Numerical Tensor Calculus

Wolfgang Hackbusch

Max-Planck-Institut für *Mathematik in den Naturwissenschaften*

and

Christians-Albrechts-Universität zu Kiel

wh@mis.mpg.de

https://www.mis.mpg.de/scicomp/hackbusch.en.html

Creswick, February 8, 2018

Overview

Tensors

- Where do large-scale tensors appear?
- Tensor operations
- High-dimensional problems in practice

Tensor Representations

- r-Term Format (Canonical Format)
- Tensor Subspace Format (Tucker Format)
- Matricisation and Tucker Ranks
- HOSVD: Higher Order Singular-Value Decomposition

Hierarchical Format

- Dimension Partition Tree
- Algorithmic Realisation
- Operations

Solution of Linear Systems

Multivariate Cross Approximation

Tensorisation

1 Tensors

1.1 Where do large-scale tensors appear in Numerical Analysis?

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:

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

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:

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

Tensor product of vectors $v^{(j)} \in \mathbb{R}^{n_j}$:

$$\left(\bigotimes_{j=1}^{d} v^{(j)} \right) [i_1, i_2, \ldots, i_d] := \prod_{j=1}^{d} v^{(j)}[i_j].$$

Challenge: How to treat tensors when $N = n^d$ is huge ($N \gg$ memory space)?

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 $(1 \leq j \leq d)$.

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 $\mathbf{A} \in \otimes^d \mathbb{R}^{n \times n} \simeq \mathbb{R}^{n^d \times n^d}$.

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 $\mathbf{u_i} = \sum_{\mathbf{0} \le \mathbf{k} \le \mathbf{i}} \mathbf{v_{i-k}} \mathbf{w_k}$

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

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 (including the TT format)

2.1 r-Term Format (Canonical Format)

By definition, each 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

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

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

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

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 instabilities, etc.)

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 $\mathbf{a} \in \bigotimes_{j=1}^{d} \mathbb{K}^{r_j}$ has $\prod_{j=1}^{d} r_j$ entries. Disadvantage for large d.

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

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

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:

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

Sometimes, $\mathbf{r} := (\text{rank}_1(\mathbf{v}), \ldots, \text{rank}_d(\mathbf{v}))$ 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 $\mathcal{T}_{\mathbf{r}}$:

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

Also for $\dim V_j = \infty$, $\mathrm{rank}_j(\mathbf{v})$ can be defined.

Under rather general assumptions on the norms of V_j and $\mathbf{V}$ one proves that

$$\mathbf{v}_n \rightharpoonup \mathbf{v} \qquad \Rightarrow \qquad \mathrm{rank}_j(\mathbf{v}) \le \underline{\lim}_{n\to\infty} \mathrm{rank}_j(\mathbf{v}_n).$$

Conclusion: 1) $\mathcal{T}_{\mathbf{r}}$ is weakly closed.
2) If $\mathbf{V}$ is a reflexive Banach space, $\inf_{\mathbf{u}\in\mathcal{T}_{\mathbf{r}}} \|\mathbf{v}-\mathbf{u}\| = \|\mathbf{v}-\mathbf{u}_{\mathrm{best}}\|$ has a solution $\mathbf{u}_{\mathrm{best}} \in \mathcal{T}_{\mathbf{r}}$.

2.2.4 HOSVD: Higher Order Singular-Value Decomposition

Diagonalisation:

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

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

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:

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

Conclusions 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 $\mathbf{U}_\alpha := \mathbf{U}_\alpha^{\mathrm{min}}(\mathbf{v})$.

For $\alpha \subset D := \{1, \ldots, d\}$ and $\alpha^c := D \backslash \alpha$, the minimal subspaces $U_\alpha^{\mathrm{min}}(\mathbf{v})$ and $U_{\alpha^c}^{\mathrm{min}}(\mathbf{v})$ satisfy

$$\mathbf{v} \in U_\alpha^{\mathrm{min}}(\mathbf{v}) \otimes U_{\alpha^c}^{\mathrm{min}}(\mathbf{v})$$

with minimal dimension.

