Ordering heuristics for tensor-train decomposition

D. Bigoni (**dabi@mit.edu**), A. Spantini, Y.M. Marzouk

Massachusetts Institute of Technology

SIAM Annual Meeting

Portland, OR – July 10, 2018

Function Approximation – why?

Global knowledge of a potentially
complex or/and
expensive or/and
high-dimensional function

Function Approximation – why?

Global knowledge of a potentially
complex or/and
expensive or/and
high-dimensional function

Function Approximation – how?

Exploit properties of f:

- regularity
- effective dimensionality
- sparsity
- separability (low-rank)

Outline

Tensor-train

Construction

Ordering heuristics

Tensor-train decomposition / Matrix product state

[Bigoni et al., 2016 / Oseledets, 2011 / Fannes et al., 1992 and Klümper et al., 1993]

1. SVD on f with $\mathcal{X} = \mathcal{I}_1$, $\mathcal{Y} = \mathcal{I}_2 \times \mathcal{I}_3$

$$f(x_1, x_2, x_3) = \sum_{\alpha_1=1}^{\infty} \gamma_1(x_1; \alpha_1)\varphi_1(\alpha_1; x_2, x_3)$$

Tensor-train decomposition / Matrix product state

[Bigoni et al., 2016 / Oseledets, 2011 / Fannes et al., 1992 and Klümper et al., 1993]

❶ SVD on f with $\mathcal{X} = \mathcal{I}_1$, $\mathcal{Y} = \mathcal{I}_2 \times \mathcal{I}_3$

$$f(x_1, x_2, x_3) = \sum_{\alpha_1=1}^{\infty} \gamma_1(x_1; \alpha_1)\varphi_1(\alpha_1; x_2, x_3)$$

❷ Truncate: $f(x_1, x_2, x_3) \approx \sum_{\alpha_1=1}^{r_1} \gamma_1(x_1; \alpha_1)\varphi_1(\alpha_1; x_2, x_3)$, $\left(\frac{\|f - f_{r_1}\|}{\|f\|} < \frac{\varepsilon}{\sqrt{2}}\right)$

Tensor-train decomposition / Matrix product state

[Bigoni et al., 2016 / Oseledets, 2011 / Fannes et al., 1992 and Klümper et al., 1993]

1. SVD on f with $\mathcal{X} = \mathcal{I}_1$, $\mathcal{Y} = \mathcal{I}_2 \times \mathcal{I}_3$

$$f(x_1, x_2, x_3) = \sum_{\alpha_1=1}^{\infty} \gamma_1(x_1; \alpha_1)\varphi_1(\alpha_1; x_2, x_3)$$

2. Truncate: $f(x_1, x_2, x_3) \approx \sum_{\alpha_1=1}^{r_1} \gamma_1(x_1; \alpha_1)\varphi_1(\alpha_1; x_2, x_3)$, $\left(\frac{\|f - f_{r_1}\|}{\|f\|} < \frac{\varepsilon}{\sqrt{2}}\right)$
3. SVD on $\varphi_1(\alpha_1; x_2, x_3)$ with $\mathcal{X} = \{1, \ldots, r_1\} \times \mathcal{I}_2$ and $\mathcal{Y} = \mathcal{I}_3$

$$\varphi_1(\alpha_1; x_2, x_3) = \sum_{\alpha_2=1}^{\infty} \gamma_2(\alpha_1; x_2; \alpha_2)\gamma_3(\alpha_2; x_3)$$

Tensor-train decomposition / Matrix product state

[Bigoni et al., 2016 / Oseledets, 2011 / Fannes et al., 1992 and Klümper et al., 1993]

1. SVD on f with $\mathcal{X} = \mathcal{I}_1$, $\mathcal{Y} = \mathcal{I}_2 \times \mathcal{I}_3$

$$f(x_1, x_2, x_3) = \sum_{\alpha_1=1}^{\infty} \gamma_1(x_1; \alpha_1)\varphi_1(\alpha_1; x_2, x_3)$$

2. Truncate: $f(x_1, x_2, x_3) \approx \sum_{\alpha_1=1}^{r_1} \gamma_1(x_1; \alpha_1)\varphi_1(\alpha_1; x_2, x_3)$, $\left(\frac{\|f - f_{r_1}\|}{\|f\|} < \frac{\varepsilon}{\sqrt{2}}\right)$
3. SVD on $\varphi_1(\alpha_1; x_2, x_3)$ with $\mathcal{X} = \{1, \ldots, r_1\} \times \mathcal{I}_2$ and $\mathcal{Y} = \mathcal{I}_3$

$$\varphi_1(\alpha_1; x_2, x_3) = \sum_{\alpha_2=1}^{\infty} \gamma_2(\alpha_1; x_2; \alpha_2)\gamma_3(\alpha_2; x_3)$$

4. Truncate: $\varphi_1(\alpha_1; x_2, x_3) \approx \sum_{\alpha_2=1}^{r_2} \gamma_2(\alpha_1; x_2; \alpha_2)\gamma_3(\alpha_2; x_3)$, $\left(\frac{\|\varphi - \varphi_{r_2}\|}{\|\varphi\|} < \frac{\varepsilon}{\sqrt{2}}\right)$

Tensor-train decomposition / Matrix product state

[Bigoni et al., 2016 / Oseledets, 2011 / Fannes et al., 1992 and Klümper et al., 1993]

1. SVD on f with $\mathcal{X} = \mathcal{I}_1$, $\mathcal{Y} = \mathcal{I}_2 \times \mathcal{I}_3$

$$f(x_1, x_2, x_3) = \sum_{\alpha_1=1}^{\infty} \gamma_1(x_1; \alpha_1)\varphi_1(\alpha_1; x_2, x_3)$$

2. Truncate: $f(x_1, x_2, x_3) \approx \sum_{\alpha_1=1}^{r_1} \gamma_1(x_1; \alpha_1)\varphi_1(\alpha_1; x_2, x_3)$, $\left(\frac{\|f - f_{r_1}\|}{\|f\|} < \frac{\varepsilon}{\sqrt{2}}\right)$
3. SVD on $\varphi_1(\alpha_1; x_2, x_3)$ with $\mathcal{X} = \{1, \ldots, r_1\} \times \mathcal{I}_2$ and $\mathcal{Y} = \mathcal{I}_3$

$$\varphi_1(\alpha_1; x_2, x_3) = \sum_{\alpha_2=1}^{\infty} \gamma_2(\alpha_1; x_2; \alpha_2)\gamma_3(\alpha_2; x_3)$$

4. Truncate: $\varphi_1(\alpha_1; x_2, x_3) \approx \sum_{\alpha_2=1}^{r_2} \gamma_2(\alpha_1; x_2; \alpha_2)\gamma_3(\alpha_2; x_3)$, $\left(\frac{\|\varphi - \varphi_{r_2}\|}{\|\varphi\|} < \frac{\varepsilon}{\sqrt{2}}\right)$
5. Assemble:

$$f(x_1, x_2, x_3) \approx \sum_{\alpha_1=1}^{r_1} \sum_{\alpha_2=1}^{r_2} \gamma_1(x_1; \alpha_1)\gamma_2(\alpha_1; x_2; \alpha_2)\gamma_3(\alpha_2; x_3)$$

