Numerical Tensor Calculus

Wolfgang HackbuschWolfgang Hackbusch

Max-Planck-Institut für *Mathematik in den Naturwissenschaften*
and University of Kiel

Inselstr. 22-26, D-04103 Leipzig, Germany
wh@mis.mpg.de

http://www.mis.mpg.de/scicomp/hackbusch_e.html

Torino, September 12+14, 2018

Overview

Introduction:

— Tensors / Tensor operations / Problems in practice

Tensor Representations

— r-Term Format / Tensor Subspace Format / HOSVD

Hierarchical Format

Solution of Linear Systems

Variational Approach

— ALS Method / Study of Examples

Multivariate Cross Approximation

PDEs with stochastic coefficients

Minimal Subspaces

Properties of the HOSVD Projection

Graph-Based Formats

— TT Format / Cyclic Matrix Product Format

Tensorisation

1 Introduction: Tensors

1.1 Where do large-scale tensors appear?

The tensor space $\mathbf{V} = V_1 \otimes V_2 \otimes \ldots \otimes V_d$ with vector spaces V_j $(1 \leq j \leq d)$ is defined as (closure of)

$$\text{span}\{v^{(1)} \otimes v^{(2)} \otimes \ldots \otimes v^{(d)} : v^{(j)} \in V_j\}.$$

Finite dimensional case:

$V_j = \mathbb{R}^{n_j} = \mathbb{R}^{I_j}$ with $I_j = \{1, \ldots, n_j\}$.

Set $\mathbf{I} := I_1 \times I_2 \times \ldots \times I_d$, then $\mathbf{V} \simeq \mathbb{R}^{\mathbf{I}}$, i.e., $\mathbf{v} = (\mathbf{v_i})_{\mathbf{i} \in \mathbf{I}}$.

Tensor product: $\mathbf{v} = v^{(1)} \otimes v^{(2)} \otimes \ldots \otimes v^{(d)} \in \mathbb{R}^{\mathbf{I}}$ with $v^{(j)} \in \mathbb{R}^{I_j}$ defined as

$$\mathbf{v_i} = \mathbf{v}_{i_1,\ldots,i_d} = \mathbf{v}[i_1, \ldots, i_d] = v^{(1)}_{i_1} \cdot v^{(2)}_{i_2} \cdot \ldots \cdot v^{(d)}_{i_d} \qquad \text{for } \mathbf{i} = (i_1, \ldots, i_d) \in \mathbf{I}.$$

1.1.1 Functions

Multivariate functions f defined on a Cartesian product

$$\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_d$$

are tensors.

For instance,

$$L^2(\Omega) = L^2(\Omega_1) \otimes L^2(\Omega_2) \otimes \ldots \otimes L^2(\Omega_d).$$

Tensor product of univariate functions:

$$\left(\bigotimes_{j=1}^{d} f_j\right)(x_1, x_2, \ldots, x_d) := \prod_{j=1}^{d} f_j(x_j).$$

The multivariate function may be the solution of a partial differential equation.

The numerical treatment replaces functions by finite-dimensional analogues ($\rightarrow$ grid functions, finite-element functions).

1.1.2 Grid Functions

Discretisation in *product grids* $\omega = \omega_1 \times \omega_2 \times \ldots \times \omega_d$,
e.g., ω_j regular grid with n_j grid points.

Total number of grid points $N = \prod_{j=1}^{d} n_j$, e.g., n^d. Tensor space:

$$\mathbb{R}^N \simeq \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2} \otimes \ldots \otimes \mathbb{R}^{n_d}.$$

Small discretisation errors require large dimensions n_j.

Challenge: Huge dimensions like in ...

1) $n = 1\,000\,000$ and $d = 3$

2) $n = 1000$ and $d = 1000 \quad \Rightarrow$

$$N = 1000^{1000} = 10^{3000}.$$

1.1.3 Matrices or Operators

Let $\mathbf{V} = V_1 \otimes V_2 \otimes \ldots \otimes V_d, \quad \mathbf{W} = W_1 \otimes W_2 \otimes \ldots \otimes W_d$ be tensor spaces,

$A_j : V_j \to W_j$ linear mappings $(1 \le j \le d)$.

The tensor product (*Kronecker product*)

$$\mathbf{A} = A_1 \otimes A_2 \otimes \ldots \otimes A_d : \mathbf{V} \to \mathbf{W}$$

is the mapping

$$\mathbf{A} : v^{(1)} \otimes v^{(2)} \otimes \ldots \otimes v^{(d)} \mapsto A_1 v^{(1)} \otimes A_2 v^{(2)} \otimes \ldots \otimes A_d v^{(d)}.$$

If $A_j \in \mathbb{R}^{n \times n}$ then $\mathbf{A} \in \mathbb{R}^{n^d \times n^d}$.

Example: Poisson problem $-\Delta u = f$ in $[0,1]^d$, $u = 0$ on Γ.

The differential operator has the form

$$L = \frac{\partial^2}{\partial x_1^2} \otimes I \otimes \ldots \otimes I + \ldots + I \otimes \ldots \otimes I \otimes \frac{\partial^2}{\partial x_d^2}.$$

Discretise by difference scheme with n grid points per direction.
The system matrix is

$$\mathbf{A} = T_1 \otimes I \otimes \ldots \otimes I + \ldots + I \otimes \ldots \otimes I \otimes T_d.$$

Challenge: Approximate the inverse of $\mathbf{A} \in \mathbb{R}^{N \times N}$,
where $n = d = 1000$, so that

$$N = n^d = 1000^{1000} = 10^{3000}.$$

Later result: required storage: $O(dn \log^2 \frac{1}{\varepsilon})$

1.2 Tensor Operations

addition: $\mathbf{v} + \mathbf{w}$,

scalar product: $\langle \mathbf{v}, \mathbf{w} \rangle$

matrix-vector multiplication: $\left(\bigotimes_{j=1}^{d} A^{(j)} \right) \left(\bigotimes_{j=1}^{d} v^{(j)} \right) = \bigotimes_{j=1}^{d} A^{(j)} v^{(j)}$,

Hadamard product: $(\mathbf{v} \odot \mathbf{w}) [\mathbf{i}] = \mathbf{v}[\mathbf{i}]\mathbf{w}[\mathbf{i}]$, pointwise product of functions

$$\left(\bigotimes_{j=1}^{d} v^{(j)} \right) \odot \left(\bigotimes_{j=1}^{d} w^{(j)} \right) = \bigotimes_{j=1}^{d} v^{(j)} \odot w^{(j)},$$

convolution: $\mathbf{v}, \mathbf{w} \in \bigotimes_{j=1}^{d} \mathbb{R}^{n} : \mathbf{u} = \mathbf{v} \star \mathbf{w}$ with $\mathbf{u_i} = \sum_{\mathbf{0} \leq \mathbf{k} \leq \mathbf{i}} \mathbf{v_{i-k}} \mathbf{w_k}$

$$\left(\bigotimes_{j=1}^{d} v^{(j)} \right) \star \left(\bigotimes_{j=1}^{d} w^{(j)} \right) = \bigotimes_{j=1}^{d} v^{(j)} \star w^{(j)}.$$

1.3 High-Dimensional Problems in Practice

1) boundary value problems $Lu = f$ in cubes or $\mathbb{R}^3 \Rightarrow d = 3$, n_j large

2) Hartree-Fock equations (as 1))

3) Schrödinger equation ($d = 3\times$ number of electrons + antisymmetry)

4) bvp $L(p)u = f$ with parameters $p = (p_1, \ldots, p_m) \Rightarrow d = m + 1$

5) bvp with stochastic coefficients $\Rightarrow$ as 4) with $m = \infty$

6) coding of a d-variate function in Cartesian product $\Rightarrow d = d$

7) ...

8) Lyapunov equation $(A \otimes I + I \otimes A)\,\mathbf{x} = \mathbf{b}$

2 Tensor Representations

How to represent tensors with n^d entries by few data?

Classical formats:

- r-Term Format (Canonical Format)

- Tensor Subspace Format (Tucker Format)

More recent:

- Hierarchical Tensor Format

2.1 r-Term Format (Canonical Format)

By definition, any algebraic tensor $\mathbf{v} \in \mathbf{V} = V_1 \otimes V_2 \otimes \ldots \otimes V_d$ has a representation

$$\mathbf{v} = \sum_{\rho=1}^{r} v_\rho^{(1)} \otimes v_\rho^{(2)} \otimes \ldots \otimes v_\rho^{(d)} \qquad \text{with } v_\rho^{(j)} \in V_j$$

and suitable r. Set

$$\mathcal{R}_r := \left\{ \sum_{\rho=1}^{r} v_\rho^{(1)} \otimes v_\rho^{(2)} \otimes \ldots \otimes v_\rho^{(d)} : v_\rho^{(j)} \in V_j \right\}$$

Storage: rdn (for $n = \max \dim V_j$).
If r is of moderate size, this format is advantageous.
Often, a tensor $\mathbf{v}$ is replaced by an approximation $\mathbf{v}_\varepsilon \in \mathcal{R}_r$ with $r = r(\varepsilon)$.

$$\text{rank}(\mathbf{v}) := \min\{r : \mathbf{v} \in \mathcal{R}_r\}, \qquad \mathcal{R}_r := \{\mathbf{v} \in \mathbf{V} : \text{rank}(\mathbf{v}) \leq r\}.$$

Recall the matrix $\mathbf{A}$ discretising the Laplace equation:

$$\mathbf{A} = T_1 \otimes I \otimes \ldots \otimes I + \ldots + I \otimes \ldots \otimes I \otimes T_d.$$

REMARK: $\mathbf{A} \in \mathcal{R}_d$ and rank($\mathbf{A}$) $= d$ (tensor rank, not matrix rank).

T_j: tridiagonal matrices of size $n \times n$.

Size of $\mathbf{A}$: $N \times N$ with $N = n^d$.
E.g., $n = d = 1000 \implies N = n^d = 1000^{1000} = 10^{3000}$.

We aim at the inverse of $\mathbf{A} \in \mathbb{R}^{N \times N}$.

Solution: $\mathbf{A}^{-1} \approx \mathbf{B}_r$ with $\mathbf{B}_r$ of the form

$$\mathbf{B}_r = \sum_{i=1}^{r} a_i \bigotimes_{j=1}^{d} \exp(-b_i T_j),$$

where $a_i, b_i > 0$ are explicitly known.

Proof. Approximate $1/x$ in $[1, \infty)$ by exponential sums $E_r(x) = \sum_{i=1}^{r} a_i \exp(-b_i x)$. The best approximation satisfies

$$\left\| \frac{1}{\bullet} - E_r(\cdot) \right\|_{\infty,[1,\infty)} \leq O(\exp(-cr^{1/2})).$$

For a positive definite matrix with $\sigma(\mathbf{A}) \subset [1, \infty)$, $E_r(\mathbf{A})$ approximates $\mathbf{A}^{-1}$ with

$$\left\| E_r(\mathbf{A}) - \mathbf{A}^{-1} \right\|_2 \leq O(\exp(-cr^{1/2})).$$

In the case of $\mathbf{A} = T_1 \otimes I \otimes \ldots \otimes I + \ldots + I \otimes \ldots \otimes I \otimes T_d$ one obtains

$$\mathbf{B}_r := E_r(\mathbf{A}) = \sum_{i=1}^{r} a_i \bigotimes_{j=1}^{d} \exp(-b_i T_j) \in \mathcal{R}_r.$$

Representation versus Decomposition

$P := (\times_{j=1}^{d} V_j)^r$ parameter set.

Representation of a tensor:

$$\varphi : P \longrightarrow \mathcal{R}_r \subset \mathbf{V}.$$

Injectivity of φ not required, rank$(\varphi(p)) \leq r$.

Let rank$(\mathbf{v}) = r$. Under certain conditions the representation of $\mathbf{v} = \varphi(p)$ is *essentially* unique. This allows the **decomposition**

$$\varphi^{-1} : \mathcal{R}_r \longrightarrow P.$$

Operations with Tensors and Truncations

$$\mathbf{A} = \sum_{\nu=1}^{r} \bigotimes_{j=1}^{d} A_\nu^{(j)} \in \mathcal{R}_r, \qquad \mathbf{v} = \sum_{\nu=1}^{s} \bigotimes_{j=1}^{d} v_\nu^{(j)} \in \mathcal{R}_s$$

$\Rightarrow$

$$\mathbf{w} := \mathbf{Av} = \sum_{\nu=1}^{r} \sum_{\mu=1}^{s} \bigotimes_{j=1}^{d} A_\nu^{(j)} v_\mu^{(j)} \in \mathcal{R}_{rs}$$

Because of the increased representation rank rs, one must apply a **truncation** $\mathbf{w} \mapsto \mathbf{w}' \in \mathcal{R}_{r'}$ with $r' < rs$.

Unfortunately, truncation to lower rank is not straightforward in the r-term format.

There are also other disadvantages of the r-term format ....

Numerical Difficulties because of Non-Closedness

In general, $\mathcal{R}_r$ is *not closed*. Example: a, b linearly independent and

$$\begin{aligned} \mathbf{v} &= a \otimes a \otimes b + a \otimes b \otimes a + b \otimes a \otimes a \in \mathcal{R}_3 \backslash \mathcal{R}_2 \\ \mathbf{v} &= \underbrace{(b + na) \otimes \left(a + \tfrac{1}{n} b\right) \otimes a + a \otimes a \otimes (b - na)}_{\mathbf{v}_n \in \mathcal{R}_2} - \tfrac{1}{n} b \otimes b \otimes a. \end{aligned}$$

Here, the terms of $\mathbf{v}_n$ grow like $O(n)$, while the result is of size $O(1)$.
This implies *numerical cancellation*: $\log_2 n$ binary digits of $\mathbf{v}_n$ are lost.
We say that the sequence $\{\mathbf{v}_n\}$ is unstable.

