Hierarchical Tensor Representations

R. Schneider (TUB Matheon)

Paris 2014

Acknowledgment

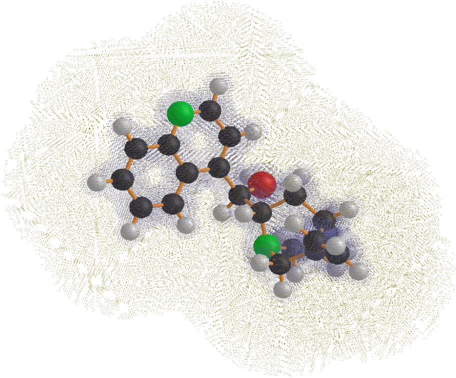
DFG Priority program SPP 1324

Extraction of essential information from complex data

Co-workers: T. Rohwedder (HUB), A. Uschmajev (EPFL Laussanne)
W. Hackbusch, B. Khoromskij, M. Espig (MPI Leipzig), I. Oseledets (Moscow) C. Lubich (Tübingen), O. Legeza (Wigner I - Budapest), Vandereycken (Princeton), M. Bachmayr, L. Grasedyck (RWTH Aachen), ...
J. Eisert (FU Berlin - Physics), F. Verstraete (U Wien), Z. Stojanac, H. Rauhhut
Students: M. Pfeffer, S. Holtz ...

I.

High-dimensional problems



PDE's in $\mathbb{R}^d$, ($d >> 3$)

Equations describing complex systems with multi-variate solution spaces, e.g.

- ▷ stationary/instationary Schrödinger type equations

$$i\hbar\frac{\partial}{\partial t}\Psi(t,\mathbf{x}) = \underbrace{(-\frac{1}{2}\Delta + V)}_{H}\Psi(t,\mathbf{x}), \quad H\Psi(\mathbf{x}) \ = \ E\Psi(\mathbf{x})$$

  describing quantum-mechanical many particle systems

- ▷ stochastic SDEs and the Fokker-Planck equation,

$$\frac{\partial p(t,\mathbf{x})}{\partial t} = \sum_{i=1}^{d}\frac{\partial}{\partial x_i}\left(f_i(t,\mathbf{x})p(t,x)\right) + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}\left(B_{i,j}(t,\mathbf{x})p(t,\mathbf{x})\right)$$

  describing mechanical systems in stochastic environment,

$$\boxed{\mathbf{x} \ = (x_1, \ldots, x_d), \text{ where usually, } d >> 3!}$$

- ▷ parametric PDEs (arising in uncertainty quantification)

$$\text{e.g.} \ \ \nabla_x a(x, y_1, \ldots, y_d)\nabla_x u(x, y_1, \ldots, y_d) = f(x)$$

$$x \in \Omega \ , \ \mathbf{y} \in \mathbb{R}^d \ , \ \ + \text{b.c. on } \partial\Omega \ .$$

Quantum physics - Fermions

For a (discs.) Hamilton operator $\mathbf{H}$ and given $h_p^q, g_{p,q}^{r,s} \in \mathbb{R}$,

$$\mathbf{H} = \sum_{p,q=1}^{d} h_p^q \mathbf{a}_p^T \mathbf{a}_q + \sum_{p,q,r,s=1}^{d} g_{r,s}^{p,q} \mathbf{a}_r^T \mathbf{a}_s^T \mathbf{a}_p \mathbf{a}_q \ .$$

the stationary (discrete) Schrödinger equation is

$$\boxed{\mathbf{H}\mathbf{U} = E\mathbf{U} \ , \quad \mathbf{U} \in \bigotimes_{j=1}^{d} \mathbb{C}^2 \simeq \mathbb{C}^{(2^d)} \ .}$$

where $A := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $A^T = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ $S := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$,

and discrete annihilation operators

$$\boxed{a_p \simeq \mathbf{a}_p := S \otimes \ldots \otimes S \otimes A_{(p)} \otimes I \otimes \ldots \otimes I}$$

and creation operators

$$\boxed{a_p^\dagger \simeq \mathbf{a}_p^T := S \otimes \ldots \otimes S \otimes A_{(p)}^T \otimes I \otimes \ldots \otimes I}$$