Tensor-train decomposition / Matrix product state

[Bigoni et al., 2016 / Oseledets, 2011 / Fannes et al., 1992 and Klümper et al., 1993]

$$f_{\mathrm{TT}}(x_1,\ldots,x_d) := \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_{d-1}=1}^{r_{d-1}} \gamma_1(x_1;\alpha_1)\gamma_2(\alpha_1;x_1;\alpha_2)\cdots\gamma_d(\alpha_{d-1};x_d)$$

Number of basis: $r_1 + \sum_{i=1}^{d-2} r_i r_{i+1} + r_{d-1} \approx \mathcal{O}(dr^2)$

Tensor-train decomposition / Matrix product state

Construction – Density Matrix Renormalization Group

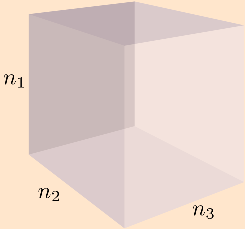
[White, 1993, Savostyanov et al., 2011 and 2014]

$$\boldsymbol{\mathcal{X}} := \bigtimes_{j=1}^{d} \mathbf{x}_j \ , \quad \boldsymbol{\mathcal{W}} := \bigotimes_{j=1}^{d} \mathbf{w}_j \quad \text{and} \quad \boldsymbol{\mathcal{A}} := f(\boldsymbol{\mathcal{X}})\sqrt{\boldsymbol{\mathcal{W}}}$$

TT-DMRG-cross

For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ find $\boldsymbol{\mathcal{I}} = \{\mathcal{I}_1, \dots, \mathcal{I}_{d-2}\}$ and $\boldsymbol{\mathcal{J}} = \{\mathcal{J}_1, \dots, \mathcal{J}_{d-2}\}$ s.t.

$$\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}_{\text{TT}-\text{dmrg}}(\boldsymbol{\mathcal{I}}, \boldsymbol{\mathcal{J}})\|_F \leq \varepsilon \|\boldsymbol{\mathcal{A}}\|_F \quad \left(\text{i.e., } \|f - f_{\text{TT}}\|_{L^2_\mu} \lesssim \varepsilon \|f\|_{L^2_\mu} \right)$$



Tensor-train decomposition / Matrix product state

Construction – Density Matrix Renormalization Group

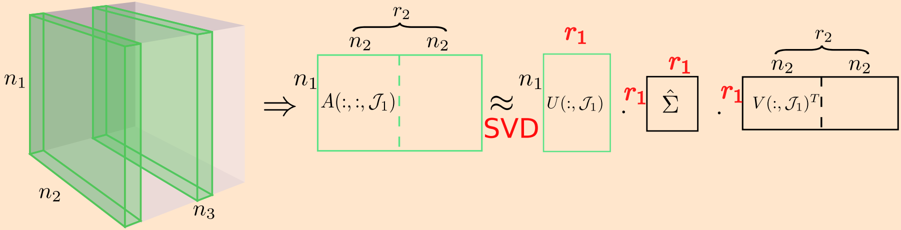
[White, 1993, Savostyanov et al., 2011 and 2014]

$$\boldsymbol{\mathcal{X}} := \bigtimes_{j=1}^{d} \mathbf{x}_j \,, \quad \boldsymbol{\mathcal{W}} := \bigotimes_{j=1}^{d} \mathbf{w}_j \quad \text{and} \quad \boldsymbol{\mathcal{A}} := f(\boldsymbol{\mathcal{X}})\sqrt{\boldsymbol{\mathcal{W}}}$$

TT-DMRG-cross

For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ find $\boldsymbol{\mathcal{I}} = \{\mathcal{I}_1, \ldots, \mathcal{I}_{d-2}\}$ and $\boldsymbol{\mathcal{J}} = \{\mathcal{J}_1, \ldots, \mathcal{J}_{d-2}\}$ s.t.

$$\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}_{\mathrm{TT-dmrg}}(\boldsymbol{\mathcal{I}}, \boldsymbol{\mathcal{J}})\|_F \leq \varepsilon \|\boldsymbol{\mathcal{A}}\|_F \quad \left(\text{i.e., } \|f - f_{\mathrm{TT}}\|_{L^2_\mu} \lesssim \varepsilon \|f\|_{L^2_\mu} \right)$$



Tensor-train decomposition / Matrix product state

Construction – Density Matrix Renormalization Group

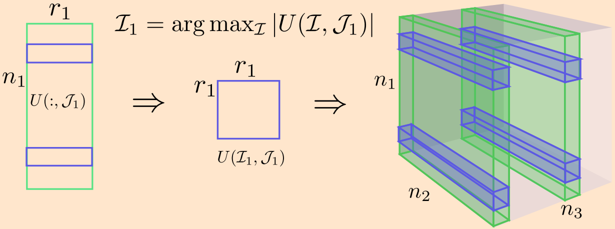
[White, 1993, Savostyanov et al., 2011 and 2014]

$$\mathcal{X} := \bigtimes_{j=1}^{d} \mathbf{x}_j \ , \quad \mathcal{W} := \bigotimes_{j=1}^{d} \mathbf{w}_j \quad \text{and} \quad \mathcal{A} := f(\mathcal{X})\sqrt{\mathcal{W}}$$

TT-DMRG-cross

For $\mathcal{A} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ find $\mathcal{I} = \{\mathcal{I}_1, \ldots, \mathcal{I}_{d-2}\}$ and $\mathcal{J} = \{\mathcal{J}_1, \ldots, \mathcal{J}_{d-2}\}$ s.t.

$$\|\mathcal{A} - \mathcal{A}_{\mathrm{TT-dmrg}}(\mathcal{I}, \mathcal{J})\|_F \le \varepsilon \|\mathcal{A}\|_F \quad \left(\text{i.e., } \|f - f_{\mathrm{TT}}\|_{L^2_\mu} \lesssim \varepsilon \|f\|_{L^2_\mu} \right)$$



Tensor-train decomposition / Matrix product state

Construction – Density Matrix Renormalization Group

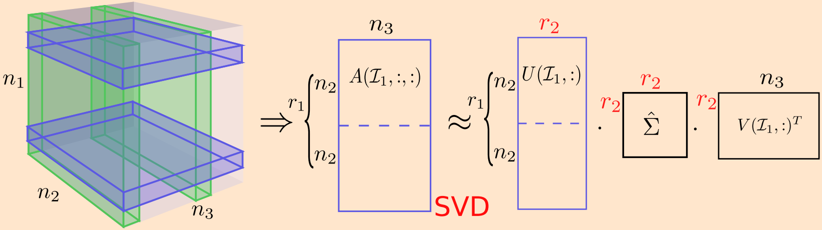
[White, 1993, Savostyanov et al., 2011 and 2014]

$$\boldsymbol{\mathcal{X}} := \bigtimes_{j=1}^{d} \mathbf{x}_j \ , \quad \boldsymbol{\mathcal{W}} := \bigotimes_{j=1}^{d} \mathbf{w}_j \quad \text{and} \quad \boldsymbol{\mathcal{A}} := f(\boldsymbol{\mathcal{X}})\sqrt{\boldsymbol{\mathcal{W}}}$$