Proposition: Suppose $\dim(V_j) < \infty$ and $\mathbf{v} \in \mathbf{V} = \bigotimes_{j=1}^{d} V_j$.
A stable sequence $\mathbf{v}_n \in \mathcal{R}_r$ with $\lim \mathbf{v}_n = \mathbf{v}$ exists if and only if $\mathbf{v} \in \mathcal{R}_r$.

Conclusion: If $\mathbf{v} = \lim \mathbf{v}_n \notin \mathcal{R}_r$, the sequence $\mathbf{v}_n \in \mathcal{R}_r$ is unstable.

Best approximation problem: Let $\mathbf{v}^* \in \mathbf{V}$. Try to find $\mathbf{v} \in \mathcal{R}_r$ with

$$\|\mathbf{v}^* - \mathbf{v}\| = \inf\{\|\mathbf{v}^* - \mathbf{w}\| : \mathbf{w} \in \mathcal{R}_r\}.$$

This optimisation problem need **not be solvable.**

De Silva–Lim (2008): Tensors without a best approximation have a positive measure ($\mathbb{K} = \mathbb{R}$).

2.2 Tensor Subspace Format (Tucker Format)

2.2.1 Definition of $\mathcal{T}_{\mathbf{r}}$

Implementational description: $\mathcal{T}_{\mathbf{r}}$ with $\mathbf{r} = (r_1, \ldots, r_d)$ contains all tensors of the form

$$\mathbf{v} = \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} \mathbf{a}[i_1, \ldots, i_d] \bigotimes_{j=1}^{d} b_{i_j}^{(j)}$$

with some vectors $\{b_{i_j}^{(j)} : 1 \leq i_j \leq r_j\} \subset V_j$ possibly with $r_j \ll n_j$ and $\mathbf{a}[i_1, \ldots, i_d] \in \mathbb{R}$.
The core tensor $\mathbf{a}$ has $\prod_{j=1}^{d} r_j$ entries.

Algebraic description:

Tensor space $\mathbf{V} = V_1 \otimes V_2 \otimes \ldots \otimes V_d$. Choose subspaces $U_j \subset V_j$ and consider the tensor subspace $\mathbf{U} = \bigotimes_{j=1}^{d} U_j$. Then

$$\mathcal{T}_{\mathbf{r}} := \bigcup_{\dim(U_j) \leq r_j} \bigotimes_{j=1}^{d} U_j.$$

Short Notation

$$\mathbf{v} = \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} \mathbf{a}[i_1, \ldots, i_d] \bigotimes_{j=1}^{d} b_{i_j}^{(j)}.$$

Define matrices $B^{(j)} := [b_1^{(j)} \cdots b_{r_j}^{(j)}]$ and $\mathbf{B} := \bigotimes_{j=1}^{d} B^{(j)}$.

Then

$$\mathbf{v} = \mathbf{Ba}$$

2.2.2 Matricisation and Tucker Ranks

Let $\mathbf{V} = \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2} \otimes \ldots \otimes \mathbb{R}^{n_d}$, fix $j \in \{1, \ldots, d\}$, set $n_{[j]} := \prod_{k \neq j} n_k$.

The j-th *matricisation* maps a tensor $\mathbf{v} \in \mathbf{V}$ into a matrix

$$M_j \in \mathbb{R}^{n_j \times n_{[j]}}$$

defined by

$$M_j[i_j, \mathbf{i}_{[j]}] := \mathbf{v}[i_1, \ldots, i_d] \qquad \text{for } \mathbf{i}_{[j]} := (i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_d).$$

The isomorphism $\mathcal{M}_j : \mathbf{V} \to \mathbb{R}^{n_j \times n_{[j]}}$ is called the j-th *matricisation*.

Tucker rank or j-th rank:

$$r_j = \mathrm{rank}_j(\mathbf{v}) := \mathrm{rank}(\mathcal{M}_j(\mathbf{v})) \qquad \text{for } 1 \leq j \leq d.$$

Sometimes, $\mathbf{r} := (r_1, \ldots, r_d)$ is called the *multilinear rank* of $\mathbf{v}$.

Example: $\mathbf{v} \in \mathbf{V} := \mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2$. Then $\mathcal{M}_2(\mathbf{v})$ belongs to $\mathbb{R}^{2\times 8}$:

$$\mathcal{M}_2(\mathbf{v}) = \begin{pmatrix} \mathbf{v}_{1111} & \mathbf{v}_{1112} & \mathbf{v}_{1121} & \mathbf{v}_{1122} & \mathbf{v}_{2111} & \mathbf{v}_{2112} & \mathbf{v}_{2121} & \mathbf{v}_{2122} \\ \mathbf{v}_{1211} & \mathbf{v}_{1212} & \mathbf{v}_{1221} & \mathbf{v}_{1222} & \mathbf{v}_{2211} & \mathbf{v}_{2212} & \mathbf{v}_{2221} & \mathbf{v}_{2222} \end{pmatrix}.$$

2.2.3 Important Properties

Alternative definition of $\mathcal{T}_{\mathbf{r}}$:

$$\mathcal{T}_{\mathbf{r}} = \left\{ \mathbf{v} \in \mathbf{V} : \text{rank}_j(\mathbf{v}) \leq r_j \text{ for all } 1 \leq j \leq d \right\}.$$

Later we shall prove:

- Also for $\dim V_j = \infty$, $\text{rank}_j(\mathbf{v})$ can be defined.
- $\mathcal{T}_{\mathbf{r}}$ is weakly closed.
- If $\mathbf{V}$ is a reflexive Banach space, the best approximation problem

  $$\inf_{\mathbf{u} \in \mathcal{T}_{\mathbf{r}}} \|\mathbf{v} - \mathbf{u}\| = \|\mathbf{v} - \mathbf{u}_{\text{best}}\|$$

  has a solution $\mathbf{u}_{\text{best}} \in \mathcal{T}_{\mathbf{r}}$.

Choice of Vectors $b_i^{(j)}$

Let $\mathbf{v} \in \bigotimes_{j=1}^{d} U_j$. Representation of U_j by

1) generating system $\{b_i^{(j)}\}$ with $U_j = \mathrm{span}_i\, b_i^{(j)}$,

2) basis $\{b_i^{(j)}\}_{i=1}^{r_j}$ $\quad (r_j = \dim U_j)$

3) orthonormal basis (good numerical properties!)

4) special orthonormal basis: HOSVD basis

2.2.4 HOSVD: Higher Order Singular Value Decomposition

Diagonalisation:

$$\mathbb{R}^{n_j \times n_j} \ni \mathcal{M}_j(\mathbf{v})\mathcal{M}_j(\mathbf{v})^{\mathsf{T}} = \sum_{i=1}^{\mathrm{rank}_j(\mathbf{v})} (\sigma_i^{(j)})^2\, b_i^{(j)} (b_i^{(j)})^{\mathsf{H}}.$$

$\sigma_i^{(j)}$: j-th singular values; $\{b_i^{(j)} : 1 \le i \le \mathrm{rank}_j(\mathbf{v})\}$: *HOSVD basis*(orthonormal!).

Truncation: Let $\mathbf{v} = \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} \mathbf{a}[i_1, \ldots, i_d] \bigotimes_{j=1}^{d} b_{i_j}^{(j)} \in \mathcal{T}_{\mathbf{r}}$ with HOSVD basis vectors $b_i^{(j)}$. For $\mathbf{s} = (s_1, \ldots, s_d) \le \mathbf{r}$ set

$$\mathbf{u}_{\mathsf{HOSVD}} = \sum_{i_1=1}^{s_1} \cdots \sum_{i_d=1}^{s_d} \mathbf{a}[i_1, \ldots, i_d] \bigotimes_{j=1}^{d} b_{i_j}^{(j)} \in \mathcal{T}_{\mathbf{s}}.$$

Quasi-optimality:

$$\|\mathbf{v} - \mathbf{u}_{\mathsf{HOSVD}}\| \le \left(\sum_{j=1}^{d} \sum_{i=s_j+1}^{r_j} \left(\sigma_i^{(j)}\right)^2 \right)^{1/2} \le d^{1/2} \|\mathbf{v} - \mathbf{u}_{\mathsf{best}}\| \quad (\mathbf{u}_{\mathsf{best}} \in \mathcal{T}_{\mathbf{s}}).$$

Conclusion concerning the traditional formats:

1. r-term format $\mathcal{R}_r$
   - advantage: low storage cost rdn
   - disadvantage: difficult truncation, numerical instability may occur

2. tensor subspace format $\mathcal{T}_{\mathbf{r}}$
   - advantage: stable and quasi-optimal truncation
   - disadvantage: exponentially expensive storage for core tensor $\mathbf{a}$

The next format combines the advantages.

3 Hierarchical Format

3.1 Dimension Partition Tree

Example: $\mathbf{v} \in \mathbf{V} = V_1 \otimes V_2 \otimes V_3 \otimes V_4$. There are subspaces such that

$$\mathbf{v} \in \mathrm{span}\{\mathbf{v}\} \subset \mathbf{U}_{\{1,2\}} \otimes \mathbf{U}_{\{3,4\}} \subset \mathbf{V}$$

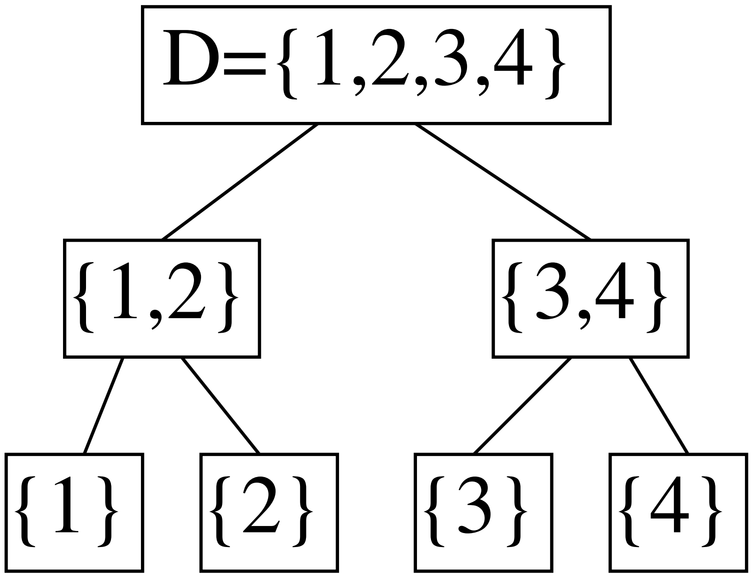
$$\mathbf{U}_{\{1,2\}} \subset U_1 \otimes U_2 \qquad\qquad \mathbf{U}_{\{3,4\}} \subset U_3 \otimes U_4$$

$$U_1 \subset V_1 \qquad U_2 \subset V_2 \qquad U_3 \subset V_3 \qquad U_4 \subset V_4$$

Optimal subspaces are $\mathbf{U}_\alpha := U_\alpha^{\mathrm{min}}(\mathbf{v})$.

Dimension partition tree:

Any binary tree with root $D := \{1, \ldots, d\}$ and leaves $\{1\}, \{2\}, ..., \{d\}$.



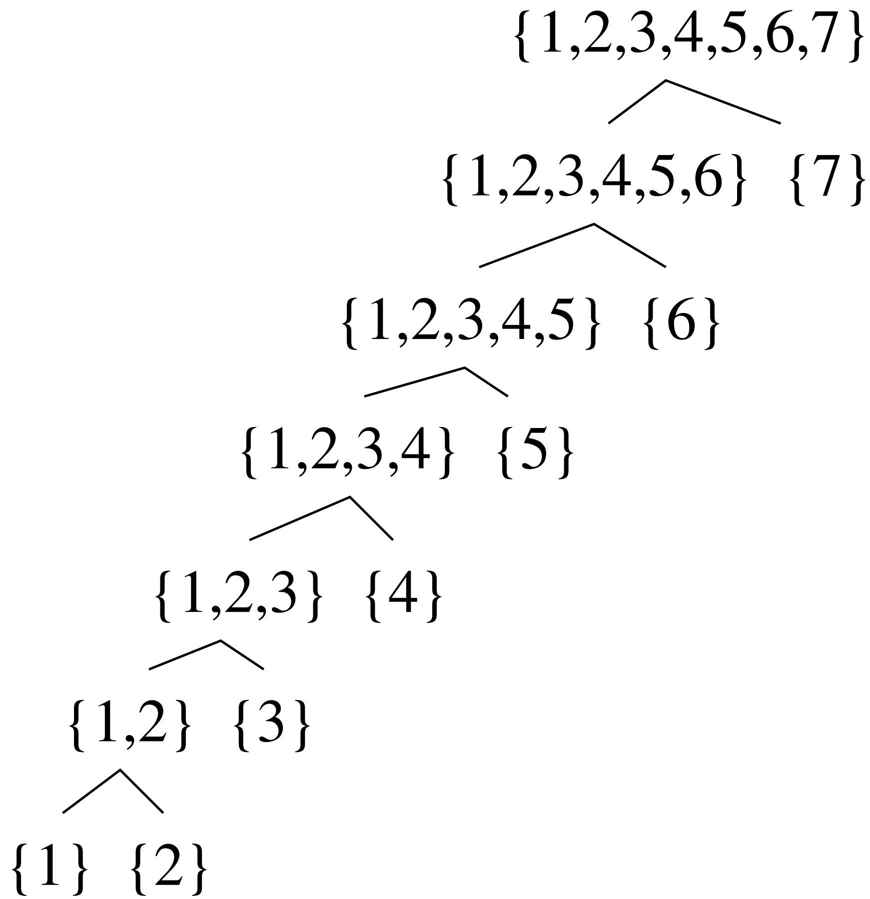


Figure 1: Balanced tree and linear tree

The hierarchical format based on the linear tree is also called the TT format.

