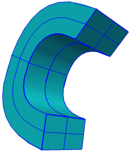
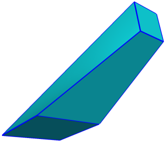
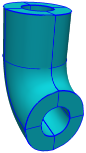
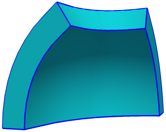
- Singular values of all unfoldings of $f(\hat{\boldsymbol{x}}) = \frac{\partial}{\partial x_1} G_1(\hat{\boldsymbol{x}})$:



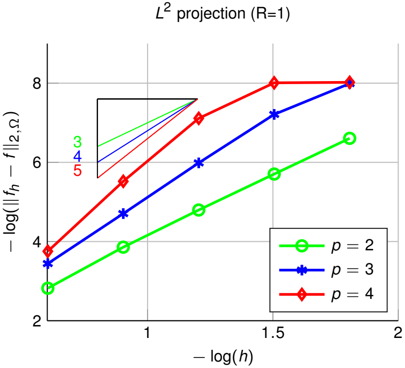


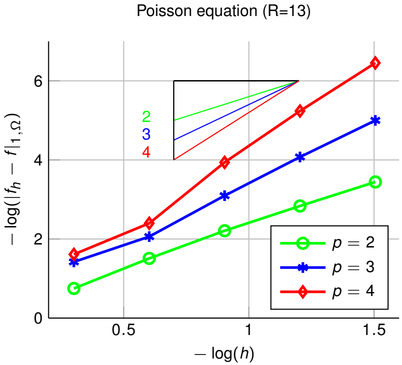


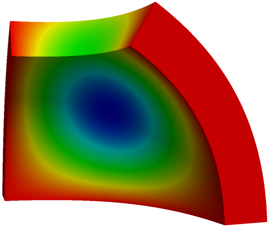
1. Tensor decomposition and rank profiles

Shape	Rank profile		Shape	Rank profile	
	$\omega(\hat{\boldsymbol{x}})$	$K(\hat{\boldsymbol{x}})$		$\omega(\hat{\boldsymbol{x}})$	$K(\hat{\boldsymbol{x}})$
 $\boldsymbol{p} = (1,1,2),\ 2 \times 2 \times 3$	1 $6 \times 6 \times 3$	$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ $5 \times 5 \times 9$	 $\boldsymbol{p} = (1,1,1),\ 2 \times 2 \times 2$	2 $3 \times 3 \times 3$	$\begin{bmatrix} 4 & 4 & 0 \\ 4 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ $5 \times 5 \times 5$
 $\boldsymbol{p} = (2,2,2),\ 4 \times 4 \times 4$	1 $11 \times 11 \times 11$	$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $17 \times 17 \times 17$	 $\boldsymbol{p} = (1,1,1),\ 2 \times 2 \times 2$	2 $3 \times 3 \times 3$	$\begin{bmatrix} 3 & 3 & 3 \\ 3 & 8 & 4 \\ 3 & 4 & 4 \end{bmatrix}$ $5 \times 5 \times 5$
 $\boldsymbol{p} = (2,1,2),\ 9 \times 2 \times 5$	4 $24 \times 3 \times 12$	$\begin{bmatrix} 6 & 6 & 1 \\ 6 & 6 & 1 \\ 1 & 1 & 9 \end{bmatrix}$ $36 \times 5 \times 18$	 $\boldsymbol{p} = (2,2,1),\ 3 \times 3 \times 2$	1 $6 \times 6 \times 3$	$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \end{bmatrix}$ $9 \times 9 \times 5$

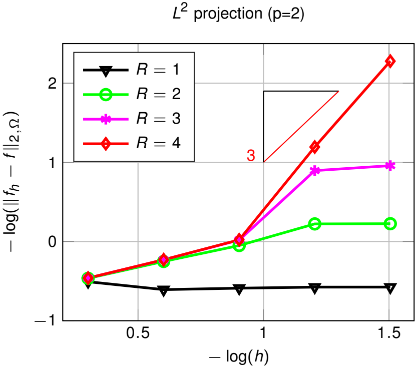
2. Convergence rates

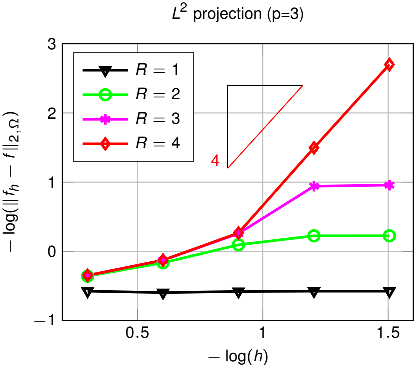


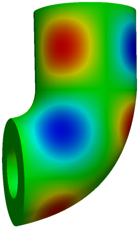




3. Rank truncation



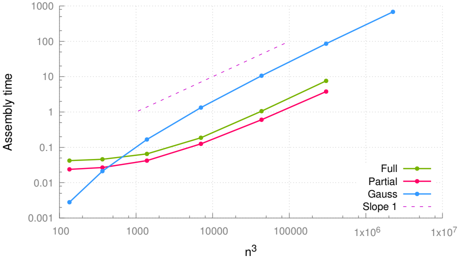




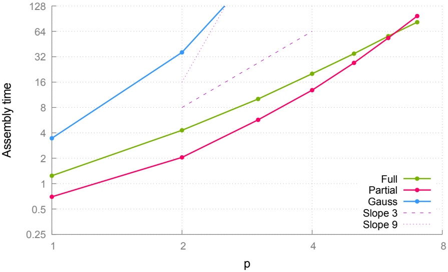
4. Computing times for Stiffness matrix

❒ Computation of “full” matrix

Fix $p = 2$



Fix $50 \times 50 \times 50$ DOFs



❒ Computation of Kronecker factors only