TT-DMRG-cross

For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ find $\boldsymbol{\mathcal{I}} = \{\mathcal{I}_1, \ldots, \mathcal{I}_{d-2}\}$ and $\boldsymbol{\mathcal{J}} = \{\mathcal{J}_1, \ldots, \mathcal{J}_{d-2}\}$ s.t.

$$\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}_{\mathrm{TT-dmrg}}(\boldsymbol{\mathcal{I}}, \boldsymbol{\mathcal{J}})\|_F \leq \varepsilon \|\boldsymbol{\mathcal{A}}\|_F \quad \left(\text{i.e., } \|f - f_{\mathrm{TT}}\|_{L^2_\mu} \lesssim \varepsilon \|f\|_{L^2_\mu}\right)$$



Tensor-train decomposition / Matrix product state

Construction – Density Matrix Renormalization Group

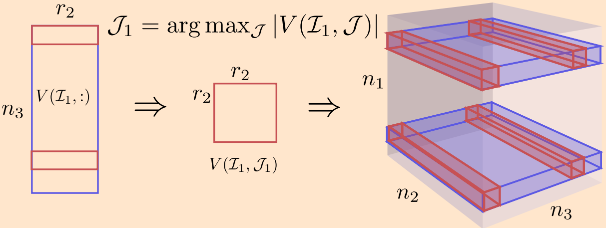
[White, 1993, Savostyanov et al., 2011 and 2014]

$$\boldsymbol{\mathcal{X}} := \bigtimes_{j=1}^{d} \mathbf{x}_j \ , \quad \boldsymbol{\mathcal{W}} := \bigotimes_{j=1}^{d} \mathbf{w}_j \quad \text{and} \quad \boldsymbol{\mathcal{A}} := f(\boldsymbol{\mathcal{X}})\sqrt{\boldsymbol{\mathcal{W}}}$$

TT-DMRG-cross

For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ find $\boldsymbol{\mathcal{I}} = \{\mathcal{I}_1, \ldots, \mathcal{I}_{d-2}\}$ and $\boldsymbol{\mathcal{J}} = \{\mathcal{J}_1, \ldots, \mathcal{J}_{d-2}\}$ s.t.

$$\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}_{\mathrm{TT-dmrg}}(\boldsymbol{\mathcal{I}}, \boldsymbol{\mathcal{J}})\|_F \leq \varepsilon \|\boldsymbol{\mathcal{A}}\|_F \quad \left(\text{i.e., } \|f - f_{\mathrm{TT}}\|_{L^2_\mu} \lesssim \varepsilon \|f\|_{L^2_\mu} \right)$$



Tensor-train decomposition / Matrix product state

Construction – Density Matrix Renormalization Group

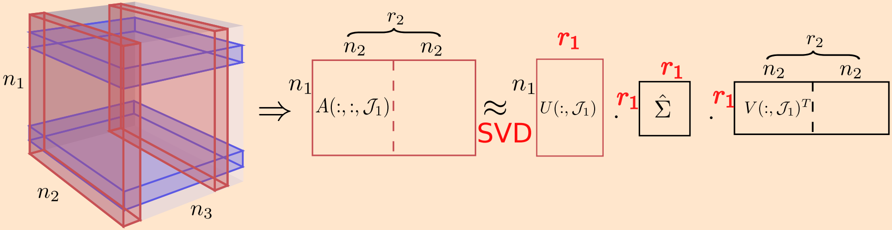
[White, 1993, Savostyanov et al., 2011 and 2014]

$$\boldsymbol{\mathcal{X}} := \bigtimes_{j=1}^{d} \mathbf{x}_j \, , \quad \boldsymbol{\mathcal{W}} := \bigotimes_{j=1}^{d} \mathbf{w}_j \quad \text{and} \quad \boldsymbol{\mathcal{A}} := f(\boldsymbol{\mathcal{X}})\sqrt{\boldsymbol{\mathcal{W}}}$$

TT-DMRG-cross

For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ find $\boldsymbol{\mathcal{I}} = \{\mathcal{I}_1, \ldots, \mathcal{I}_{d-2}\}$ and $\boldsymbol{\mathcal{J}} = \{\mathcal{J}_1, \ldots, \mathcal{J}_{d-2}\}$ s.t.

$$\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}_{\mathrm{TT-dmrg}}(\boldsymbol{\mathcal{I}}, \boldsymbol{\mathcal{J}})\|_F \leq \varepsilon \|\boldsymbol{\mathcal{A}}\|_F \quad \left(\text{i.e., } \|f - f_{\mathrm{TT}}\|_{L^2_\mu} \lesssim \varepsilon \|f\|_{L^2_\mu} \right)$$



Tensor-train decomposition / Matrix product state

Construction – Density Matrix Renormalization Group

[White, 1993, Savostyanov et al., 2011 and 2014]

$$\boldsymbol{\mathcal{X}} := \bigtimes_{j=1}^{d} \mathbf{x}_j \ , \quad \boldsymbol{\mathcal{W}} := \bigotimes_{j=1}^{d} \mathbf{w}_j \quad \text{and} \quad \boldsymbol{\mathcal{A}} := f(\boldsymbol{\mathcal{X}})\sqrt{\boldsymbol{\mathcal{W}}}$$

TT-DMRG-cross

For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ find $\boldsymbol{\mathcal{I}} = \{\mathcal{I}_1, \ldots, \mathcal{I}_{d-2}\}$ and $\boldsymbol{\mathcal{J}} = \{\mathcal{J}_1, \ldots, \mathcal{J}_{d-2}\}$ s.t.

$$\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}_{\mathrm{TT-dmrg}}(\boldsymbol{\mathcal{I}}, \boldsymbol{\mathcal{J}})\|_F \leq \varepsilon \, \|\boldsymbol{\mathcal{A}}\|_F \quad \left(\text{i.e., } \|f - f_{\mathrm{TT}}\|_{L^2_\mu} \lesssim \varepsilon \|f\|_{L^2_\mu} \right)$$

+ Evaluation of $\mathcal{O}(ndr^2)$ elements for $d \gg 1$
+ Rank revealing

The ordering problem

An example

$$f(x_1, x_2, x_3) = 2x_1x_2 + 2x_2x_3 + \frac{1}{x_1 + x_3 + 1}\ , \quad 0 \leq x_i \leq 1$$

An example

$$f(x_1, x_2, x_3) = 2x_1x_2 + 2x_2x_3 + \frac{1}{x_1 + x_3 + 1}, \quad 0 \leq x_i \leq 1$$

Permutation set: $\Sigma := \{\sigma : \{1, \dots, d\} \to \{1, \dots, d\} | \sigma \text{ is a bijection}\}$