3.2 Algorithmic Realisation

Typical situation: $\mathbf{U}_{\{1,2\}} \subset U_1 \otimes U_2$ (nestedness property).

Bases: $U_1 = \underset{1\le i\le r_1}{\operatorname{span}} \{b_i^{(1)}\}$, $U_2 = \underset{1\le j\le r_2}{\operatorname{span}} \{b_j^{(2)}\}$, $\mathbf{U}_{\{1,2\}} = \underset{1\le \ell\le r_{\{1,2\}}}{\operatorname{span}} \{\mathbf{b}_\ell^{(\{1,2\})}\}$.

$$\mathbf{b}_\ell^{(\{1,2\})} = \sum_{i=1}^{r_{\{1,2\}}} c_{ij}^{(\{1,2\},\ell)}\, b_i^{(1)} \otimes b_j^{(2)}$$

Only the basis vectors $b_\nu^{(j)}$ of $U_j \subset V_j$ $(1 \le j \le d)$ are explicitly stored,
for the other nodes store the coefficient matrices

$$C^{(\alpha,\ell)} = \left(c_{ij}^{(\alpha,\ell)}\right)_{ij} \in \mathbb{R}^{r_{\alpha_1}\times r_{\alpha_2}}.$$

The tensor is represented by $\mathbf{v} = c_1 \mathbf{b}_1^{(\{1,\ldots,d\})}$.

Storage: $(d-1)r^3 + drn$ for $\left[C^{(\alpha,\ell)}, c_1, b_\nu^{(j)}\right]$ $(r := \max_\alpha \dim U_\alpha;\ n := \max_j \dim(V_j))$

3.3 Operations - Example: scalar product

Let $\mathbf{v}, \mathbf{w} \in \mathbf{V}$ be given by the data $\left(C'^{(\alpha,\ell)}, c'_1, b'^{(j)}_\nu\right)$ and $\left(C''^{(\alpha,\ell)}, c''_1, b''^{(j)}_\nu\right)$ resp.

$$\mathbf{v} = c'_1 \mathbf{b}'^{(D)}_1, \;\; \mathbf{w} = c''_1 \mathbf{b}''^{(D)}_1 \quad \Rightarrow \quad \langle \mathbf{v}, \mathbf{w} \rangle = c'_1 c''_1 \left\langle \mathbf{b}'^{(D)}_1, \mathbf{b}''^{(D)}_1 \right\rangle .$$

Determine the scalar products $\beta^{(\alpha)}_{ij} := \left\langle \mathbf{b}'^{(\alpha)}_i, \mathbf{b}''^{(\alpha)}_j \right\rangle$ recursively by

$$\begin{aligned}
\beta^{(\alpha)}_{ij} &= \left\langle \mathbf{b}'^{(\alpha)}_i, \mathbf{b}''^{(\alpha)}_j \right\rangle = \left\langle \sum_{k,\ell} c'^{(\alpha,i)}_{k,\ell} b'^{(\alpha_1)}_k \otimes b'^{(\alpha_2)}_\ell, \sum_{p,q} c''^{(\alpha,j)}_{p,q} b''^{(\alpha_1)}_p \otimes b''^{(\alpha_2)}_q \right\rangle \\
&= \sum_{k,\ell} \sum_{p,q} c'^{(\alpha,i)}_{k,\ell} c''^{(\alpha,j)}_{p,q} \left\langle b'^{(\alpha_1)}_k, b''^{(\alpha_1)}_p \right\rangle \left\langle b'^{(\alpha_2)}_\ell, b''^{(\alpha_2)}_q \right\rangle \\
&= \sum_{k,\ell} \sum_{p,q} c'^{(\alpha,i)}_{k,\ell} c''^{(\alpha,j)}_{p,q} \beta^{(\alpha_1)}_{kp} \beta^{(\alpha_2)}_{\ell q}
\end{aligned}$$

(α_1, α_2: sons of α; $\beta^{(\alpha)}_{kp}$ explicitly computable for leaves $\alpha = \{j\}$).

3.4 Basis Transformation

Set $\mathbf{B}_\alpha = [\mathrm{b}_1^{(\alpha)} \cdots \mathrm{b}_{r_\alpha}^{(\alpha)}]$. Let $\mathbf{B}'_\alpha = [\mathrm{b}_1'^{(\alpha)} \cdots \mathrm{b}_{r_\alpha}'^{(\alpha)}]$ be another basis.

The sons of α are denoted by α_1 and α_2.

The relation

$$\mathbf{B}'_{\alpha_i} T^{(\alpha_i)} = \mathbf{B}_{\alpha_i} \qquad (i = 1, 2)$$

corresponds to

$$C'^{(\alpha,\ell)} = T^{(\alpha_1)} C^{(\alpha,\ell)} (T^{(\alpha_2)})^{\mathsf{T}} \qquad \text{for } 1 \le \ell \le r_\alpha.$$

Two directions:

1) Given $\mathbf{B}_{\alpha_i}$, the new bases $\mathbf{B}'_{\alpha_i} := \mathbf{B}_{\alpha_i}(T^{(\alpha_i)})^{-1}$ lead to new coefficient matrices $C'^{(\alpha,\ell)} := T^{(\alpha_1)} C^{(\alpha,\ell)} (T^{(\alpha_2)})^{\mathsf{T}}$.

2) Given $\mathbf{B}'_{\alpha_i}$ and a decomposition $C'^{(\alpha,\ell)} = T^{(\alpha_1)} \cdot C^{(\alpha,\ell)} \cdot (T^{(\alpha_2)})^{\mathsf{T}}$, $C^{(\alpha,\ell)}$ corresponds to $\mathbf{B}_{\alpha_i} := \mathbf{B}'_{\alpha_i} T^{(\alpha_i)}$.

3.5 Orthonormalisation

REMARK Let α be a vertex with sons α_1 and α_2.
The basis $\{\mathbf{b}_\ell^{(\alpha)}\}$ is orthonormal, if
(a) the bases $\{\mathbf{b}_i^{(\alpha_1)}\}$ and $\{\mathbf{b}_j^{(\alpha_2)}\}$ of the sons are orthonormal and
(b) the matrices $C^{(\alpha,\ell)}$ are orthonormal with respect to the Frobenius scalar product:

$$\left\langle C^{(\alpha,\ell)}, C^{(\alpha,m)} \right\rangle_{\mathsf{F}} = \sum\nolimits_{ij} \left\langle c_{ij}^{(\alpha,\ell)}, c_{ij}^{(\alpha,m)} \right\rangle = \delta_{\ell m}.$$

Algorithm:
(a) Orthonormalise the explicitly given bases at the leaves (e.g., by QR).
(b) As soon as $\{\mathbf{b}_i^{(\alpha_1)}\}$ and $\{\mathbf{b}_j^{(\alpha_2)}\}$ are orthonormal, orthonormalise the matrices $\{C^{(\alpha,\ell)}\}$.
The new matrices $C_{\mathsf{new}}^{(\alpha,\ell)}$ define the new orthonormal basis $\{\mathbf{b}_{\ell,\mathsf{new}}^{(\alpha)}\}$.

3.6 HOSVD and HOSVD Bases

We recall: The HOSVD basis $\{\mathbf{b}_\ell^{(\alpha)}\}$ consists of the normalised eigenvectors of $M_\alpha M_\alpha^\mathsf{T}$, where $M_\alpha := \mathcal{M}_\alpha(\mathbf{v})$ is the α-matricisation of the tensor $\mathbf{v}$.
Instead of $\{\mathbf{b}_\ell^{(\alpha)}\}$ we need the corresponding coefficient matrices $\{C_{\mathsf{HOSVD}}^{(\alpha,\ell)}\}$.

Step 1: Orthonormalisation of the bases.
Step 2: Recursion from the root to the leaves:
2a) Start at the root: $\sigma_1^{(root)} := |c_1^{(root)}|$ where $\mathbf{v} = c_1^{(root)}\mathbf{b}_1^{(root)}$.
2b) Set

$$E_{\alpha_1} := \sum_{i=1}^{r_\alpha} (\sigma_i^{(\alpha)})^2 C^{(\alpha,i)}(C^{(\alpha,i)})^\mathsf{H}, \quad E_{\alpha_2} := \sum_{i=1}^{r_\alpha} (\sigma_i^{(\alpha)})^2 ((C^{(\alpha,i)})^\mathsf{H} C^{(\alpha,i)})^\mathsf{T}.$$

Diagonalisation yields

$$E_{\alpha_1} = U\Sigma_{\alpha_1}^2 U^\mathsf{H}, \qquad E_{\alpha_2} = V\Sigma_{\alpha_2}^2 V^\mathsf{H} \qquad \text{with} \quad \Sigma_{\alpha_i} = \mathrm{diag}\{\sigma_\nu^{(\alpha_i)}\}.$$

$\mathbf{B}_{\alpha_1}^{\mathsf{HOSVD}} := \mathbf{B}_{\alpha_1}U$ and $\mathbf{B}_{\alpha_2}^{\mathsf{HOSVD}} = \mathbf{B}_{\alpha_2}V$ are the desired HOSVD bases.

Arithmetical cost: $O(dr^4 + dnr^2)$.

3.7 HOSVD Truncation

Represent the tensor $\mathbf{v}$ with respect to the HOSVD bases $\left\{b_\ell^{(\alpha)} : 1 \leq \ell \leq r_\alpha\right\}$.

Choose smaller dimensions

$$s_\alpha \leq r_\alpha.$$

Omit all terms corresponding to $\left\{b_\ell^{(\alpha)} : s_\alpha < \ell \leq r_\alpha\right\}$. Result: $\mathbf{v}_{\mathsf{HOSVD}}$.

Then the following estimates hold:

$$\|\mathbf{v} - \mathbf{v}_{\mathsf{HOSVD}}\| \leq \left(\sum_\alpha \sum_{\nu \geq s_\alpha + 1} (\sigma_\nu^{(\alpha)})^2\right)^{1/2} \leq (2d-3)^{1/2} \|\mathbf{v} - \mathbf{v}_{\mathsf{best}}\|.$$

4 Solution of Linear Systems

Linear system

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

where $\mathbf{x}, \mathbf{b} \in \mathbf{V} = \bigotimes_{j=1}^{d} V_j$ and $\mathbf{A} \in \bigotimes_{j=1}^{d} \mathcal{L}(V_j, V_j) \subset \mathcal{L}(\mathbf{V}, \mathbf{V})$ are represented in one of the formats (e.g., $\mathbf{A}$: r-term format, $\mathbf{x}, \mathbf{b}$: hierarchical format):

Standard linear iteration:

$$\mathbf{x}^{m+1} = \mathbf{x}^{m} - \mathbf{B}\left(\mathbf{A}\mathbf{x}^{m} - \mathbf{b}\right).$$

$\Rightarrow$ representation ranks blow up.

Therefore truncations T are used ('truncated iteration'):

$$\mathbf{x}^{m+1} = T\left(\mathbf{x}^{m} - \mathbf{B}\left(T\left(\mathbf{A}\mathbf{x}^{m} - \mathbf{b}\right)\right)\right).$$

Cost per step: $nd\times$ powers of the involved representation ranks.

$$\mathbf{x}^{m+1} = T\left(\mathbf{x}^m - \mathbf{B}\left(T\left(\mathbf{A}\mathbf{x}^m - \mathbf{b}\right)\right)\right)$$

Choice of $\mathbf{B}$:

If $\mathbf{A}$ corresponds to an elliptic pde of order 2, the discretisation of Δ is spectrally equivalent $\Rightarrow \mathbf{B} = \mathbf{B}_r$ from above has a simple r-term format.

Obvious variants: cg-like methods

Literature:

Khoromskij 2009, Kressner-Tobler 2010, Kressner-Tobler 2011 (SIAM), Kressner-Tobler 2011 (CMAM), Osedelets-Tyrtyshnikov-Zamarashkin 2011, Ballani-Grasedyck 2013, Savas-Eldén 2013

Remark: For $d = 2$, these linear systems may be written as matrix equations:

$$(A \otimes I + I \otimes A)\,\mathbf{x} = \mathbf{b} \quad \Leftrightarrow \quad AX + XA = B \qquad \text{(Lyapunov)}$$

(cf. Benner-Breiten 2013).

5 Variational Approach

Define

$$\Phi(\mathbf{x}) := \langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle - 2 \langle \mathbf{b}, \mathbf{x} \rangle$$

if $\mathbf{A}$ is positive definite or

$$\Phi(\mathbf{x}) := \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$$

or

$$\Phi(\mathbf{x}) := \|\mathbf{B}(\mathbf{A}\mathbf{x} - \mathbf{b})\|^2$$

and try to minimise $\Phi(\mathbf{x})$ over all parameters of $\mathbf{x}$ is a fixed format.

Literature:

Espig-Hackbusch-Rohwedder-Schneider, Falcó-Nouy,
Holtz-Rohwedder-Schneider, Mohlenkamp, Osedelets,...

5.1 Formulation of the Problem, ALS Method

Let

$$\Phi(\mathbf{u}) = \min$$

be a minimisation problem over the whole tensor space $\mathbf{u} \in \mathbf{V}$.

Approximation: Choose any format $\mathcal{F} \subset \mathbf{V}$. Solve

$$\Phi(\mathbf{u}) = \min \text{ over all } \mathbf{v} \in \mathcal{F}.$$

This is the minimisation over all parameters in the representation of $\mathbf{v} \in \mathcal{F}$.

Difficulty: While the original problem may be convex, the new problem is not.