Curse of dimensions

For simplicity of presentation: discrete tensor product spaces

$$\mathcal{H} = \mathcal{H}_d \; := \; \bigotimes_{i=1}^d V_i, \qquad \text{e.g.:} \quad \mathcal{V} \; = \; \bigotimes_{i=1}^d \mathbb{R}^{n_i} \; = \; \mathbb{R}^{(\Pi_{i=1}^d n_i)}$$

we consider tensors as multi-index arrays ($\mathcal{I}_i = 1, \ldots, n_i$)

$$U = \left(U_{x_1, x_2, \ldots, x_d}\right)_{x_i = 1, \ldots, n_i \;,\; i=1, \ldots, d} \in \mathcal{V} \;\; ,$$

or equivalently <u>functions of discrete variables</u> ($\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$)

$$U : \times_{i=1}^d \mathcal{I}_i \to \mathbb{K} \;\; , \; \mathbf{x} = (x_1, \ldots, x_d) \mapsto U = U[x_1, \ldots, x_d] \in \mathcal{H} \; ,$$

$d = 1$: n-tuples $(U_x)_{x=1}^n$, or $x \mapsto U[x]$, or $d = 2$: matrices $\left(U_{x,y}\right)$ or $(x, y) \mapsto U[x, y]$.

If not specified otherwise, $\|.\| = \sqrt{\langle ., . \rangle}$ denotes the ℓ_2- norm.

$$\boxed{\dim \mathcal{H}_d = \mathcal{O}(n^d) \quad -- \;\; \text{Curse of dimensionality!}}$$

e.g. $n = 100, d = 10 \rightsquigarrow 100^{10}$ basis functions,
$\rightsquigarrow$ coefficient vectors of 800×10^{18} Bytes $=$ 800 Exabytes
$n = 2$, $d = 500$: then $2^{500} >>$ the estimated number of atoms in the universe!

Setting - Tensors of order d

Goal: Problems posed on tensor spaces,

$$\mathcal{H} \;:=\; \bigotimes_{i=1}^{d} V_i, \qquad \text{e.g.:} \quad \mathcal{H} \;=\; \bigotimes_{i=1}^{d} \mathbb{R}^n \;=\; \mathbb{R}^{(n^d)}$$

Notation: $\mathbf{x} = (x_1, \dots, x_d) \mapsto U = U[x_1, \dots, x_d] \in \mathcal{H}$ For simplicity we will consider only the Hilbert spaces $\ell_2(\mathcal{I})$!

Main problem:

$$\boxed{\dim \mathcal{V} = \mathcal{O}(n^d) \quad -- \;\; \text{Curse of dimensionality!}}$$

e.g. $n = 100, d = 10 \rightsquigarrow 100^{10}$ basis functions,
$\rightsquigarrow$ coefficient vectors of 800×10^{18} Bytes $=$ 800 Exabytes

Approach: Some higher order tensors can be constructed (data-) sparsely from lower order quantities.

As for matrices, incomplete SVD:

$$A[x_1, x_2] \approx \sum_{k=1}^{r} \sigma_k \big(u_k[x_1] \otimes v_k[x_2] \big)$$

Setting - Tensors of order d

Goal: Problems posed on tensor spaces,

$$\mathcal{H} := \bigotimes_{i=1}^{d} V_i, \qquad \text{e.g.:} \quad \mathcal{H} = \bigotimes_{i=1}^{d} \mathbb{R}^n = \mathbb{R}^{(n^d)}$$

Notation: $\mathbf{x} = (x_1, \ldots, x_d) \mapsto U = U[x_1, \ldots, x_d] \in \mathcal{H}$ For simplicity we will consider only the Hilbert spaces $\ell_2(\mathcal{I})$!

Main problem:

$$\boxed{\dim \mathcal{V} = \mathcal{O}(n^d) \quad -- \quad \text{Curse of dimensionality!}}$$

e.g. $n = 100, d = 10 \rightsquigarrow 100^{10}$ basis functions,

$\rightsquigarrow$ coefficient vectors of 800×10^{18} Bytes $=$ 800 Exabytes

Approach: Some higher order tensors can be constructed (data-) sparsely from lower order quantities.

As for matrices, incomplete SVD:

$$A[x_1, x_2] \approx \sum_{k=1}^{r} \sigma_k \big(u_k[x_1] \otimes v_k[x_2]\big)$$

Setting - Tensors of order d

Goal: Problems posed on tensor spaces,

$$\mathcal{H} := \bigotimes_{i=1}^{d} V_i, \qquad \text{e.g.:} \quad \mathcal{H} = \bigotimes_{i=1}^{d} \mathbb{R}^n = \mathbb{R}^{(n^d)}$$

Notation: $\mathbf{x} = (x_1, \ldots, x_d) \mapsto U = U[x_1, \ldots, x_d] \in \mathcal{H}$ For simplicity we will consider only the Hilbert spaces $\ell_2(\mathcal{I})$!