$\sigma^{-1}([1,2,3])$	TT-ranks	#F.eval.	% Fill lev.
$[1,2,3]$	$[\mathbf{1},\mathbf{7},\mathbf{7},\mathbf{1}]$	8676	26.48%
$[1,3,2]$	$[1,7,2,1]$	6767	20.65%
$[2,1,3]$	$[1,2,7,1]$	6721	20.51%
$[2,3,1]$	$[1,7,2,1]$	4822	14.72%
$[3,1,2]$	$[1,2,7,1]$	6840	20.87%
$[3,2,1]$	$[\mathbf{1},\mathbf{7},\mathbf{7},\mathbf{1}]$	8080	24.66%

An example

$$f(x_1, x_2, x_3) = 2x_1x_2 + 2x_2x_3 + \frac{1}{x_1 + x_3 + 1} , \quad 0 \leq x_i \leq 1$$

Sensitivity analysis: two-way interactions with method of Sobol'

$$s_{ij}^2 := \frac{\mathbb{V}_{x_i,x_j}\left[\mathbb{E}_{x_k}\left[f(x_1, x_2, x_3)\right] - \mathbb{E}\left[f(x_1, x_2, x_3)\right]\right]}{\mathbb{V}\left[f(x_1, x_2, x_3)\right]} , \quad k \neq i, j$$

Variables	s_{ij}^2
x_1, x_2	0.060
x_1, x_3	0.003
x_2, x_3	0.060

A more involved example...

Permutation : $\sigma \in \Sigma := \{\sigma : \{0, \dots, d-1\} \to \{0, \dots, d-1\} | \sigma \text{ is a bijection}\}$

Partition : $p \leq d$, $\mathbf{n} = [n_0 < n_1 < \dots < n_p]$, $d = \sum_{i=1}^{p} d_i = \sum_{i=1}^{p} (n_i - n_{i-1})$

Let $g : \mathbb{R}^p \to \mathbb{R}$ be a low-rank function.
Let $h_i : \mathbb{R}^{d_i} \to \mathbb{R}$ for $i \in \{1, \dots, p\}$ be high-rank/low-rank functions.

$$f(\mathbf{x}) := g\left(h_1(x_{\sigma(n_0)}, \dots, x_{\sigma(n_1)}), \dots, h_p(x_{\sigma(n_{p-1})}, \dots, x_{\sigma(n_p)})\right)$$

e.g. $g(\mathbf{y}) := y_0 \cdots y_{p-1}$ and $y_i := h_i(\mathbf{z}) := \left(1 + \sum_{j=0}^{d_i - 1} z_j\right)^{-d_i+1}$

A more involved example...

$$f(\mathbf{x}) := g\left(h_1(x_{\sigma(n_0)}, \ldots, x_{\sigma(n_1)}), \ldots, h_p(x_{\sigma(n_{p-1})}, \ldots, x_{\sigma(n_p)})\right)$$

$$\text{e.g. } g(\mathbf{y}) := y_0 \cdots y_{p-1} \quad \text{and} \quad h_i(\mathbf{z}) := \left(1 + \sum_{j=0}^{d_i-1} z_j\right)^{-d_i+1}$$

$$d = 10\ , \quad \sigma([0, \ldots, 9]) := [[6, 7, 5, 0], [8, 3, 4, 1], [2, 9]]$$

Discretization: Legendre polynomials, $n = 32$ Gauss points.

A more involved example...

$$f(\mathbf{x}) := g\left(h_1(x_{\sigma(n_0)}, \ldots, x_{\sigma(n_1)}), \ldots, h_p(x_{\sigma(n_{p-1})}, \ldots, x_{\sigma(n_p)})\right)$$

$$\text{e.g. } g(\mathbf{y}) := y_0 \cdots y_{p-1} \quad \text{and} \quad h_i(\mathbf{z}) := \left(1 + \sum_{j=0}^{d_i-1} z_j\right)^{-d_i+1}$$

$$d = 10\ , \quad \sigma([0, \ldots, 9]) := [[6,7,5,0],[8,3,4,1],[2,9]]$$

Discretization: Legendre polynomials, $n = 32$ Gauss points.

	Blind `TT-dmrg`	**Ordered** `TT-dmrg`
Ranks	$[1,5,19,38,33,31,29,27,12,4,1]$	$[1,5,5,5,1,5,6,6,1,4,1]$
# f **eval.**	$1942411 \approx 1.9 \times 10^6$	$70088 \approx 7.0 \times 10^4$
% f **eval.**	$1.72 \times 10^{-7}\%$	$6.23^{-9}\%$

Objective

Use **minimal** information to estimate
the **optimal** ordering
for **minimal** tensor-train storage

Objective

Use **minimal** information to estimate
the **optimal** ordering
for **minimal** tensor-train storage

$$\sigma^* = \arg\min_{\sigma \in \Sigma} c(\sigma) \ ,$$

$$c(\sigma) := r_1^\sigma + \sum_{i=2}^{d-1} r_{i-1}^\sigma r_i^\sigma + r_{d-1}^\sigma \ .$$

Objective

Use **minimal** information to estimate
the **optimal** ordering
for **minimal** tensor-train storage

$$\sigma^* = \arg\min_{\sigma \in \Sigma} c(\sigma) \ ,$$

$$c(\sigma) := r_1^\sigma + \sum_{i=2}^{d-1} r_{i-1}^\sigma r_i^\sigma + r_{d-1}^\sigma \ .$$

The problem is combinatorial and can be challenging to solve a priori!

Objectives

Additive assumption : $f^A(\mathbf{x}) = \sum_{i=0}^{d-2} \sum_{j=i+1}^{d-1} h_{ij}(x_i, x_j)$

Use **second order rank** information to estimate
the **optimal** ordering
for **minimal** tensor-train storage
of a particular class of functions

$$\sigma^* = \arg\min_{\sigma \in \Sigma} c(\sigma) \ ,$$

$$c(\sigma) := \tilde{r}_1^\sigma + \sum_{i=2}^{d-1} \tilde{r}_{i-1}^\sigma \tilde{r}_i^\sigma + \tilde{r}_{d-1}^\sigma \ .$$

Objectives

Additive assumption : $f^A(\mathbf{x}) = \sum_{i=0}^{d-2} \sum_{j=i+1}^{d-1} h_{ij}(x_i, x_j)$

Use **second order rank** information to estimate
the **optimal** ordering
for **minimal** tensor-train storage
of a particular class of functions

$$\sigma^* = \arg\min_{\sigma \in \Sigma} c(\sigma) \; ,$$

$$c(\sigma) := \tilde{r}_1^\sigma + \sum_{i=2}^{d-1} \tilde{r}_{i-1}^\sigma \tilde{r}_i^\sigma + \tilde{r}_{d-1}^\sigma \; .$$

The problem is still combinatorial but computable!