Example: $\Phi(\mathbf{u}) = \|\mathbf{v}-\mathbf{u}\|^2$ over all $\mathbf{u} \in \mathcal{R}_1 = \mathcal{T}_{(1,\ldots,1)}$. $\mathbf{v} \in \mathbf{V}$ is arbitrary.

Ansatz:

$$\mathbf{u} = u^{(1)} \otimes u^{(2)} \otimes \ldots \otimes u^{(d)}, \quad u^{(j)} \in V_j = \mathbb{R}^{n_j}$$

Necessary condition: $\nabla\Phi(\mathbf{u}) = 0$ (multilinear system of equations).

ALS = alternating least-squares method:

1) solve $\nabla\Phi(u^{(1)} \otimes u^{(2)} \otimes \ldots \otimes u^{(d)}) = 0$ w.r.t. $u^{(1)} \Rightarrow$ solution: $\hat{u}^{(1)}$,

2) solve $\nabla\Phi(\hat{u}^{(1)} \otimes u^{(2)} \otimes \ldots \otimes u^{(d)}) = 0$ w.r.t. $u^{(2)} \Rightarrow$ solution: $\hat{u}^{(2)}$,

⋮

d) solve $\nabla\Phi(\hat{u}^{(1)} \otimes \ldots \otimes \hat{u}^{(d-1)} \otimes u^{(d)}) = 0$ w.r.t. $u^{(d)} \Rightarrow$ solution: $\hat{u}^{(d)}$

All partial steps are linear problems and easy to solve.

One ALS iteration is given by $\mathbf{u}_0 = u^{(1)} \otimes \ldots \otimes u^{(d)} \mapsto \mathbf{u}_1 = \hat{u}^{(1)} \otimes \ldots \otimes \hat{u}^{(d)}$.
This defines a ALS sequence $\{\mathbf{u}_m : m \in \mathbb{N}_0\}$.

Questions: Does $\mathbf{u}_m$ converge? To what limit? Convergence speed?

5.2 First Results

Mohlenkamp (2013, Linear Algebra Appl. **438**):

- The sequence $\{\mathbf{u}_m : m \in \mathbb{N}_0\}$ is bounded,
- $\|\mathbf{u}_m - \mathbf{u}_{m+1}\| \to 0$,
- $\sum_{m=0}^{\infty} \|\mathbf{u}_m - \mathbf{u}_{m+1}\|^2 < \infty$,
- the set S of accumulation points of $\{\mathbf{u}_m\}$ is connected and compact.

Conclusion: If S contains an isolated point $\mathbf{u}^*$, it follows that $\mathbf{u}_m \to \mathbf{u}^*$.

Note that, in general, the limit may depend on the starting value!

5.3 Study of Examples

5.3.1 Case of $d = 2$

$$\mathbf{v} := \binom{1}{0} \otimes \binom{1}{0} + 2\binom{0}{1} \otimes \binom{0}{1}, \quad \Phi(\mathbf{u}) = \|\mathbf{v} - \mathbf{u}\|^2 .$$

1) $\mathbf{u}^{**} = 2\binom{0}{1} \otimes \binom{0}{1}$ is the global minimiser and an attractive fixed point.
2) $\mathbf{u}^{*} = \binom{1}{0} \otimes \binom{1}{0}$ is a fixed point of the ALS iteration:

$$\Phi(\mathbf{u}^* + \delta_1 \otimes \binom{1}{0}) = \Phi(\mathbf{u}^*) + \|\delta_1\|^2 .$$

But $\Phi\left(\binom{1}{t} \otimes \binom{1}{t}\right) = \Phi(\mathbf{u}^*) - t^2\left(2 - t^2\right)$

$\Rightarrow \mathbf{u}^*$ is a saddle point and a repulsive fixed point.

Conclusion: Almost all starting values lead to $\mathbf{u}_m \to \mathbf{u}^{**}$.

5.3.2 Case of $d \geq 3$

For $a \perp b$ with $\|a\| = \|b\| = 1$ consider $\Phi(\mathbf{u}) = \|\mathbf{v} - \mathbf{u}\|^2$ with

$$\mathbf{v} = \otimes^3 a + 2 \otimes^3 b.$$

Again $\mathbf{u}^* = \otimes^3 a$ and $\mathbf{u}^{**} = 2 \otimes^3 b$ are fixed points, $\Phi(\mathbf{u}^{**}) < \Phi(\mathbf{u}^*)$.
But now both are local minima (attractive fixed points)!
Additional saddle point (repulsive fixed point): $\mathbf{u}^{***} = c \otimes^3 (a + \frac{1}{2} b)$.

The sequence $\{\mathbf{u}_m\}$ corresponding to the starting value

$$\mathbf{u}_0 = c^{(0)} \left(a + t_1^{(0)} b\right) \otimes \left(a + t_2^{(0)} b\right) \otimes \left(a + t_3^{(0)} b\right)$$

is completely defined by $t_2^{(0)}$ and $t_3^{(0)}$. The characteristic value is

$$\tau_m := \left| t_2^{(m)} \right|^\alpha \left| t_3^{(m)} \right|^\beta \quad \text{with} \quad \alpha = 5^{1/2} - 1, \; \beta = 2.$$

(A) $\tau_0 > 2^{-\gamma}$, $\gamma = 5^{1/2} + 1 \Rightarrow \mathbf{u}_m \to \mathbf{u}^{**}$ (global minimiser),
(B) $\tau_0 < 2^{-\gamma} \Rightarrow \mathbf{u}_m \to \mathbf{u}^*$ (local minimiser),
(C) $\tau_0 = 2^{-\gamma} \Rightarrow \mathbf{u}_m \to \mathbf{u}^{***}$ (saddle point, global minimiser on the manifold $\tau = 2^{-\gamma}$).

We recall:

Conclusion: If the set of accumulation points of $\{\mathbf{u}_m\}$ contains an isolated point $\mathbf{u}^*$, it follows that $\mathbf{u}_m \to \mathbf{u}^*$.

Wang–Chu (2014): Global convergence for almost all $\mathbf{u}_0$.

Uschmajew (2015):

Analysis based on the Łojasiewicz inequality yields:
All sequences $\mathbf{u}_m$ converge to some $\mathbf{u}^*$ with $\nabla\Phi(\mathbf{u}^*) = 0$.

Łojasiewicz (1965, Ensembles semi-analytiques): If Φ is analytic,

$$\exists\theta \in (0, 1/2] \quad |\Phi(x) - \Phi(x_*)|^{1-\theta} \le \|\nabla\Phi(x)\|$$

in some neighbourhood of x_*.

Convergence speed?

The proof by the Łojasiewicz inequality is not constructive.

Espig–Khachatryan (2015): Study of sequences for $\Phi(\mathbf{u}) = \|\mathbf{v} - \mathbf{u}\|^2$ with

$$\begin{aligned} \mathbf{v} &= \otimes^3 a + \lambda\,(a \otimes a \otimes b + a \otimes b \otimes a + b \otimes a \otimes a)\,, \\ & a \perp b, \qquad \|a\| = \|b\| = 1. \end{aligned}$$

Depending on the value of λ it is shown that the convergence can be

- sublinear ($\lambda = 1/2$),
- linear ($\lambda < 1/2$).

For $\mathbf{v} = \otimes^3 a + 2 \otimes^3 b$, $\mathbf{u}_m \to \otimes^3 a$ or $2 \otimes^3 b$, we have

- superlinear convergence (of order $2 + 5^{1/2} > 1$)

Study of the general case: Gong–Mohlenkamp–Young 2017

6 Multivariate Cross Approximation

Matrix Case

Problem: given $M \in \mathbb{K}^{I \times J}$, find a rank-$r$ matrix R_r close to M evaluating only $O(r(\#I + \#J))$ entries.

Choose r rows (index subset $\tau := \{i_1, \dots, i_r\} \subset I$) and r columns (index subset $\sigma := \{j_1, \dots, j_r\} \subset J$).

$$M = \begin{bmatrix} & * & & & * & & * & & & \\ & * & & & * & & * & & & \\ * & * & * & * & * & * & * & * & * & * \\ & * & & & * & & * & & & \\ & * & & & * & & * & & & \\ * & * & * & * & * & * & * & * & * & * \\ & * & & & * & & * & & & \\ * & * & * & * & * & * & * & * & * & * \\ & * & & & * & & * & & & \\ & * & & & * & & * & & & \end{bmatrix}$$

Then, a matrix R_r of rank r with

$$R[i,j] = M[i,j] \quad \text{for all index pairs with either } i \in \tau \text{ or } j \in \sigma$$

is given by

$$R_r = M|_{I\times\sigma} \cdot (M|_{\tau\times\sigma})^{-1} \cdot M|_{\tau\times J},$$

provided that the $r \times r$ matrix $M|_{\tau\times\sigma}$ is regular.

$$M = \begin{bmatrix} & * & & & * & & * & & & \\ & * & & & * & & * & & & \\ * & * & * & * & * & * & * & * & * & * \\ & * & & & * & & * & & & \\ & * & & & * & & * & & & \\ * & * & * & * & * & * & * & * & * & * \\ & * & & & * & & * & & & \\ * & * & * & * & * & * & * & * & * & * \\ & * & & & * & & * & & & \\ & * & & & * & & * & & & \end{bmatrix}$$

If $\operatorname{rank}(M) = r$, there exist subsets τ, σ such that $M|_{\tau\times\sigma}$ is regular and $R_r = M$.

Adaptive Cross Approximation (ACA): adaptive choice of τ, σ.

Generalisation to order $d \geq 3$

- hierarchical format

- Apply the previous idea to all matricisations

$$M := \mathcal{M}_\alpha(\mathbf{v}).$$

- M is large, but the matrix $(M|_{\tau\times\sigma})^{-1}$ is still of size $r \times r$.

Then:

Required number of evaluations of the tensor is $O\left(\sum_j \#I_j)\right)$.

If $\mathbf{v}$ has hierarchical rank $\mathfrak{r} := (\text{rank}_\alpha(\mathbf{v}))_{\alpha\in T_D}$, it can be reconstructed in $\mathcal{H}_\mathfrak{r}$ exactly.

Suited for applications to multivariate functions.

EXAMPLE: Approximation of a special multilinear function

Boundary-element application. Solution of $-\Delta u = 0$ in $\Omega \subset \mathbb{R}^3$ with boundary Γ. Ansatz functions: piecewise constant functions for a triangulation $\mathcal{T}$.

Galerkin matrix:

$$M_{\Delta'\Delta''} = \iint_{\Delta'} \iint_{\Delta''} \frac{\mathrm{d}\Gamma_{\mathbf{x}} \mathrm{d}\Gamma_{\mathbf{y}}}{\|\mathbf{x} - \mathbf{y}\|} \qquad (\Delta', \Delta'' \in \mathcal{T}).$$

Difficult cases: $\Delta' \cap \Delta'' \neq \emptyset$.

Case of one common side.
W.l.o.g. the corners of Δ' are $(0,0,0)$, $(1,0,0)$, $(x,y,0)$, while those of Δ'' are $(0,0,0)$, $(1,0,0)$, (ξ,η,τ).

$\Rightarrow M_{\Delta'\Delta''} = f(x,y,\xi,\eta,\tau)$.

Tensor approximation faster than quadrature by a factor of 630 to 2800 (cf. Ballani 2012).

7 PDEs with stochastic coefficients

Literature: Espig-Hackbusch-Litvinenko-Matthies-Wähnert: *Efficient low-rank approximation of the stochastic Galerkin matrix in tensor formats.* Comput. Math. Appl. **67** (2014) 818–829

7.1 Formulation of the problem

Boundary value problem in $D \subset \mathbb{R}^d$ $(1 \le d \le 3)$:

$$\begin{aligned} \operatorname{div} \kappa(x,\omega) \operatorname{grad} u &= f \quad \text{for } x \in D, \omega \in \Omega, \\ u &= 0 \text{ on } \partial D. \end{aligned}$$

Assumption (log-normal distribution):

$$\kappa(x,\omega) = \exp(\gamma(x,\omega)), \qquad \gamma \text{ Gaussian random field.}$$

Solution $u = u(x,\omega) \in L^2(\Omega, H_0^1(D))$.

Weak formulation: $a(u,v) = f(v)$ for all $v \in L^2(\Omega, H_0^1(D))$.

Stochastic quantities:

Mean functions:

$$m_\kappa(x) := \mathbb{E}\left(\kappa(x,\cdot)\right),$$
$$m_\gamma(x) := \mathbb{E}\left(\gamma(x,\cdot)\right),$$

covariance functions:

$$\Gamma_\kappa(x,y) := \mathbb{E}\left[(\kappa(x,\cdot) - m_\kappa(x))\,(\kappa(y,\cdot) - m_\kappa(y))\right],$$
$$\Gamma_\gamma(x,y) := \mathbb{E}\left[(\gamma(x,\cdot) - m_\gamma(x))\,(\gamma(y,\cdot) - m_\gamma(y))\right].$$

Interconnection:

$$\begin{aligned} m_\gamma(x) &= 2\log m_\kappa(x) - \frac{1}{2}\log\left(\Gamma_\kappa(x,x) + m_\kappa(x)^2\right), \\ \Gamma_\gamma(x,y) &= \log\left(1 + \frac{\Gamma_\kappa(x,y)}{m_\kappa(x)m_\kappa(y)}\right). \end{aligned}$$

Singular value decompositions (sums restricted to positive singular values):

$$\tilde{\kappa}(x,\omega) := \kappa(x,\omega) - m_\kappa(x) = \sum_k (\lambda_k)^{1/2} \kappa_k(x)\Phi_k(\omega),$$

$$\tilde{\gamma}(x,\omega) := \gamma(x,\omega) - m_\gamma(x) = \sum_k \left(\lambda'_k\right)^{1/2} \gamma_k(x)\theta_k(\omega).$$

The $L^2(D)$ orthonormal system $\{\kappa_k\}$ are the eigenfunctions of the Hilbert-Schmidt operator

$$C_\kappa \in \mathcal{L}(L^2(D), L^2(D)), \quad (C_\kappa\varphi)(x) = \int_D \Gamma_\kappa(x,y)\varphi(y)\mathrm{d}y,$$

$$C_\kappa \kappa_k = \lambda_k \kappa_k.$$

Similarly,

$$C_\gamma \in \mathcal{L}(L^2(D), L^2(D)), \quad (C_\gamma\varphi)(x) = \int_D \Gamma_\gamma(x,y)\varphi(y)\mathrm{d}y,$$

$$C_\gamma \gamma_k = \lambda'_k \gamma_k.$$

Furthermore,

$$\theta_k(\omega) = \left(\lambda'_k\right)^{-1/2} \int_D \left[\gamma(x,\omega) - m_\gamma(x)\right] \gamma_k(x)\mathrm{d}x$$

are jointly normal distributed and orthonormal random variables in $L^2(\Omega)$.