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

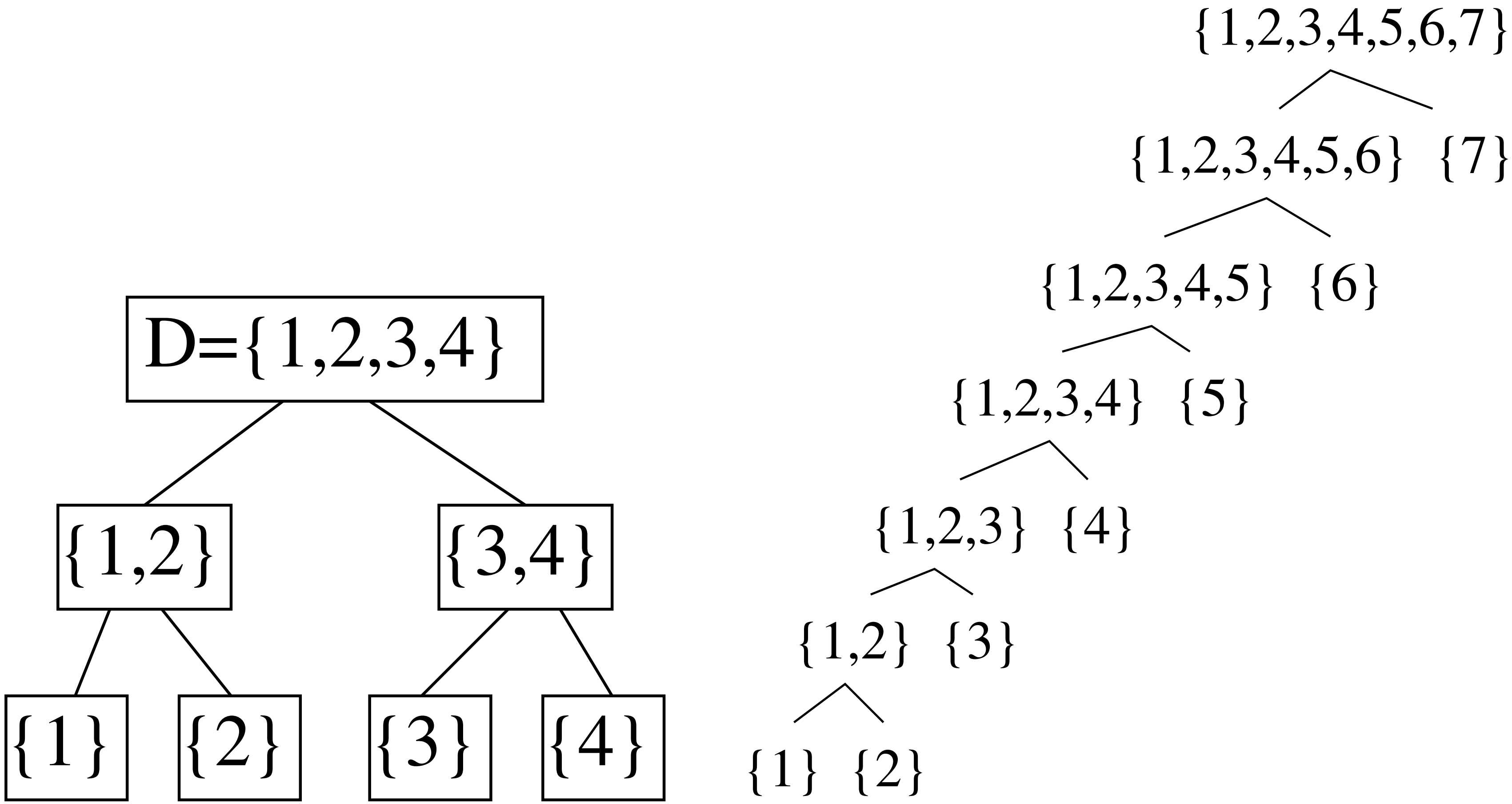

Figure 1: Balanced tree and linear tree

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

$$\mathbf{b}_\ell^{(\{1,2\})} = \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} c_{ij}^{(\{1,2\},\ell)}\, b_i^{(1)} \otimes b_j^{(2)}$$

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 $\mathbf{v} = c_1 \mathbf{b}_1^{(\{1,\dots,d\})}$.

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

3.3 Truncation, Operations

Operations are typically recursive w.r.t. to the tree structure.

They involve only the data $\left[C^{(\alpha,\ell)}, c_1, b_\nu^{(j)}\right]$.

The typical operation cost is

$$O(dr^4 + dnr^2).$$

The HOSVD truncation is based on SVDs involving the coefficient matrices

$$C^{(\alpha,\ell)} \in \mathbb{R}^{r_{\alpha_1} \times r_{\alpha_2}}.$$

(α_1, α_2: sons of α).

3.4 Operations - Example: scalar product

Let $\mathbf{v},\mathbf{w}\in\mathbf{V}$ be given by $\left(C'^{(\alpha,\ell)},c_1',b_\nu'^{(j)}\right)$ and $\left(C''^{(\alpha,\ell)},c_1'',b_\nu''^{(j)}\right)$ resp.

$$\mathbf{v}=c_1'\mathbf{b}_1'^{(D)},\ \mathbf{w}=c_1''\mathbf{b}_1''^{(D)}\quad\Rightarrow\quad\langle\mathbf{v},\mathbf{w}\rangle=c_1'c_1''\left\langle\mathbf{b}_1'^{(D)},\mathbf{b}_1''^{(D)}\right\rangle.$$