Objectives

Use **second order rank** information to estimate
the **sub-optimal** ordering
for **favourable** tensor-train storage
of a particular class of functions

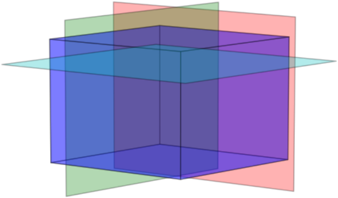
$$\sigma^* \approx \arg\min_{\sigma\in\Sigma} c(\sigma) \ ,$$

$$c(\sigma) := \tilde{r}_1^\sigma + \sum_{i=2}^{d-1} \tilde{r}_{i-1}^\sigma \tilde{r}_i^\sigma + \tilde{r}_{d-1}^\sigma \ .$$

Second order rank information

Let allow ourself to **sample "few" slices** of the tensor.

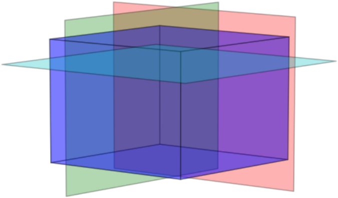
function $\textsc{Order2RankEst}(\boldsymbol{\mathcal{A}}, n_{\min}, n_{\max})$
    for $0 \leq i < j < d$ **do**
        $R[i,j] \leftarrow \mathbb{E}\left[\mathsf{rank}(\boldsymbol{\mathcal{A}}^{(i,j)})\right]$
        $V[i,j] \leftarrow \mathbb{V}\left[\mathsf{rank}(\boldsymbol{\mathcal{A}}^{(i,j)})\right]$
    end for
    return (R, V)
end function

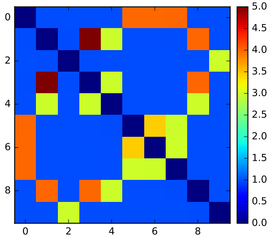


Second order rank information

Let allow ourself to **sample "few" slices** of the tensor.

function $\textsc{Order2RankEst}(\boldsymbol{\mathcal{A}}, n_{\min}, n_{\max})$
  for $0 \leq i < j < d$ **do**
    $R[i,j] \leftarrow \mathbb{E}\left[\text{rank}(\boldsymbol{\mathcal{A}}^{(i,j)})\right]$
    $V[i,j] \leftarrow \mathbb{V}\left[\text{rank}(\boldsymbol{\mathcal{A}}^{(i,j)})\right]$
  end for
  return (R, V)
end function





$$f(\mathbf{x}) := g\left(h_1(x_{\sigma(n_0)}, \ldots, x_{\sigma(n_1)}), \ldots, h_p(\ldots)\right)$$

$$g(\mathbf{y}) := y_0 \cdots y_{p-1} \quad \text{and} \quad h_i(\mathbf{z}) := \left(1 + \sum_{j=0}^{d_i-1} z_j\right)^{-d_i+1}$$

$$\sigma([0,\ldots,9]) := [[6,7,5,0],[8,3,4,1],[2,9]]$$

Second order interaction assumption

Additive assumption : $f^A(\mathbf{x}) = \sum_{i=0}^{d-2} \sum_{j=i+1}^{d-1} h_{ij}(x_i, x_j)$

- Rank matrix R contains the **numerical ranks** r_{ij} such that

$$h_{ij}(x_i, x_j) \approx \sum_{k_{ij}=1}^{r_{ij}} \phi_i^{(i,j)}(k_{ij}; x_i)\phi_j^{(i,j)}(k_{ij}; x_j)$$

- Variance matrix V expresses the **confidence** on the assumption

Second order interaction assumption

Additive assumption : $f^A(\mathbf{x}) = \sum_{i=0}^{d-2} \sum_{j=i+1}^{d-1} h_{ij}(x_i, x_j)$

- Rank matrix R contains the **numerical ranks** r_{ij} such that

$$h_{ij}(x_i, x_j) \approx \sum_{k_{ij}=1}^{r_{ij}} \phi_i^{(i,j)}(k_{ij}; x_i)\phi_j^{(i,j)}(k_{ij}; x_j)$$

- Variance matrix V expresses the **confidence** on the assumption

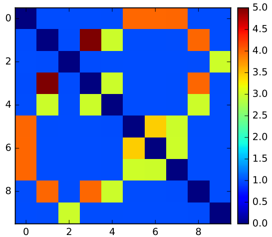
What is the **upper bound** on

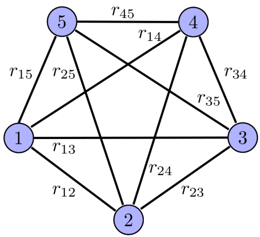
$$c(\sigma) := \tilde{r}_1^\sigma + \sum_{i=2}^{d-1} \tilde{r}_{i-1}^\sigma \tilde{r}_i^\sigma + \tilde{r}_{d-1}^\sigma \ ,$$

for a given permutation/ordering σ?

Second order interaction assumption – The rank graph

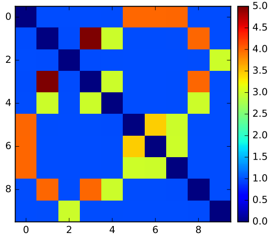
$$f^A(\mathbf{x}) = \sum_{i<j}^{d-1} h_{ij}(x_i, x_j)$$

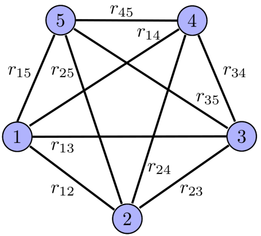




Second order interaction assumption – The rank graph

$$f^A(\mathbf{x}) = \sum_{i<j}^{d-1} h_{ij}(x_i, x_j)$$



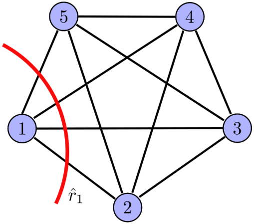


Let $\mathcal{X}_1 = (x_1)$ and $\mathcal{Y}_1 = (x_2, \ldots, x_d)$.

$$f_1^A(\mathcal{X}_1, \mathcal{Y}_1) = g_c(\mathcal{X}_1, \mathcal{Y}_1) + g_r(\mathcal{Y}_1) \ ,$$

$$g_c(\mathcal{X}_1, \mathcal{Y}_1) = \sum_{j=2}^{d} h_{1j}(x_1, x_j) \ ,$$

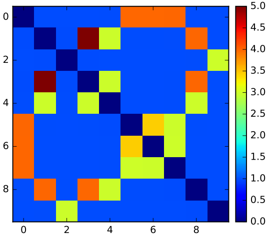
$$g_r(\mathcal{Y}_1) = \sum_{i=2}^{d-1} \sum_{j=i+1}^{d} h_{ij}(x_i, x_j) \ .$$