Uniform ellipticity:

In the following, we assume that

$$\sum_k \left(\lambda_k'\right)^{1/2} \|\gamma_k\|_\infty < \infty.$$

Then one can show that

$$0 < \underline{\kappa} \leq \kappa(x, \omega)$$

holds almost surely and for almost all $x \in D$.

Consequence: Sufficiently small perturbations of $\kappa(x, \omega)$ do not change the ellipticity of the problem.

Multivariate Hermite polynomials $L^2(\Omega)$:

$$H_{\boldsymbol{\iota}}(\mathbf{x}) := \prod_{k=1}^{\infty} h_{\iota_k}(x_k) \quad \text{for } \boldsymbol{\iota} \in \ell_0(\mathbb{N});$$

$$h_i : i\text{-th Hermite polynomial},$$

$$\ell_0(\mathbb{N}) := \{\boldsymbol{\iota} = (\iota_k)_{k\in\mathbb{N}} : \iota_k \in \mathbb{N}_0,\ \iota_k = 0 \text{ for almost all } k \in \mathbb{N}\};$$

Set

$$\boldsymbol{\theta} = (\theta_k)_{k\in\mathbb{N}} \text{ orthonormal system in } L^2(\Omega).$$

Then $\left\{(\boldsymbol{\iota}!)^{-1/2} H_{\boldsymbol{\iota}}(\boldsymbol{\theta}) : \boldsymbol{\iota} \in \ell_0(\mathbb{N})\right\}$ is an orthonormal basis in $L^2(\Omega)$ and

$$\mathbb{E}\left(\kappa(x,\cdot)\,(\boldsymbol{\iota}!)^{-1/2} H_{\boldsymbol{\iota}}(\boldsymbol{\theta})\right) = m_\kappa(x) \prod_k \left(\left(\lambda'_k\right)^{1/2} \gamma_k(x)\right)^{\iota_k} (\iota_k!)^{-1/2}$$

(cf. Janson: Gaussian Hilbert Spaces, 1997; Ullmann: PhD thesis 2008).

The expansion of

$$\tilde{\kappa} = \kappa - m_\kappa = \sum_{\boldsymbol{\iota} \in \ell_0(\mathbb{N})} \sum_{\ell \in \mathbb{N}} \xi_\ell^{(\boldsymbol{\iota})} (\boldsymbol{\iota}!)^{-1/2} \kappa_\ell \otimes H_{\boldsymbol{\iota}}(\boldsymbol{\theta}) \in L^2(D \times \Omega)$$

into the orthonormal basis

$$\left\{ (\boldsymbol{\iota}!)^{-1/2} \kappa_\ell \otimes H_{\boldsymbol{\iota}}(\boldsymbol{\theta}) : \boldsymbol{\iota} \in \ell_0(\mathbb{N}), \ell \in \mathbb{N} \right\}$$

has the coefficients

$$\begin{aligned} \xi_{\ell,\boldsymbol{\iota}} &= (\boldsymbol{\iota}!)^{-1/2} \int_D \kappa_\ell(x) \mathbb{E}\left[\kappa(x,\cdot) H_{\boldsymbol{\iota}}(\boldsymbol{\theta})\right] \mathrm{d}x \\ &= \int_D \kappa_\ell(x) m_\kappa(x) \prod\nolimits_k \left(\left(\lambda_k'\right)^{1/2} \gamma_k(x) \right)^{\iota_k} (\iota_k!)^{-1/2} \, \mathrm{d}x \\ &\quad - \delta_{0\boldsymbol{\iota}} \int_D \kappa_\ell(x) m_\kappa(x) \mathrm{d}x \end{aligned}$$

($\delta_{0\boldsymbol{\iota}}$: Kronecker delta).

7.2 Discretisation

Spatial discretisation: subspace $V_N \subset H_0^1(D)$ spanned by

$$\{\varphi_1, \dots, \varphi_N\}.$$

Stochastic discretisation: subspace $S_J \subset L^2(\Omega)$ spanned by

$$\{H_{\boldsymbol{\iota}}(\boldsymbol{\theta}) : \boldsymbol{\iota} \in J\} \quad \text{with} \quad \#J < \infty, \quad p_k = \max\{\iota_k : \boldsymbol{\iota} \in J\}.$$

Galerkin discretisation:

$$\begin{aligned} & a(\varphi_i \otimes H_{\boldsymbol{\alpha}}(\boldsymbol{\theta}), \varphi_j \otimes H_{\boldsymbol{\beta}}(\boldsymbol{\theta})) \\ = \; & \delta_{\boldsymbol{\alpha}\boldsymbol{\beta}} \int_D m_\kappa(x) \left\langle \nabla\varphi_i(x), \nabla\varphi_j(x) \right\rangle \mathrm{d}x \\ & + \sum_{\ell=1}^{\infty} \xi_\ell^{(\boldsymbol{\iota})} \cdot \mathbb{E}(H_{\boldsymbol{\iota}}(\boldsymbol{\theta}) H_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) H_{\boldsymbol{\beta}}(\boldsymbol{\theta})) \cdot \int_D \kappa_\ell(x) \left\langle \nabla\varphi_i(x), \nabla\varphi_j(x) \right\rangle \mathrm{d}x \end{aligned}$$

Stochastic Galerkin matrix:

$$\mathbf{K} := \Big(a(\varphi_i \otimes H_{\boldsymbol{\alpha}}, \varphi_j \otimes H_{\boldsymbol{\beta}})\Big)_{(i,\boldsymbol{\alpha}),(j,\boldsymbol{\beta})}$$
$$= K_0 \otimes \boldsymbol{\Delta}_{\mathbf{0}} + \sum_{\ell} \sum_{\boldsymbol{\iota} \in J} \xi_\ell^{(\boldsymbol{\iota})} K_\ell \otimes \bigotimes_{k=1}^{K} \boldsymbol{\Delta}_{\iota_k} \quad \in \quad \mathbb{R}^{N \times N} \otimes \bigotimes_{k=1}^{K} \mathbb{R}^{(p_k+1)\times(p_k+1)}$$

with

$$K := \max\{k : \iota_k > 0 \text{ for some } \boldsymbol{\iota} \in J\},$$
$$(\boldsymbol{\Delta}_{\iota_k})_{\alpha\beta} := \mathbb{E}(H_{\iota_k}(\theta_k) H_\alpha(\theta_k) H_{\boldsymbol{\beta}}(\theta_k)), \qquad \boldsymbol{\Delta}_{\iota_k} \in \mathbb{R}^{(p_k+1)\times(p_k+1)},$$
$$(K_\ell)_{ij} := \int_D \kappa_\ell(x) \Big\langle \nabla\varphi_i(x), \nabla\varphi_j(x) \Big\rangle \,\mathrm{d}x, \qquad K_\ell \in \mathbb{R}^{N \times N},$$
$$(K_0)_{ij} := \int_D m_\kappa(x) \Big\langle \nabla\varphi_i(x), \nabla\varphi_j(x) \Big\rangle \,\mathrm{d}x, \qquad K_0 \in \mathbb{R}^{N \times N}.$$

The size of the stochastic Galerkin matrix is

$$\left(N \cdot \prod_{k=1}^{K} (p_k + 1) \right) \times \left(N \cdot \prod_{k=1}^{K} (p_k + 1) \right).$$

Truncation of $\ell \in \mathbb{N}$ in

$$\mathbf{K} = K_0 \otimes \mathbf{\Delta_0} + \sum_{\ell \in \mathbb{N}} \sum_{\boldsymbol{\iota} \in J} \xi_\ell^{(\boldsymbol{\iota})} K_\ell \otimes \bigotimes_{k=1}^{K} \mathbf{\Delta}_{\iota_k}$$

to $\ell \in \{1, \ldots, M\}$ yields a finite expression

$$\mathbf{K} \approx \mathbf{L} := K_0 \otimes \mathbf{\Delta}_0 + \sum_{\ell=1}^{M} \sum_{\boldsymbol{\iota} \in J} \xi_\ell^{(\boldsymbol{\iota})} K_\ell \otimes \bigotimes_{k=1}^{K} \mathbf{\Delta}_{\iota_k}.$$

The approximation error is proportional to $\sum_{\ell=M+1}^{\infty} \lambda_\ell \to 0.$

Question: What is a suitable representation of the huge matrix $\mathbf{L}$ or its approximation?

Later numerical example:
$N = 1000,\ p = 10,\ K = 20 \quad \Rightarrow \quad N \cdot (p+1)^{20} \approx 6.7 \times 10^{23}.$

7.3 Tensor rank of the stochastic Galerkin matrix

$1 + M \cdot \#J$ terms are involved in

$$\mathbf{L} := K_0 \otimes \mathbf{\Delta}_0 + \sum_{\ell=1}^{M} \sum_{\boldsymbol{\iota} \in J} \xi_{\ell,\boldsymbol{\iota}} K_\ell \otimes \bigotimes_{k=1}^{K} \mathbf{\Delta}_{\iota_k}.$$

Assume that we can approximate the tensor

$$\boldsymbol{\xi} \in \mathbb{R}^M \otimes \bigotimes_{k=1}^{K} \mathbb{R}^{p_k+1}$$

by $\boldsymbol{\eta}$ in R-term format: $\boldsymbol{\eta} = \sum_{j=1}^{R} \left[y_j^{(0)} \otimes \bigotimes_{k=1}^{K} y_j^{(k)} \right]$; i.e.,

$$\eta_{\ell,\boldsymbol{\iota}} = \sum_{j=1}^{R} \left[\left(y_j^{(0)} \right)_\ell \cdot \prod_{k=1}^{K} \left(y_j^{(k)} \right)_{\iota_k} \right] \qquad \text{with } y_j^{(0)} \in \mathbb{R}^M \text{ and } y_j^{(k)} \in \mathbb{R}^{p_k+1}.$$

Then

$$\hat{\mathbf{L}} = K_0 \otimes \mathbf{\Delta}_0 + \sum_{j=1}^{R} \left(\sum_{\ell=1}^{M} \left(y_j^{(0)} \right)_\ell K_\ell \right) \otimes \bigotimes_{k=1}^{K} \left(\sum_{\iota_k} \left(y_j^{(k)} \right)_{\iota_k} \mathbf{\Delta}_{\iota_k} \right),$$

i.e., $\hat{\mathbf{L}}$ has an $(1+R)$-term representation: $\hat{\mathbf{L}} \in \mathcal{R}_{1+R}$.
$\Rightarrow$ also the other ranks (Tucker, hierarchical format, TT) are $\leq 1 + R$.

Interludio:

$V_j = \mathbb{K}^{I_j}$, $\mathbf{V} = \bigotimes_j V_j$.

For each $i_j \in I_j$ is associated to a function $f_{i_j}^{(j)}$.

The tensor $\mathbf{v} \in \mathbf{V}$ is defined by

$$\mathbf{v}[i_1, \ldots, i_d] = \int_D \prod_{j=1}^{d} f_{i_j}^{(j)}(x)\mathrm{d}x.$$

Then quadrature yields

$$\mathbf{v}[i_1, \ldots, i_d] \approx \tilde{\mathbf{v}}[i_1, \ldots, i_d] := \sum_{\ell=1}^{R} \omega_\ell \prod_{j=1}^{d} f_{i_j}^{(j)}(x_\ell).$$

Set $v_\ell^{(j)} := \left(f_i^{(j)}(x_\ell) \right)_{i \in I_j} \in V_j$. Then $\tilde{\mathbf{v}}[i_1, \ldots, i_d] = \sum_{\ell=1}^{R} \omega_\ell \prod_{j=1}^{d} v_\ell^{(j)}[i_j]$,

i.e.,

$$\tilde{\mathbf{v}} = \sum_{\ell=1}^{R} \omega_\ell \bigotimes_{j=1}^{d} v_\ell^{(j)} \in \mathcal{R}_R.$$

Explicit description of $\boldsymbol{\xi}$:

$$\xi_{\ell,\boldsymbol{\iota}} := \int_D \kappa_\ell(x) m_\kappa(x) \prod_{k=1}^{K} \left\{ \left[(\lambda'_k)^{\frac{1}{2}} \gamma_k(x) \right]^{\iota_k} (\iota_k!)^{-\frac{1}{2}} \right\} \mathrm{d}x - \delta_{0,\boldsymbol{\iota}} \int_D \kappa_\ell(x) m_\kappa(x) \mathrm{d}x.$$

Apply a quadrature to $\int_D \cdots \mathrm{d}x$: $\quad \xi_{\ell,\boldsymbol{\iota}} \approx \eta_{\ell,\boldsymbol{\iota}} :=$

$$\sum_{j=1}^{R} \omega_j \kappa_\ell(x_j) m_\kappa(x_j) \prod_{k=1}^{K} \left\{ \left[(\lambda'_k)^{1/2} \gamma_k(x_j) \right]^{\iota_k} (\iota_k!)^{-1/2} \right\} - \delta_{\mathbf{0},\boldsymbol{\iota}} \sum_{j'=1}^{R} \omega_{j'} \kappa_\ell(x_{j'}) m_\kappa(x_{j'}).$$

This yields the desired $(R+1)$-term representation of $\boldsymbol{\eta}$:

$$\left(y_j^{(0)}\right)_\ell := \omega_j \, \kappa_\ell(x_j) \, m_\kappa(x_j),$$
$$\left(y_j^{(k)}\right)_{\iota_k} := \left[(\lambda'_k)^{1/2} \gamma_k(x_j) \right]^{\iota_k} (\iota_k!)^{-1/2} \quad (1 \le k \le K)$$

for $1 \le j \le R$.