Main problem:

$$\boxed{\dim \mathcal{V} = \mathcal{O}(n^d) \quad -- \;\; \text{Curse of dimensionality!}}$$

e.g. $n = 100, d = 10 \rightsquigarrow 100^{10}$ basis functions,
$\rightsquigarrow$ coefficient vectors of 800×10^{18} Bytes $=$ 800 Exabytes

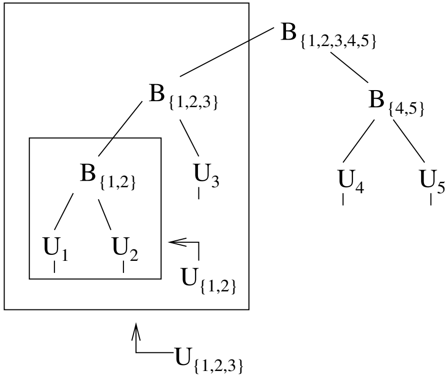
Approach: Some higher order tensors can be constructed (data-) sparsely from lower order quantities.

$\rightsquigarrow$ **Canonical decomposition** for order-d-tensors:

$$U[x_1, \ldots, x_d] \approx \sum_{k=1}^{r} \big(\otimes_{i=1}^{d} u^i[x_i, k] \big).$$

I.

Subspace approximation and novel tensor formats



(Format $\approx$ representation closed under linear algebra manipulations)

Subspace approximation $d = 2$

Let $F : \mathcal{K} \to V$, $y \mapsto F_y \in V$ and $\mathcal{K}$ be compact. (Provided it make sense,) the Kolmogorov r-width is

$$
\begin{aligned}
d_{r,\infty}(F) &:= \inf_{\{U : \dim U \leq r, U \subset V\}} \sup_{y \in \mathcal{K}} \inf_{f_y \in U} \| F_y - f_y \| \\
d_{r,2}(F) &:= \inf_{\{U : \dim U \leq r, U \subset V\}} \Big(\int_{\mathcal{K}} \inf_{f_y \in U} \| F_y - f_y \|^2 dy \Big)^{\frac{1}{2}}
\end{aligned}
$$

Theorem (E. Schmidt (07))

$V := \mathbb{R}^{n_1}$, $\mathcal{K} := \{1 \ldots, n_2\}$, $(x, y) \to F_y(x) := \mathbf{U}[x, y] \in \mathbb{R}^{n_1 \times n_2}$, then the best approximation in the library of all subspaces of dimension at most r is provided by the singular value decomposition (SVD, Schmidt decomposition) and

$$
d_{r,2}(F) = \inf_{\{\mathbf{V} \in U_1 \otimes U_2 : U_1 \subset \mathbb{R}^{n_1}, U_2 \subset \mathbb{R}^{n_2} \ ; \ dim\ U_1 \leq r\}} \| \mathbf{U} - \mathbf{V} \|
$$

Tucker decomposition - sub-space approximation

We are seeking subspaces $U_i \subset V_i$ fitting best to a given tensor $X \in \bigotimes_{i=1}^{d} V_i$, in the sense

$$\boxed{\|X - U\|^2 := \inf_{\{V \in U_1 \otimes \cdots \otimes U_d \,:\, \dim\, U_i \leq r_i\}} \|X - V\|^2}$$

i.e we are minimizing over subspaces $U_i \in \mathcal{G}(V_i, r_i)$,

$$\mathcal{G}(V, r) := \{U \subset V \text{ subspace } : \dim U = r\} \quad \text{Grasmannian}$$

$$U_i = \text{ span } \{\mathbf{b}^i_{k_i} : k_i = r_i\} \subset V_i \quad , \quad \text{rank tuple } \mathbf{r} = (r_1, \dots, r_d) \,.$$

$$\Rightarrow \quad C[k_1, \dots, k_d] = \langle U, \mathbf{b}^1_{k_1} \otimes \cdots \otimes \mathbf{b}^d_{k_d} \rangle \quad \text{core tensor}$$

$$U[x_1, .., x_d] = \sum_{k_1=1}^{r_1} \dots \sum_{k_d=1}^{r_d} C[k_1, .., k_d] \bigotimes_{i=1}^{d} \mathbf{b}^i_{k_i}[x_i]$$

Subspace approximation

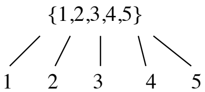
Subspace approximation

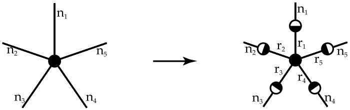
- Tucker format (MCSCF, MCTDH(F)) - robust
  But complexity $\mathcal{O}(r^d + ndr)$

  Is there a robust tensor format, but polynomial in d?