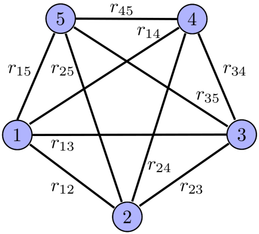


$$\hat{r}_1 := \sum_{j=2}^{d} r_{1j} + 1$$

Second order interaction assumption – The rank graph

$$f^A(\mathbf{x}) = \sum_{i<j}^{d-1} h_{ij}(x_i, x_j)$$





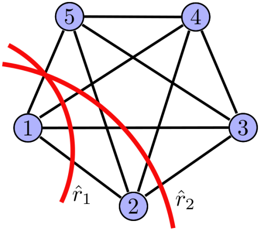
Let $\mathcal{X}_2 = (x_1, x_2)$ and $\mathcal{Y}_2 = (x_3, \dots, x_d)$.

$$f_2^A(\mathcal{X}_2, \mathcal{Y}_2) = g_l(\mathcal{X}_2) + g_c(\mathcal{X}_2, \mathcal{Y}_2) + g_r(\mathcal{Y}_2) \ ,$$

$$g_l(\mathcal{X}_2) = h_{12}(x_1, x_2)$$

$$g_c(\mathcal{X}_2, \mathcal{Y}_2) = \sum_{i=1}^{2} \sum_{j=3}^{d} h_{ij}(x_i, x_j) \ ,$$

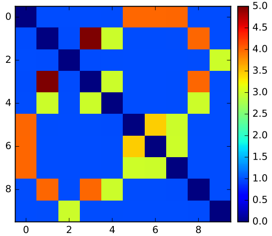
$$g_r(\mathcal{Y}_2) = \sum_{i=3}^{d-1} \sum_{j=i+1}^{d} h_{ij}(x_i, x_j) \ .$$

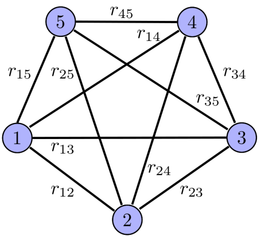


$$\hat{r}_2 := \sum_{i=1}^{2} \sum_{j=3}^{d} r_{ij} + 2$$

Second order interaction assumption – The rank graph

$$f^A(\mathbf{x}) = \sum_{i<j}^{d-1} h_{ij}(x_i, x_j)$$



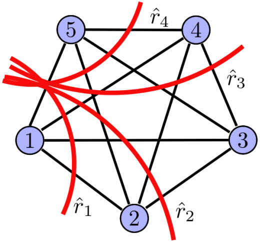


Let $\mathcal{X}_{d-1} = (x_1, \ldots, x_{d-1})$ and $\mathcal{Y}_{d-1} = (x_d)$.

$$f^A_{d-1}(\mathcal{X}_{d-1}, \mathcal{Y}_{d-1}) = g_l(\mathcal{X}_{d-1}) + g_c(\mathcal{X}_{d-1}, \mathcal{Y}_{d-1}) \ ,$$

$$g_l(\mathcal{X}_{d-1}) = \sum_{i=1}^{d-2} \sum_{j=i+1}^{d-1} h_{ij}(x_i, x_j)$$

$$g_c(\mathcal{X}_{d-1}, \mathcal{Y}_{d-1}) = \sum_{i=1}^{d-1} h_{id}(x_i, x_d) \ ,$$



$$\hat{r}_{d-1} := \sum_{i=1}^{d-1} r_{id} + 1$$

Upper bound of tensor-train storage/efficiency

Additive assumption : $f^A(\mathbf{x}) = \sum_{i=0}^{d-2} \sum_{j=i+1}^{d-1} h_{ij}(x_i, x_j)$

$$c(\sigma) \leq \hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma + C \ ,$$

$$\hat{r}_k^\sigma := \sum_{i=1}^{k} \sum_{j=k+1}^{d} r_{\sigma(i)\sigma(j)}$$

Approaches to the "relaxed" ordering problem

Additive assumption : $f^A(\mathbf{x}) = \sum_{i=0}^{d-2} \sum_{j=i+1}^{d-1} h_{ij}(x_i, x_j)$

$$\sigma^* = \arg\min_{\sigma \in \Sigma} \hat{c}(\sigma) \ ,$$

$$\hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma \ .$$

Approaches to the "relaxed" ordering problem

Additive assumption : $f^A(\mathbf{x}) = \sum_{i=0}^{d-2} \sum_{j=i+1}^{d-1} h_{ij}(x_i, x_j)$

$$\sigma^* = \arg\min_{\sigma\in\Sigma} \hat{c}(\sigma) \ ,$$

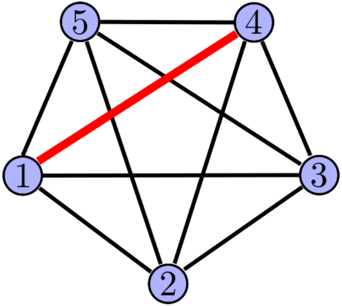
$$\hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma \ .$$

1. Hierarchical Clustering
2. Dynamic programming (greedy, branch and bound)

Hierarchical clustering approach – The intuition

Let $d = 5$, $r_{14} \gg r_{ij}$ and $\sigma := \mathbb{I}$, i.e.

$$\sigma([1,2,3,4,5]) = [1,2,3,4,5]$$

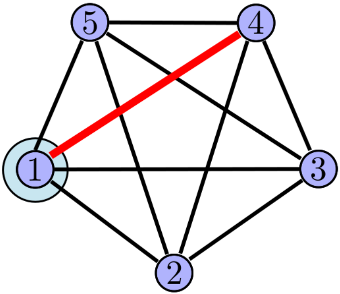


$$\hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma \, .$$

Hierarchical clustering approach – The intuition

Let $d = 5$, $r_{14} \gg r_{ij}$ and $\sigma := \mathbb{I}$, i.e.

$$\sigma([1,2,3,4,5]) = [1,2,3,4,5]$$



1. $\hat{r}_1^\sigma = r_{1,2} + r_{1,3} + \mathbf{r_{1,4}} + r_{1,5}$

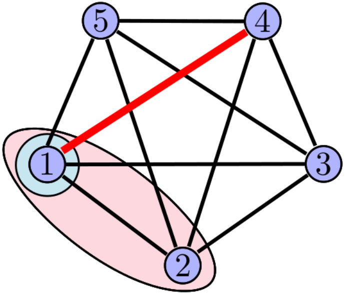
$$\hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma \ .$$

Hierarchical clustering approach – The intuition

Let $d = 5$, $r_{14} \gg r_{ij}$ and $\sigma := \mathbb{I}$, i.e.

$$\sigma([1,2,3,4,5]) = [1,2,3,4,5]$$

1. $\hat{r}_1^\sigma = r_{1,2} + r_{1,3} + \mathbf{r_{1,4}} + r_{1,5}$
2. $\hat{r}_2^\sigma = r_{1,3} + \mathbf{r_{1,4}} + r_{1,5} + r_{2,3} + r_{2,4} + r_{2,5}$



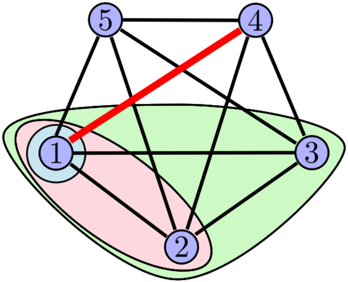
$$\hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma \ .$$