The additional term for $j = 0$ is

$$\left(y_0^{(0)}\right)_\ell := -\sum_{j'=1}^{R} \omega_{j'} \, \kappa_\ell(x_{j'}) \, m_\kappa(x_{j'}), \qquad \left(y_0^{(k)}\right)_{\iota_k} := \delta_{0,\iota_k}.$$

The error $\|\boldsymbol{\xi} - \boldsymbol{\eta}\|_F$ (quadrature error) does not depend on J (i.e., on K and p_k).

Final problem:

$$\hat{\mathbf{L}}\mathbf{u} = \left(\sum_{j=0}^{R} \hat{K}_j \otimes \hat{\mathbf{\Delta}}_j \right) \mathbf{u} = \mathbf{f}.$$

Let B the approximate inverse of the discrete Laplacian. Then

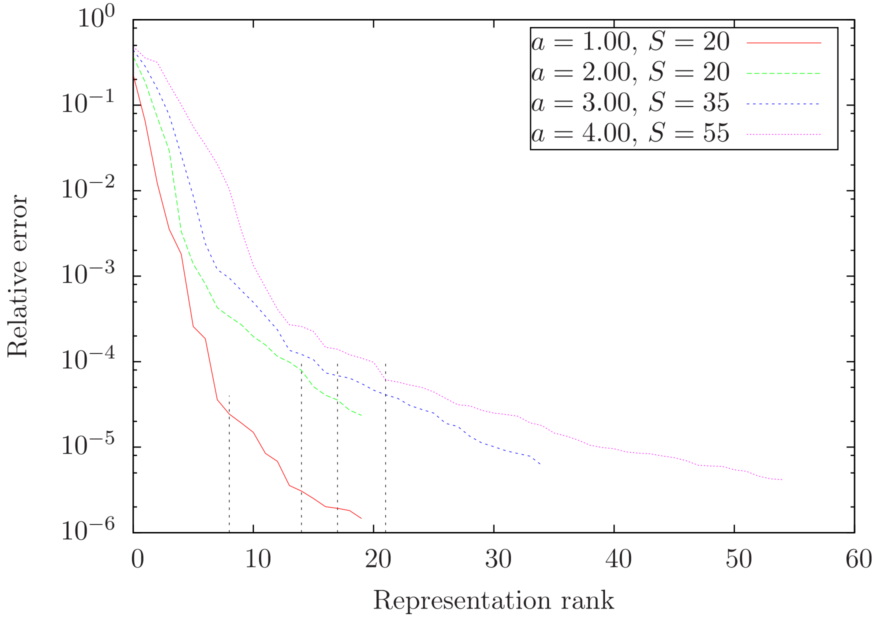
$$\sigma(B\hat{K}_j) = O(1)$$

and $(B \otimes I)\,\hat{\mathbf{L}}$ is well-conditioned.

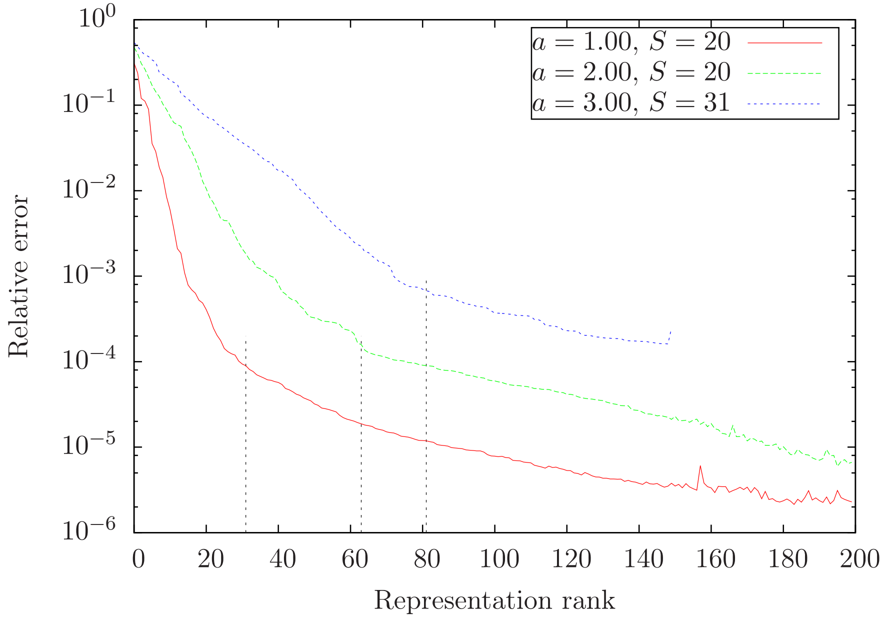
Numerical results with

$$\Gamma_\kappa(x, y) = \exp(-a^2 \|x - y\|^2), \quad \frac{1}{a} \text{ covariance length,}$$

Gaussian quadrature with S points per direction:



$D = (0, 1)$



$D = (0, 1)^2$

8 Minimal Subspaces

8.1 Definitions

We recall the definition of the **algebraic tensor space:**

$$\mathbf{V} := \mathrm{span}\left\{ \bigotimes_{j=1}^{d} v^{(j)} : v^{(j)} \in V_j \right\} =: {}_a\bigotimes_{j=1}^{d} V_j$$

Here, $\dim(V_j) = \infty$ may hold.

Question: Given $\mathbf{v} \in \mathbf{V}$, are there minimal subspaces $U_j^{\min}(\mathbf{v}) \subset V_j$ such that

$$\begin{aligned} \mathbf{v} &\in \bigotimes_{j=1}^{d} U_j^{\min}(\mathbf{v}), \\ \mathbf{v} &\in \bigotimes_{j=1}^{d} U_j \implies U_j^{\min}(\mathbf{v}) \subset U_j. \end{aligned}$$

Such subspaces are the optimal choice for the tensor subspace representation (Tucker).

Elementary results:

1) There are finite-dimensional U_j with $\mathbf{v} \in \bigotimes_{j=1}^d U_j$, more precisely

$$\dim(U_j) \le \mathrm{rank}(\mathbf{v}).$$

2) $\mathbf{v} \in \bigotimes_{j=1}^d U_j'$ and $\mathbf{v} \in \bigotimes_{j=1}^d U_j''$ imply $\mathbf{v} \in \bigotimes_{j=1}^d \left(U_j' \cap U_j''\right)$.

3) The intersection of all U_j with $\mathbf{v} \in \bigotimes_{j=1}^d U_j$ yields $U_j^{\min}(\mathbf{v})$.

Characterisation of $U_j^{\min}(\mathbf{v})$ in the finite-dimensional case:

$$U_j^{\min}(\mathbf{v}) = \mathrm{range}(M_j), \qquad \text{where } M_j := \mathcal{M}_j(\mathbf{v}) \text{ (matricisation)}.$$

The characterisation in the general case needs some notation.

V_j' dual space of V_j. Consider $\varphi^{[j]} := \bigotimes_{k \ne j} \varphi^{(k)}$ with $\varphi^{(k)} \in V_k'$.
$\varphi^{[j]}$ can be regarded as a map from $\mathbf{V} = \bigotimes_{k=1}^d V_k$ onto V_j via

$$\varphi^{[j]}\left(\bigotimes_{k=1}^d v^{(k)}\right) = \left(\prod_{k \ne j} \varphi^{(k)}(v^{(k)})\right) v^{(j)}.$$

If V_j is a normed space, V_j^* denotes the continuous dual space ($V_j^* \subset V_j'$).

Characterisations:

$$U_j^{\mathsf{min}}(\mathbf{v}) = \left\{ \varphi^{[j]}(\mathbf{v}) : \ \varphi^{[j]} \in {}_a\bigotimes_{k\neq j} V_k' \right\},$$

$$U_j^{\mathsf{min}}(\mathbf{v}) = \left\{ \varphi(\mathbf{v}) : \ \varphi \in \left({}_a\bigotimes_{k\neq j} V_k \right)' \right\},$$

although ${}_a\bigotimes_{k\neq j} V_k'$ is strictly smaller than $({}_a\bigotimes_{k\neq j} V_k)'$ in the general infinite-dimensional case.

If V_k and/or ${}_a\bigotimes_{k\neq j} V_k$ are normed spaces, even

$$U_j^{\mathsf{min}}(\mathbf{v}) = \left\{ \varphi^{[j]}(\mathbf{v}) : \ \varphi^{[j]} \in {}_a\bigotimes_{k\neq j} V_k^* \right\},$$

$$U_j^{\mathsf{min}}(\mathbf{v}) = \left\{ \varphi(\mathbf{v}) : \ \varphi \in \left({}_a\bigotimes_{k\neq j} V_k \right)^* \right\}$$

holds.

8.2 Topological Tensor Space

$\left(V_j, \|\cdot\|_j\right)$ are Banach spaces. The topological tensor space $\mathbf{V} := {}_{\|\cdot\|}\bigotimes_{j=1}^{d} V_j$ is the completion of the algebraic tensor space ${}_a\bigotimes_{j=1}^{d} V_j$ w.r.t. a norm $\|\cdot\|$.

A necessary condition for reasonable topological tensor spaces is the continuity of the tensor product, i.e.,

$$\left\| \bigotimes_{j=1}^{d} v^{(j)} \right\| \leq C \prod_{j=1}^{d} \left\| v^{(j)} \right\|_j$$

for some $C < \infty$ and all $v^{(j)} \in V_j$.

DEFINITION: $\|\cdot\|$ is called a crossnorm if

$$\left\| \bigotimes_{j=1}^{d} v^{(j)} \right\| = \prod_{j=1}^{d} \left\| v^{(j)} \right\|_j .$$

REMARK: There are different crossnorms $\|\cdot\|$ for the same $\|\cdot\|_j$!

Reasonable Crossnorms

$\|\cdot\|_j^*$: dual norm corresponding to $\|\cdot\|_j$, i.e. $\|\varphi\|_j^* = \max\{|\varphi(v)| / \|v\|_j : 0 \neq v \in V_j\}$.

DEFINITION: $\|\cdot\|$ is called a reasonable crossnorm if

$$\left\| \bigotimes_{j=1}^d v^{(j)} \right\| = \prod_{j=1}^d \left\| v^{(j)} \right\|_j \qquad \text{for } v^{(j)} \in V_j \quad \text{and}$$
$$\left\| \bigotimes_{j=1}^d \varphi^{(j)} \right\|^* = \prod_{j=1}^d \left\| \varphi^{(j)} \right\|_j^* \qquad \text{for } \varphi^{(j)} \in V_j^*.$$

There are two extreme reasonable crossnorm. The strongest is the projective norm

$$\|\mathbf{v}\|_\wedge := \inf\left\{ \sum_{i=1}^m \prod_{j=1}^d \left\| v_i^{(j)} \right\|_j : \mathbf{v} = \sum_{i=1}^m \bigotimes_{j=1}^d v_i^{(j)} \right\}$$

The weakest is ....

DEFINITION. For $\mathbf{v} \in \mathbf{V} = {}_a\bigotimes_{j=1}^d V_j$ define $\|\cdot\|_\vee$ by

$$\|\mathbf{v}\|_\vee := \sup\left\{ \frac{\left|\left(\varphi^{(1)} \otimes \varphi^{(2)} \otimes \ldots \otimes \varphi^{(d)}\right)(\mathbf{v})\right|}{\|\varphi^{(j)}\|_1^* \|\varphi^{(j)}\|_2^* \cdot \ldots \cdot \|\varphi^{(j)}\|_d^*} : 0 \neq \varphi^{(j)} \in V_j^*, 1 \leq j \leq d \right\}.$$

(injective norm [Grothendieck 1953]).

THEOREM. A norm $\|\cdot\|$ on ${}_a\bigotimes_{j=1}^d V_j$, for which

$$\bigotimes_{j=1}^d \;:\; V_1 \times \ldots \times V_d \to {}_a\bigotimes_{j=1}^d V_j \quad \text{and}$$
$$\bigotimes_{j=1}^d \;:\; V_1^* \times \ldots \times V_d^* \to {}_a\bigotimes_{j=1}^d V_j^*$$

are continuous, cannot be weaker than $\|\cdot\|_\vee$, i.e.,

$$\|\cdot\| \gtrsim \|\cdot\|_\vee . \qquad \text{(norm)}$$

We recall the definition of $\varphi^{[j]} := \bigotimes_{k\neq j} \varphi^{(k)}$ ($\varphi^{(k)} \in V_k'$) by

$$\varphi^{[j]} \left(\bigotimes_{k=1}^{d} v^{(k)} \right) = \left(\prod_{k\neq j} \varphi^{(k)}(v^{(k)}) \right) v^{(j)}.$$

LEMMA. $\varphi \in {}_a\bigotimes_{k\in\{1,\ldots,d\}\setminus\{j\}} V_j^*$ is continuous, i.e., $\varphi \in \mathcal{L}\left({}_\vee\bigotimes_{k=1}^{d} V_k, V_j \right)$. Its norm is

$$\|\varphi\|_{V_j \leftarrow {}_\vee\bigotimes_{k=1}^{d} V_k} = \prod_{k\in\{1,\ldots,d\}\setminus\{j\}} \|v_k^*\|_k^* \,.$$

Consequence: $\varphi \in {}_a\bigotimes_{k\in\{1,\ldots,d\}\setminus\{j\}} V_j^*$ is well defined for topological tensors $\mathbf{v} \in {}_\vee\bigotimes_{k=1}^{d} V_k$. The same conclusion holds for stronger norms than $\|\cdot\|_\vee$, in particular for all *reasonable crossnorms*.