Univariate bases $x_i \mapsto \left(U_i[k_i, x_i]\right)_{k_i=1}^{r_i}$ ($\to$ Graßmann man.)

$$U[x_1, .., x_d] = \sum_{k_1=1}^{r_1} \dots \sum_{k_d=1}^{r_d} B[k_1, .., k_d] \bigotimes_{i=1}^{d} \mathbf{U}^i[k_i, x_i]$$





Subspace approximation

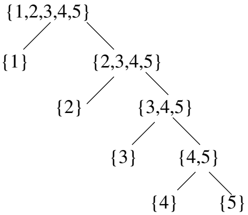
- ▷ Tucker format (MCSCF, MCTDH(F)) - robust
  But complexity $\mathcal{O}(r^d + ndr)$

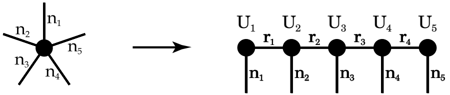
  Is there a robust tensor format, but polynomial in d?

- ▷ Hierarchical Tucker format
  (HT; Hackbusch/Kühn, Grasedyck, Meyer et al., Thoss & Wang, Tree-tensor networks)

- ▷ Tensor Train (TT-)format $\simeq$ Matrix product states (MPS)

$$U[\mathbf{x}] = \sum_{k_1=1}^{r_1} \dots \sum_{k_{d-1}=1}^{r_{d-1}} \prod_{i=1}^{d} B^i[k_{i-1}, x_i, k_i] = \mathbf{B}_1[x_1] \cdots \mathbf{B}_d[x_d]$$





Hierarchical tensor (HT) format

- ▷ Canonical decomposition
- ▷ Subspace approach (Hackbusch/Kühn, 2009)

(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n \times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t_1} \times k_{t_2}}$)

Hierarchical tensor (HT) format

- ▷ Canonical decomposition not closed, no embedded manifold!
- ▷ Subspace approach (Hackbusch/Kühn, 2009)

(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n \times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t_1} \times k_{t_2}}$)

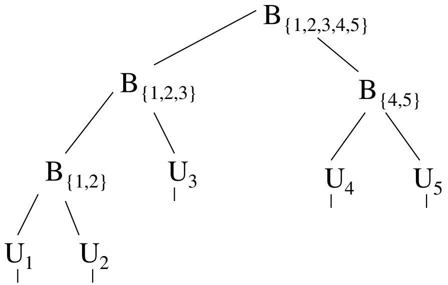
Hierarchical tensor (HT) format

- Canonical decomposition not closed, no embedded manifold!
- Subspace approach (Hackbusch/Kühn, 2009)

(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n\times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t\times k_{t_1}\times k_{t_2}}$)

Hierarchical tensor (HT) format

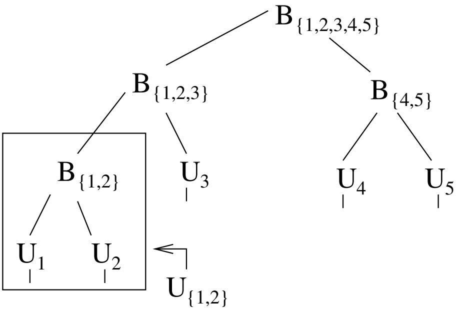
- ▷ Canonical decomposition   not closed, no embedded manifold!
- ▷ Subspace approach (Hackbusch/Kühn, 2009)



(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n \times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t_1} \times k_{t_2}}$)