Hierarchical clustering approach – The intuition

Let $d = 5$, $r_{14} \gg r_{ij}$ and $\sigma := \mathbb{I}$, i.e.

$$\sigma([1,2,3,4,5]) = [1,2,3,4,5]$$

1. $\hat{r}_1^\sigma = r_{1,2} + r_{1,3} + \mathbf{r_{1,4}} + r_{1,5}$
2. $\hat{r}_2^\sigma = r_{1,3} + \mathbf{r_{1,4}} + r_{1,5} + r_{2,3} + r_{2,4} + r_{2,5}$
3. $\hat{r}_3^\sigma = \mathbf{r_{1,4}} + r_{1,5} + r_{2,4} + r_{2,5} + r_{3,4} + r_{3,5}$



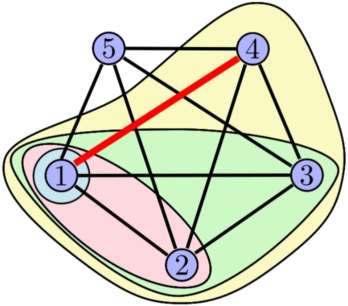
$$\hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma \ .$$

Hierarchical clustering approach – The intuition

Let $d = 5$, $r_{14} \gg r_{ij}$ and $\sigma := \mathbb{I}$, i.e.

$$\sigma([1,2,3,4,5]) = [1,2,3,4,5]$$

1. $\hat{r}_1^\sigma = r_{1,2} + r_{1,3} + \mathbf{r_{1,4}} + r_{1,5}$
2. $\hat{r}_2^\sigma = r_{1,3} + \mathbf{r_{1,4}} + r_{1,5} + r_{2,3} + r_{2,4} + r_{2,5}$
3. $\hat{r}_3^\sigma = \mathbf{r_{1,4}} + r_{1,5} + r_{2,4} + r_{2,5} + r_{3,4} + r_{3,5}$
4. $\hat{r}_4^\sigma = r_{1,5} + r_{2,5} + r_{3,5} + r_{4,5}$

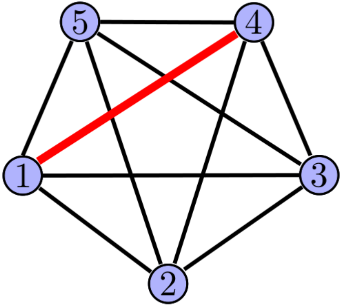


$$\hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma \,.$$

Hierarchical clustering approach – The intuition

Let $d = 5$, $r_{14} \gg r_{ij}$ and $\sigma := \mathbb{I}$, i.e.

$$\sigma([1,2,3,4,5]) = [1,4,5,2,3]$$



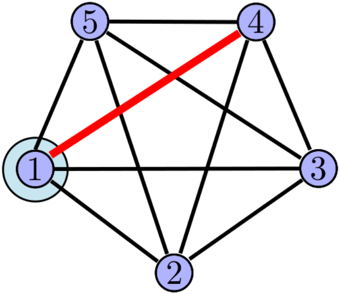
$$\hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma \ .$$

Hierarchical clustering approach – The intuition

Let $d = 5$, $r_{14} \gg r_{ij}$ and $\sigma := \mathbb{I}$, i.e.

$$\sigma([1,2,3,4,5]) = [1,4,5,2,3]$$

1. $\hat{r}_1^\sigma = r_{1,2} + r_{1,3} + \mathbf{r_{1,4}} + r_{1,5}$



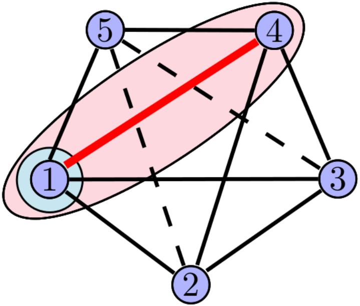
$$\hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma \ .$$

Hierarchical clustering approach – The intuition

Let $d = 5$, $r_{14} \gg r_{ij}$ and $\sigma := \mathbb{I}$, i.e.

$$\sigma([1,2,3,4,5]) = [1,4,5,2,3]$$

1. $\hat{r}_1^\sigma = r_{1,2} + r_{1,3} + \mathbf{r_{1,4}} + r_{1,5}$
2. $\hat{r}_2^\sigma = r_{1,2} + r_{1,3} + r_{1,5} + r_{2,4} + r_{3,4} + r_{4,5}$



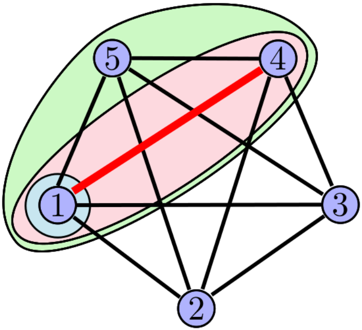
$$\hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma \ .$$

Hierarchical clustering approach – The intuition

Let $d = 5$, $r_{14} \gg r_{ij}$ and $\sigma := \mathbb{I}$, i.e.

$$\sigma([1,2,3,4,5]) = [1,4,5,2,3]$$

1. $\hat{r}_1^\sigma = r_{1,2} + r_{1,3} + \mathbf{r_{1,4}} + r_{1,5}$
2. $\hat{r}_2^\sigma = r_{1,2} + r_{1,3} + r_{1,5} + r_{2,4} + r_{3,4} + r_{4,5}$
3. $\hat{r}_3^\sigma = r_{1,2} + r_{1,3} + r_{2,4} + r_{3,4} + r_{2,5} + r_{3,5}$

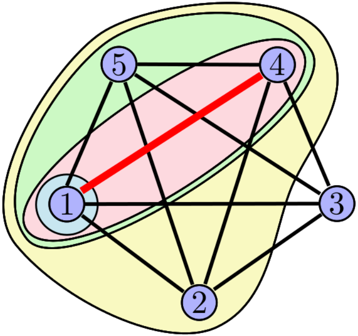


$$\hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma .$$

Hierarchical clustering approach – The intuition

Let $d = 5$, $r_{14} \gg r_{ij}$ and $\sigma := \mathbb{I}$, i.e.

$$\sigma([1,2,3,4,5]) = [1,4,5,2,3]$$



1. $\hat{r}_1^\sigma = r_{1,2} + r_{1,3} + \mathbf{r_{1,4}} + r_{1,5}$
2. $\hat{r}_2^\sigma = r_{1,2} + r_{1,3} + r_{1,5} + r_{2,4} + r_{3,4} + r_{4,5}$
3. $\hat{r}_3^\sigma = r_{1,2} + r_{1,3} + r_{2,4} + r_{3,4} + r_{2,5} + r_{3,5}$
4. $\hat{r}_4^\sigma = r_{1,3} + r_{2,3} + r_{3,4} + r_{3,5}$

$$\hat{c}(\sigma) := \hat{r}_1^\sigma + \sum_{i=2}^{d-1} \hat{r}_{i-1}^\sigma \hat{r}_i^\sigma + \hat{r}_{d-1}^\sigma \,.$$

Hierarchical clustering approach – The method

Hierarchical clustering: generate a hierarchy of clusters minimizing the "distance" within the clusters and maximizing the "distance" among the clusters.