Assume $\|\cdot\| \gtrsim \|\cdot\|_\vee$.

MAIN THEOREM. For $\mathbf{v}_n \in {}_a\bigotimes_{j=1}^d V_j$ assume $\mathbf{v}_n \rightharpoonup \mathbf{v} \in {}_{\|\cdot\|}\bigotimes_{j=1}^d V_j$. Then

$$\dim U_j^{\mathsf{min}}(\mathbf{v}) \leq \liminf_{n\to\infty} \dim U_j^{\mathsf{min}}(\mathbf{v}_n) \qquad \text{for all } 1 \leq j \leq d.$$

THEOREM. The sets $\mathcal{T}_\mathbf{r}$ and $\mathcal{H}_\mathbf{r}$ are weakly closed.

PROOF. Let $\mathbf{v}_n \in \mathcal{T}_\mathbf{r}$, i.e., there are subspaces $U_{j,n}$ with $\mathbf{v}_n \in \bigotimes_{j=1}^d U_{j,n}$ and $\dim U_{j,n} \leq r_j$. Note that $U_j^{\mathsf{min}}(\mathbf{v}_n) \subset U_{j,n}$ with $\dim U_j^{\mathsf{min}}(\mathbf{v}_n) \leq r_j$.

If $\mathbf{v}_n \rightharpoonup \mathbf{v}$, then $\dim U_{j,\mathsf{min}}(\mathbf{v}) \leq r_j$ and therefore $\mathbf{v} \in \mathcal{T}_\mathbf{r}$. Similar for $\mathcal{H}_\mathbf{r}$.

Application to Best Approximation

THEOREM. Let $(X, \|\cdot\|)$ be a reflexive Banach space with a weakly closed subset $\emptyset \neq M \subset X$. Then for any $x \in X$ there exists a best approximation $v \in M$ with

$$\|x - v\| = \inf\{\|x - w\| : w \in M\}.$$

LEMMA A. If $x_n \rightharpoonup x$, then $\|x\| \leq \liminf_{n\to\infty} \|x_n\|$.

LEMMA B. If X is a reflexive Banach space, any bounded sequence $x_n \in X$ has a subsequence x_{n_i} converging weakly to some $x \in X$.

PROOF of the Theorem. Choose $w_n \in M$ with $\|x - w_n\| \to \inf\{\|x - w\| : w \in M\}$. Since $(w_n)_{n\in\mathbb{N}}$ is a bounded sequence in X, LEMMA B ensures the existence of a subsequence $w_{n_i} \rightharpoonup v \in X$. v belongs to M because $w_{n_i} \in M$ and M is weakly closed. Since also $x - w_{n_i} \rightharpoonup x - v$, LEMMA A shows $\|x - v\| \leq \liminf \|x - w_{n_i}\| \leq \inf\{\|x - w\| : w \in M\}$.

Conclusion for $M \in \{\mathcal{T}_\mathbf{r}, \mathcal{H}_\mathbf{r}\}$:

COROLLARY. Let $\|\cdot\|$ satisfy $\|\cdot\| \gtrsim \|\cdot\|_\vee$ and let $(V, \|\cdot\|)$ be reflexive. Then best approximations in the formats $\mathcal{T}_\mathbf{r}$ and $\mathcal{H}_\mathbf{r}$ exist.

9 Properties of the HOSVD Projection

We recall: The Tucker and hierarchical representation may be based on the HOSVD bases $\left\{b_\ell^{(\alpha)} : 1 \leq \ell \leq r_\alpha\right\}$. The HOSVD projection is of the form

$$P = P_\alpha \otimes P_{\alpha^c} \quad \text{with } P_\alpha b_\ell^{(\alpha)} = \begin{cases} b_\ell^{(\alpha)} & \text{for } 1 \leq \ell \leq s_\alpha, \\ 0 & \text{for } s_\alpha < \ell \leq r_\alpha \end{cases}$$

Let

$$\mathbf{u}_{\text{HOSVD}} = P\mathbf{v}.$$

LEMMA. Let $\phi_j \mathbf{v} = 0$ for some $\phi_j = id \otimes \ldots \otimes \varphi_j \otimes id \otimes \ldots \otimes id$, $\varphi_j \in V_j'$. Then also $\phi_j \mathbf{u}_{\text{HOSVD}} = 0$.

LEMMA. If $\mathbf{v} \in \mathbf{V}$ belongs to the domain of ϕ_j, then also $\mathbf{u}_{\text{HOSVD}}$ belongs to the domain and satisfies

$$\left\|\phi_j \mathbf{u}_{\text{HOSVD}}\right\| \leq \left\|\phi_j \mathbf{v}\right\|.$$

Application: $\left\|\partial^k \mathbf{u}_{\text{HOSVD}} / \partial x_j^k\right\|_{L^2} \leq \left\|\partial^k \mathbf{v} / \partial x_j^k\right\|_{L^2}.$

L^∞ Estimates

Problem:

- HOSVD projection uses the underlying Hilbert norm (L^2)
- Pointwise evaluations require the maximum norm (L^∞)

Gagliardo-Nirenberg inequality:

$$\|\varphi\|_\infty \le c_m^\Omega \, |\varphi|_m^{\frac{d}{2m}} \, \|\varphi\|_{L^2}^{1-\frac{d}{2m}}, \qquad \text{where}$$

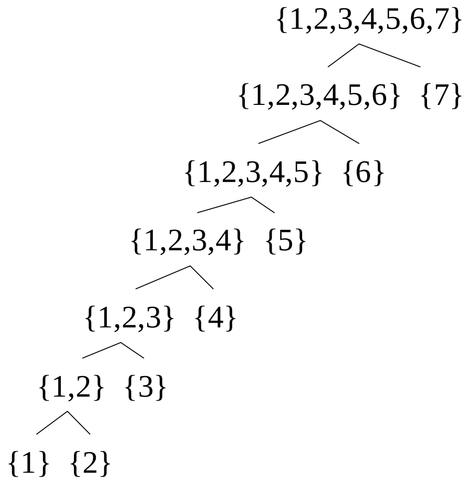
$$|\varphi|_m := \left(\int_\Omega \sum_{j=1}^{d} \left| \frac{\partial^m \varphi}{\partial x_j^m} \right|^2 \mathrm{d}x \right)^{1/2}.$$

For $\Omega = \mathbb{R}^d$ we have

$$\lim_{m \to \infty} c_m^\Omega = \pi^{-d/2}.$$

10 Graph-Based Formats

10.1 Matrix-Product (TT) Format

A particular binary tree is 

. It leads to the TT format (Oseledets–Tyrtyshnikov 2005) and coincides with the description of the matrix product states (Vidal 2003, Verstraete–Cirac 2006) used in physics:

Each component $\mathbf{v}[i_1, \ldots, i_d]$ of $\mathbf{v} \in \mathbf{V} = \bigotimes_{j=1}^{d} \mathbb{K}^{n_j}$ is expressed by

$$\mathbf{v}[i_1 i_2 \cdots i_d] = M^{(1)}[i_1] \cdot M^{(2)}[i_2] \cdot \ldots \cdot M^{(d-1)}[i_{d-1}] \cdot M^{(d)}[i_d] \in \mathbb{K},$$

where $M^{(j)}[i]$ are matrices of size $r_{j-1} \times r_j$ with $r_0 = r_d = 1$.

To avoid the special roles of the vectors $M^{(1)}[i_1], M^{(d)}[i_d]$ and to describe periodic situations, the Cyclic Matrix-Product format $\mathcal{C}(d,(r_j),(n_j))$, $n_j = \dim V_j$, is used in physics:

$$\mathbf{v}[i_1 i_2 \cdots i_d] = \mathrm{trace}\{M^{(1)}[i_1] \cdot M^{(2)}[i_2] \cdot \ldots \cdot M^{(d-1)}[i_{d-1}] \cdot M^{(d)}[i_d]\}$$
$$= \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} M_{k_d k_1}{}^{(1)}[i_1] \cdot M^{(2)}_{k_1 k_2}[i_2] \cdot \ldots \cdot M^{(d-1)}[i_{d-1}] \cdot M_{k_{d-1} k_d}{}^{(d)}[i_d].$$

Tensor Network: tensor representations based on general graphs.

THEOREM (Landsberg–Qi–Ye 2012) Formats based on a graph$\neq$tree are in general not closed.

10.2 Intermezzo: Algebra Structure Tensors

V algebra, i.e. vector space with additional operation $\circ$, $\{b_k\}$ basis of V.
The operation is completely described by the coefficients s_{ijk} in

$$b_i \circ b_j = \sum_k s_{ijk} b_k.$$

Let $b_i^* \in V'$ the dual element with $\left\langle \sum_k \alpha_k b_k, b_i^* \right\rangle = \alpha_i$. Then

$$\mathrm{s} := \sum_{i,j,k} s_{ijk} b_i^* \otimes b_j^* \otimes b_k \in V' \otimes V' \otimes V$$

is the *structure tensor* of the algebra.

Remark. For $v, w \in V$ we have $v \circ w = (v \otimes w \otimes id)\, \mathrm{s}$.

Proof: Let $v = \sum_i v_i b_i$ and $w = \sum_j w_j b_j$. Then

$$\begin{aligned}(v \otimes w \otimes id)\, \mathrm{s} &= \sum_{i,j,k} s_{ijk} \langle v, b_i^* \rangle \left\langle w, b_j^* \right\rangle b_k = \sum_{i,j,k} s_{ijk} v_i w_j b_k = \sum_{i,j} v_i w_j \sum_k s_{ijk} b_k \\ &= \sum_{i,j} v_i w_j b_i \circ b_j = \left(\sum_i v_i b_i \right) \circ \left(\sum_j w_j b_j \right) = v \circ w.\end{aligned}$$

Matrix Multiplication

Consider $V = \mathbb{K}^{2\times 2}$, $\circ = *$ is the matrix multiplication.

The basis of V is $\{E_{pq} : 1 \le p, q \le 2\}$, where $E_{pq}[i,j] = \begin{cases} 1 & \text{for } (i,j) = (p,q) \\ 0 & \text{otherwise.} \end{cases}$

LEMMA. The structure tensor of the matrix multiplication is

$$\mathbf{m} := \sum_{i_1=1}^{2} \sum_{i_2=1}^{2} \sum_{i_3=1}^{2} E^*_{i_1,i_2} \otimes E^*_{i_2,i_3} \otimes E_{i_1,i_3} \in V' \otimes V' \otimes V.$$

Proof. Let $A, B \in \mathbb{K}^{2\times 2}$ and $C := AB$. Then

$$(A \otimes B \otimes id)\mathbf{m} = \sum_{i_1,i_2,i_3=1}^{2} A_{i_1,i_2} B_{i_2,i_3} E_{i_1,i_3} = \sum_{i_1,i_3=1}^{2} C_{i_1,i_3} E_{i_1,i_3} = C.$$

THEOREM: $\text{rank}(\mathbf{m}) = \underline{\text{rank}}(\mathbf{m}) = 7$.

Strassen, 1969: $\text{rank}(\mathbf{m}) \le 7$; Winograd, 1971: $\text{rank}(\mathbf{m}) = 7$;
Landsberg, 2012: $\underline{\text{rank}}(\mathbf{m}) = 7$.

10.3 Cyclic Matrix-Product Format

We recall the Cyclic Matrix-Product Format $\mathcal{C}(d,(r_j),(n_j))$

$$\mathbf{v}[i_1 i_2 \cdots i_d] = \mathrm{trace}\{M^{(1)}[i_1] \cdot M^{(2)}[i_2] \cdot \cdots \cdot M^{(d-1)}[i_{d-1}] \cdot M^{(d)}[i_d]\}$$
$$= \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} M^{(1)}_{k_d k_1}[i_1] \cdot M^{(2)}_{k_1 k_2}[i_2] \cdot \ldots \cdot M^{(d-1)}[i_{d-1}] \cdot M^{(d)}_{k_{d-1} k_d}[i_d].$$

A subcase is the site-independent format Matrix-Product Format $\mathcal{C}_{\mathrm{ind}}(d,r,n)$ with

$$\begin{aligned} M^{(j)}[i] &= M[i] \\ r_j &= r, \\ V_j &= V \qquad \text{for all } j, \\ n &= \dim V. \end{aligned}$$

THEOREM (Landsberg–Qi–Ye 2012) Formats based on a graph$\neq$tree are in general not closed.

10.4 Result for $d = 3$, $\mathbf{V} = \otimes^3 \mathbb{K}^{2\times 2}$, $r_1 = r_2 = r_3 = 2$ by Harris–Michałek–Sertöz 2018

Let

$$\mathbf{m} := \sum_{k_1,k_2,k_3=1}^{2} E_{k_3,k_1} \otimes E_{k_1 k_2} \otimes E_{k_2,k_3} \in \bigotimes_{j=1}^{3} \mathbb{K}^{2\times 2}.$$

$\{E_{pq} : 1 \leq p, q \leq 2\}$ is the canonical basis of $\mathbb{K}^{2\times 2}$.