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

$$\begin{aligned}\beta_{ij}^{(\alpha)}&=\left\langle\mathbf{b}_i'^{(\alpha)},\mathbf{b}_j''^{(\alpha)}\right\rangle=\left\langle\sum_{k,\ell}c_{k,\ell}'^{(\alpha,i)}b_k'^{(\alpha_1)}\otimes b_\ell'^{(\alpha_2)},\sum_{p,q}c_{p,q}''^{(\alpha,j)}b_p''^{(\alpha_1)}\otimes b_q''^{(\alpha_2)}\right\rangle\\&=\sum_{k,\ell}\sum_{p,q}c_{k,\ell}'^{(\alpha,i)}c_{p,q}''^{(\alpha,j)}\left\langle b_k'^{(\alpha_1)},b_p''^{(\alpha_1)}\right\rangle\left\langle b_\ell'^{(\alpha_2)},b_q''^{(\alpha_2)}\right\rangle\\&=\sum_{k,\ell}\sum_{p,q}c_{k,\ell}'^{(\alpha,i)}c_{p,q}''^{(\alpha,j)}\beta_{kp}^{(\alpha_1)}\beta_{\ell q}^{(\alpha_2)}\end{aligned}$$

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

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.

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

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). Used in Control Theory and Model Reduction.

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, Uschmajew,...

5 Multivariate Cross Approximation

Matrix Case

Choose suitable r rows and columns:

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

They define a rank-r matrix M_r interpolating M at these rows and columns.

If $\text{rank}(M) = r$, then $M_r = M$.

Order $d \geq 3$: In principle similar when using the hierarchical format.
Required number of evaluations of the tensor is $O\left(\sum_j n_j\right)$.

6 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.,

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

Application of tensor tools (SVD: 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 $= 2d = 2\log_2 n$.

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: $\quad O(n\exp(-2^{1/2}\pi r^{1/2}))$ for $\delta = 0$, $\qquad O(\exp(-cr))$ for $\delta = O(1)$.

Storage size:

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

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

Hierarchical Format, Matricisation

Tree:

$$\begin{array}{ccccccccccc} 1 & — & \{1,2\} & — & \{1,2,3\} & — & \ldots & \{1,\ldots,d\text{-}1\} & — & \{1,\ldots,d\} \\ & / & & / & & / & \ldots & & / & \\ 2 & & 3 & & 4 & & \ldots & d & & \end{array}$$

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

Polynomials

f polynomial of degree $p \Rightarrow \text{rank}_\alpha(\mathbf{v}) = \dim \mathcal{M}_\alpha(\mathbf{v}) \le p+1$.

hp method, i.e., piecewise polynomial

Singularity at $x = 0$, partition:

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

Local polynomials of degree $p \Rightarrow \text{rank}_\alpha(\mathbf{v}) = \dim \mathcal{M}_\alpha(\mathbf{v}) \le p+2$.

Conclusion: If any hp approximation with a piecewise polynomial $\mathbf{P}$ of degree $\le p$ exists, then the tensorised grid function $\mathbf{f}$ can be approximated by a tensor $\tilde{\mathbf{f}}$ such that the ranks are bounded by $p+2$ and

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

The data size is bounded by

$$\le 2d\,(p+2)^2\,.$$

The computation of $\tilde{\mathbf{f}}$ is completely black-box (e.g., no information about the location of the singularity required).

Error analysis for asymptotically smooth functions

Functions f satisfying

$$|f^{(k)}(x)| \le C\, k!\, x^{-k-a} \quad \text{for all } k \in \mathbb{N},\ 0 < x \le 1 \text{ and some } a > 0.$$

are called *asymptotically smooth* in $(0, 1]$.

For any $\xi \in (0, 1]$ there is a polynomial p of degree N such that

$$\|f - p\|_{[\xi/2,\xi],\infty} = \max_{\xi/2 \le x \le \xi} |f(x) - p(x)| \le \varepsilon_{N,\xi} := \frac{C}{2} \left(\frac{\xi}{4}\right)^{-a} 3^{-a-N}.$$

Proof. Choose p as Taylor approximation of degree N.

Convolution of tensorised vectors is possible.

With the suitable interpretation,

$$\left(\bigotimes_{j=1}^{d} v^{(j)}\right) \star \left(\bigotimes_{j=1}^{d} w^{(j)}\right) = \bigotimes_{j=1}^{d} \left(v^{(j)} \star w^{(j)}\right)$$
$$v^{(j)}, w^{(j)} \in \mathbb{K}^2,$$

is correct.

Literature:

W. Hackbusch: Tensor spaces and numerical tensor calculus. Springer 2012

Program for tomorrow's lecture:

I. ALS Method for Optimisation Problems

— Formulation of the Problem
— Study of Examples
— Global Convergence for Rank-1 Approximation

II. (Non-)Closedness Questions

— r-Term Format, Rank of a Tensor
— Properties of $\mathcal{R}_r$, Numerical Instability
— Strassen's Matrix Multiplication
— Matrix-Product (TT) Format, Tensor Networks
— Nonclosedness of the Cyclic Matrix-Product Format

Minimal Subspaces

— Definition
— Tensor Spaces of Linear Mappings, Functionals
— Characterisation of Minimal Subspaces in Infinite Dimensions

Topological Tensor Spaces

— Banach spaces, Crossnorms, Projective and Injective Norm
— Final Proof