Hierarchical clustering approach – The method

Hierarchical clustering: generate a hierarchy of clusters minimizing the "distance" within the clusters and maximizing the "distance" among the clusters.

Rank graph

$$\text{dist}(i,j) = r_{ij}$$

High ranks
connect
distant nodes

→

Proximity graph

$$\text{dist}(i,j) = r_{\max} - r_{ij} + 1$$

High ranks
connect
close nodes

Hierarchical clustering approach – The method

Hierarchical clustering: generate a hierarchy of clusters minimizing the "distance" within the clusters and maximizing the "distance" among the clusters.

Rank graph

$$\text{dist}(i,j) = r_{ij}$$

High ranks
connect
distant nodes

→

Proximity graph

$$\text{dist}(i,j) = r_{\max} - r_{ij} + 1$$

High ranks
connect
close nodes

Cluster distance

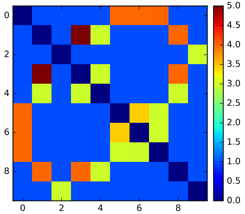
$$d(\mathcal{S},\mathcal{T}) = \sum_{\substack{s\in\mathcal{S}\\ t\in\mathcal{T}}} \frac{\text{dist}(s,t)}{|\mathcal{S}|\cdot|\mathcal{T}|}$$

→

Clustering result

Max. intra-clusters distance
corresponds to
Min. intra-clusters rank cuts

Hierarchical clustering approach – Example

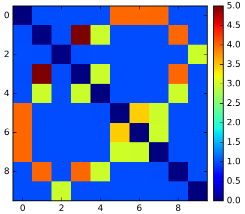


$$f(\mathbf{x}) := g\left(h_1(x_{\sigma(n_0)}, \ldots, x_{\sigma(n_1)}), \ldots, h_p(\ldots)\right)$$

$$g(\mathbf{y}) := y_0 \cdots y_{p-1} \quad \text{and} \quad h_i(\mathbf{z}) := \left(1 + \sum_{j=0}^{d_i - 1} z_j\right)^{-d_i+1}$$

$$\sigma([0, \ldots, 9]) := [[6, 7, 5, 0], [8, 3, 4, 1], [2, 9]]$$

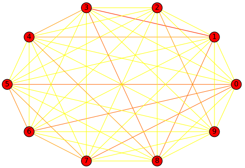
Hierarchical clustering approach – Example

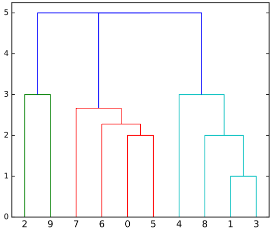


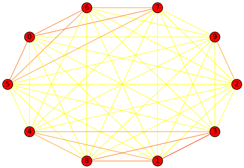
$$f(\mathbf{x}) := g\left(h_1(x_{\sigma(n_0)}, \ldots, x_{\sigma(n_1)}), \ldots, h_p(\ldots)\right)$$

$$g(\mathbf{y}) := y_0 \cdots y_{p-1} \quad \text{and} \quad h_i(\mathbf{z}) := \left(1 + \sum_{j=0}^{d_i-1} z_j\right)^{-d_i+1}$$

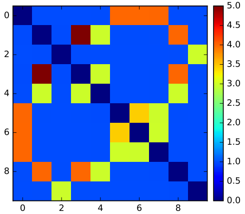
$$\sigma([0,\ldots,9]) := [[6,7,5,0],[8,3,4,1],[2,9]]$$







Hierarchical clustering approach – Example



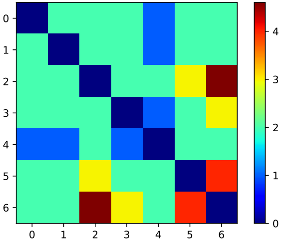
$$f(\mathbf{x}) := g\left(h_1(x_{\sigma(n_0)}, \ldots, x_{\sigma(n_1)}), \ldots, h_p(\ldots)\right)$$

$$g(\mathbf{y}) := y_0 \cdots y_{p-1} \quad \text{and} \quad h_i(\mathbf{z}) := \left(1 + \sum_{j=0}^{d_i-1} z_j\right)^{-d_i+1}$$

$$\sigma([0,\ldots,9]) := [[6,7,5,0],[8,3,4,1],[2,9]]$$

	Blind `TT-dmrg`	**Ordered** `TT-dmrg`
Ranks	$[1,5,19,38,33,31,29,27,12,4,1]$	$[1,4,1,4,5,5,1,4,5,6,1]$
# f **eval.**	$1942411 \approx 1.9 \times 10^6$	$62916 \approx 6.2 \times 10^4$
% f **eval.**	$1.72 \times 10^{-7}\%$	$5.59 \times 10^{-9}\%$

Hierarchical clustering approach – Piston Simulation Function

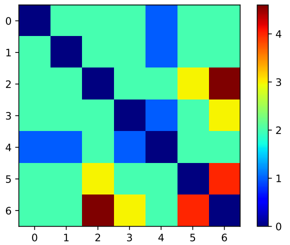


$$C(\mathbf{x}) = 2\pi \sqrt{\frac{M}{k + S^2 \frac{P_0 V_0}{T_0} \frac{T_a}{V^2}}}$$

$$V = \frac{S}{2k} \left(\sqrt{A^2 + 4k \frac{P_0 V_0}{T_0} T_a} - A \right)$$

$$A = P_0 S + 19.62 M - \frac{k V_0}{S}$$

Hierarchical clustering approach – Piston Simulation Function



$$C(\mathbf{x}) = 2\pi \sqrt{\frac{M}{k + S^2 \frac{P_0 V_0}{T_0} \frac{T_a}{V^2}}}$$

$$V = \frac{S}{2k}\left(\sqrt{A^2 + 4k\frac{P_0 V_0}{T_0} T_a} - A\right)$$

$$A = P_0 S + 19.62M - \frac{kV_0}{S}$$

	Blind `TT-dmrg`	**Ordered** `TT-dmrg`
Ranks	$[1,3,4,11,12,11,6,1]$	$[1,2,2,4,7,7,6,1]$
# f **eval.**	$633875 \approx 6.3 \times 10^5$	$218041 \approx 2.2 \times 10^5$
% f **eval.**	$2.36 \times 10^{-1}\%$	$7.50 \times 10^{-2}\%$

Conclusions

- Characterization of the ordering problem in TT/MPS decompositions
- Proposal of a pre-processing ordering heuristic
- Clustering approach to the solution of the proposed heuristic

Future works

- Better two-way interaction estimates
- Embed eigenspace information into the decomposition construction
- Make better use of the rank variance

Thanks to: U.S. DEPARTMENT OF ENERGY DARPA