LEMMA. Let $\mathbf{V} = \bigotimes_{j=1}^{3} \mathbb{K}^{2\times 2}$. The set $\mathcal{C}(d = 3, (r_j = 2), (n_j = 4))$ consists of all

$$\mathbf{v} = \Phi(\mathbf{m}) \qquad \text{with} \quad \Phi = \bigotimes_{j=1}^{3} \phi^{(j)} \quad \text{and} \quad \phi^{(j)} \in L(\mathbb{K}^{2\times 2}, \mathbb{K}^{2\times 2}).$$

REMARK. a) $\mathbf{m}$ is equivalent to the Strassen tensor of the matrix multiplication.
b) If all $\phi^{(j)}$ are bijective, $\mathbf{v} = \Phi(\mathbf{m})$ implies that $\operatorname{rank}(\mathbf{v}) = 7$.

We consider the *site-independent* case $M^{(j)}[i] = M[i]$ for all $1 \le j \le d := 3$.

Define $\psi \in L(\mathbb{K}^{2\times 2}, \mathbb{K}^{2\times 2})$ by $\psi(E_{12}) = E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $\psi(E_{pq}) = 0$ for $(p,q) \ne (1,2)$ and

$$\mathbf{v}(t) = \left(\otimes^3(\psi + t \cdot id)\right)(\mathbf{m}) = \mathbf{v}_0 + t \cdot \mathbf{v}_1 + t^2 \cdot \mathbf{v}_2 + t^3 \cdot \mathbf{v}_3 \in \mathcal{C}_{\mathsf{ind}}(3,2,4)$$

with

$$\begin{aligned} \mathbf{v}_0 &= (\otimes^3\psi)(\mathbf{m}), \qquad \mathbf{v}_1 = [\psi \otimes \psi \otimes id + \psi \otimes id \otimes \psi + id \otimes \psi \otimes \psi](\mathbf{m}), \\ \mathbf{v}_2 &= [id \otimes id \otimes \psi + id \otimes \psi \otimes id + \psi \otimes id \otimes id](\mathbf{m}), \qquad \mathbf{v}_3 = \mathbf{m}. \end{aligned}$$

$\Rightarrow$ $\mathbf{v}_0 = \mathbf{v}_1 = 0$ and

$$\begin{aligned} \mathbf{v}_2 = {} & E_{21} \otimes E_{11} \otimes E_{12} + E_{22} \otimes E_{21} \otimes E_{12} + E_{11} \otimes E_{12} \otimes E_{21} \\ & + E_{21} \otimes E_{12} \otimes E_{22} + E_{12} \otimes E_{21} \otimes E_{11} + E_{12} \otimes E_{22} \otimes E_{21}, \end{aligned}$$

$\Rightarrow$ $\mathsf{rank}(\mathbf{v}_2) \le 6$. The following limit exists:

$$\mathbf{v}_2 = \lim_{t \to 0} t^{-2}\mathbf{v}(t) \in closure(\mathcal{C}_{\mathsf{ind}}(3,2,4)).$$

The non-closedness of $\mathcal{C}_{\text{ind}}(3,2,4)$ will follow from $\mathbf{v}_2 \notin \mathcal{C}_{\text{ind}}(3,2,4)$.

For an indirect proof assume $\mathbf{v}_2 \in \mathcal{C}_{\text{ind}}(3,2,4)$. The Lemma implies that there is some $\phi \in L(\mathbb{K}^{2\times 2}, \mathbb{K}^{2\times 2})$ with $\mathbf{v}_2 = (\otimes^3 \phi)(\mathbf{m})$.
It is easy to check that the range of the matricisation

$$\mathcal{M}_1((\otimes^3\phi)(\mathbf{m})) = \phi \mathcal{M}_1(\mathbf{m})(\otimes^2\phi)^{\mathsf{T}}$$

is $\mathbb{K}^{2\times 2}$. Therefore the map ϕ must be surjective.

Since $\phi \in L(\mathbb{K}^{2\times 2}, \mathbb{K}^{2\times 2})$ is surjective, it is also injective and thus bijective.

By Remark (b) $\text{rank}(\mathbf{v}_2) = \text{rank}(\mathbf{m}) = 7$ holds in contradiction to $\text{rank}(\mathbf{v}_2) \le 6$.

This contradiction proves that $\mathbf{v}_2 \notin \mathcal{C}_{\text{ind}}(3,2,4)$.

Similarly $\mathbf{v}_2 \notin \mathcal{C}(3,(2,2,2),(4,4,4))$ follows (no site-independence).

10.5 Result for $\mathbf{V} = \otimes^d \mathbb{C}^2$, $r_j = 2$

Smallest (nontrivial) dimension: $V_j = \mathbb{C}^2$,

tensor space $\mathbf{V} = \otimes^d \mathbb{C}^2$

Site-independent cyclic format $\mathcal{C}_{\text{ind}}(d, 2, 2)$, i.e., $r = 2$

Result:

$d = 3$: $\mathcal{C}_{\text{ind}}(3, 2, 2)$ is closed (cf. Harris–Michałek–Sertöz 2018)

$d > 3$: $\mathcal{C}_{\text{ind}}(d, 2, 2)$ is not closed (cf. Tim Seynnaeve 2018)

Same for $\mathbb{K} = \mathbb{R}$

Extension to Larger Spaces

$d \geq 4$:
$\mathcal{C}_{\mathsf{ind}}(d, r = 2, n = 2)$ not closed $\Rightarrow \mathcal{C}_{\mathsf{ind}}(d, r = 2, n)$ not closed for all $n \geq 2$.

Missing case
$\mathcal{C}_{\mathsf{ind}}(3, 2, 2)$ closed, $\mathcal{C}_{\mathsf{ind}}(3, 2, 4)$ not closed
Also $\mathcal{C}_{\mathsf{ind}}(3, 2, 3)$ is not closed (Tim Seynnaeve, technical proof).

Case of $r \geq 3$

$\mathbf{V}_{cycl} := \{\mathbf{v} \in \mathbf{V} : \pi\mathbf{v} = \mathbf{v}\}$ for $\pi : (1, 2, \ldots, d) \mapsto (2, \ldots, d, 1)$

Let $d > 3,\ n \geq 2,\ \mathbb{K} = \mathbb{C}$ or d odd:
a) $\mathcal{C}_{\mathsf{ind}}(d, r, n)$ not closed for $r = 2$
b) r sufficiently large $\Rightarrow \mathcal{C}_{\mathsf{ind}}(d, r, n) = \mathbf{V}_{cycl} \Rightarrow \mathcal{C}_{\mathsf{ind}}(d, r, n)$ closed.

Another reason for closedness of $\mathcal{C}_{\mathsf{ind}}(3, 2, 2)$ (Proof: Tim Seynnaeve):

$$\mathcal{C}_{\mathsf{ind}}(3, 2, 2) = \mathbf{V}_{\mathsf{cycl}}.$$

11 Tensorisation

$V_j = \mathbb{R}^n \Rightarrow$ storage: $rdn + (d-1)r^3$. Now: $n \to O(\log n)$

Let the vector $y \in \mathbb{R}^n$ represent the grid values of a function in $(0,1]$:

$$y_\mu = f\left(\frac{\mu+1}{n}\right) \qquad (0 \le \mu \le n-1)\,.$$

Choose, e.g., $n = 2^d$, and note that $\mathbb{R}^n \cong \mathbf{V} := \bigotimes_{j=1}^d \mathbb{R}^2$.
Isomorphism by binary integer representation:
$\mu = \sum_{j=1}^d \mu_j 2^{j-1}$ with $\mu_j \in \{0,1\}$, i.e.,

$$y_\mu = \mathbf{v}[\mu_1, \mu_2, \ldots, \mu_{d-1}, \mu_d].$$

Algebraic Function Compression (black-box procedure)
1) Tensorisation: $y \in \mathbb{R}^n \longmapsto \mathbf{v} \in \mathbf{V}$ (storage size: $n = 2^d$)
2) Apply the *tensor truncation*: $\mathbf{v} \longmapsto \mathbf{v}_\varepsilon$
3) Observation: often the data size decreases from $n = 2^d$ to $O(d) = O(\log n)$.

EXAMPLE

$y \in \mathbb{C}^n$ with $y_\mu = \zeta^\mu$ leads to an *elementary tensor* $\mathbf{v} \in \mathbf{V}$, i.e.,

$$\mathbf{v} = \bigotimes_{j=1}^{d} v^{(j)} \qquad \text{with } v^{(j)} = \begin{bmatrix} 1 \\ \zeta^{2^{j-1}} \end{bmatrix} \in \mathbb{C}^2.$$

Storage size $= 2d = 2\log_2 n$.

Consequence:

All functions $f \in C((0,1])$, which can be well-approximated by r *trigonometric terms* or *exponential sums* with r terms (even with complex coefficients $\rightarrow$ Bessel functions) can be approximated by a tensor representation with data size

$$2dr = O(r \log n).$$

Example:

$f(x) = 1/(x+\delta) \in C((0,1])$, $\delta \geq 0$, can be well-approximated by exponential sums (cf. Braess-H.):

$$f(x) \approx \sum_{\nu=1}^{r} a_\nu \exp(-b_\nu x) \qquad (a_\nu, b_\nu > 0)$$

error: $O(n \exp(-2^{1/2}\pi r^{1/2}))$ if $\delta = 0$,

$O(\exp(-cr)) \qquad$ if $\delta = O(1)$.

Storage size:

$$2dr = 2r \log_2 n = O(\log^2(\varepsilon)\log(n))$$

Hierarchical Format, Matricisation

Tree:

```
1 — {1,2} — {1,2,3} — ... {1,...,d-1} — {1,...,d}
  /       /         /  ...            /
2       3         4    ...          d
```

(also called TT format)

Consider the tensorisation $\mathbf{v} \in \bigotimes_{j=1}^{d} \mathbb{R}^2$ of the vector $y = (y_0, \ldots, y_{n-1}) \in \mathbb{R}^n$.
The matricisation for $\alpha = \{1, \ldots, j\}$ $(1 \leq j \leq d-1)$ yields

$$\mathcal{M}_\alpha(\mathbf{v}) = \begin{bmatrix} y_0 & y_m & \cdots & y_{n-m} \\ y_1 & y_{m+1} & \cdots & y_{n-m+1} \\ \vdots & \vdots & & \vdots \\ y_{m-1} & y_{2m-1} & \cdots & y_{n-1} \end{bmatrix} \text{ with } m := 2^j.$$

Recall: $\mathrm{rank}_\alpha(\mathbf{v}) = \dim \mathcal{M}_\alpha(\mathbf{v})$.

p-Methods

$f(x) \approx \tilde{f}(x) = \sum_{k=1}^{r} a_k e^{2\pi i(k-1)}$ trigonometric approximation

$\Rightarrow$ tensorisation, storage $2dr = 2r \log_2 n$, error $\leq \left\| f - \tilde{f} \right\|$

Similar for $\tilde{f}(x) = \sum_{k=1}^{r} a_k \sin(2\pi ik)$ etc.

Polynomials:

$f(x) \approx P(x)$, P polynomial of degree $\leq p$

An r-term representation $\sum_{i=1}^{r} \bigotimes_{j=1}^{d} v_i^{(j)}$ does not work well.

Instead, the hierarchical format (in particular, the TT format) is used.

Conclusion for polynomial p-methods

If $\mathbf{f} \approx \mathbf{P}$ with a polynomial $\mathbf{P}$ of degree $\leq p$ ($\Rightarrow$ data size $p+1$), then the tensorised grid function $\mathbf{f}$ can be approximated by a tensor $\tilde{\mathbf{f}}$ such that the TT ranks are bounded by $\rho_j \leq p+1$:

$$\left\| \mathbf{f} - \tilde{\mathbf{f}} \right\|_2 \leq \left\| \mathbf{f} - \mathbf{P} \right\|_2$$

The data size is bounded by

$$\leq 2d\,(p+1)^2\,.$$

hp Method

Let f be an asymptotically smooth function in $(0,1]$ with possible singularity at $x=0$, e.g., $f(x)=x^x$.
Use the (best) piecewise polynomial approximation $\tilde{f}$ (by degree p) in all intervals

$$[0,\tfrac{1}{n}],\ [\tfrac{1}{n},\tfrac{2}{n}],\ [\tfrac{2}{n},\tfrac{4}{n}],\ldots,[\tfrac{1}{4},\tfrac{1}{2}],\ [\tfrac{1}{2},1].$$

Required data size of hp method: $(p+1)\log_2 n$.
Tensor ranks:

$$\begin{aligned} r_1 &\le \dim(span\{\tilde{f}|_{[(\mu-1)h,\mu h]} : 1\le\mu\le n\}) \le p+1,\\ r_2 &\le \dim(span\{\tilde{f}|_{[(\mu-1)2h,\mu 2h]} : 1\le\mu\le \tfrac{n}{2}\}) \le p+2,\\ r_3 &\le \dim(span\{\tilde{f}|_{[(\mu-1)4h,\mu 4h]} : 1\le\mu\le \tfrac{n}{4}\}) \le p+2,\\ &\vdots \end{aligned}$$

Hence, the data size of the tensorisation of $\tilde{f}$ is bounded by

$$d\,(p+2)^2 = (p+2)^2\log_2 n.$$

THEOREM (Grasedyck 2010) f asymptotically smooth with m point singularities. Then the data size of $\mathbf{v}_\varepsilon$ corresponding exactly to a piecewise polynomial approximation is characterised by

$$r = O(1) + \log_2\tfrac{1}{\varepsilon} + 2m.$$