Hierarchical tensor (HT) format

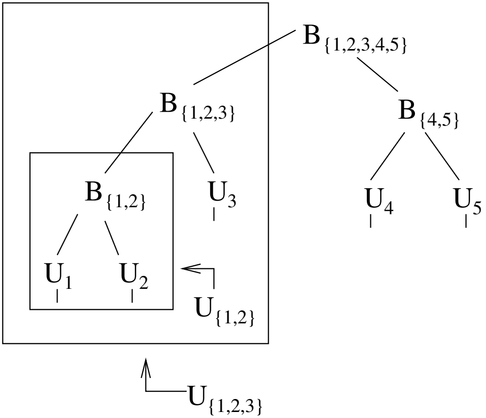
- ▷ Canonical decomposition **not closed, no embedded manifold!**
- ▷ Subspace approach (Hackbusch/Kühn, 2009)



(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n \times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t_1} \times k_{t_2}}$)

Hierarchical tensor (HT) format

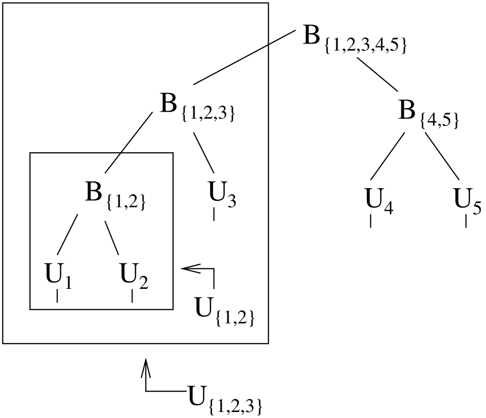
- ▷ Canonical decomposition   not closed, no embedded manifold!
- ▷ Subspace approach (Hackbusch/Kühn, 2009)



(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n\times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t_1} \times k_{t_2}}$)

Hierarchical tensor (HT) format

- ▷ Canonical decomposition   not closed, no embedded manifold!
- ▷ Subspace approach (Hackbusch/Kühn, 2009)

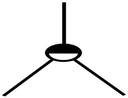


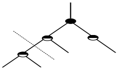
(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n \times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t_1} \times k_{t_2}}$)

Recursive definition by bases representations

$$U^{\alpha} \quad = \quad span\{\mathbf{b}_i^{(\alpha)} : 1 \leq i r_{\alpha}\}$$

$$\mathrm{b}_{\ell}^{(\alpha)} = \sum_{i=1}^{r_{\alpha_1}} \sum_{j=1}^{r_{\alpha_2}} \mathbf{c}^{\alpha}[i, j\ell]\, \mathbf{b}_i^{(\alpha_1)} \otimes \mathbf{b}_j^{(\alpha_2)} \qquad (\alpha_1, \alpha_2 \text{ sons of } \alpha \in T_D).$$





The tensor is recursively defined by the transfer or component tensors $(\ell, i, j) \mapsto \mathbf{c}^{\alpha}[i, j, \ell]$ in $\mathbb{R}^{k_t \times k_1 \times k_2}$.

$$U[\mathbf{x}] = \sum_{k_{\alpha} : \alpha \in \mathbb{T}} \bigotimes_{\alpha \in \mathbb{T}} \mathbf{c}^{\alpha}[k_{s_1(\alpha)}, k_{s_2(\alpha)}, k_{\alpha}]$$

(with obvious modifications for $\alpha = D$ or α is a leave.)

Data complexity $\boxed{\mathcal{O}(dr^3 + dnr)}$! $(r := \max\{r_{\alpha}\})$

TT - Tensors - Matrix product representation

Noteable special case of HT:

TT format (Oseledets & Tyrtyshnikov, 2009)
$\simeq$ matrix product states (MPS) in quantum physics Affleck, Kennedy, Lieb &Tagasaki (87)., Römmer & Ostlund (94), Vidal (03),
HT $\simeq$ tree tensor network states in quantum physics (Cirac, Verstraete, Eisert .....)

TT tensor U can be written as matrix product form

$$U[\mathbf{x}] = \mathbf{U}_1[x_1] \cdots \mathbf{U}_i[x_i] \cdots \mathbf{U}_d[x_d]$$

$$= \sum_{k_1=1}^{r_1} .. \sum_{k_{d-1}=1}^{r_{d-1}} U_1[x_1, k_1] U_2[k_1, x_2, k_2] ... U_{d-1}[k_{d-2} x_{d-1}, k_{d-1}] U_d[k_{d-1}, x_d, k_d]$$

with matrices or component functions

$$\mathbf{U}_i[x_i] = \left(u_{k_{i-1},k_i}[x_i] \right) \in \mathbb{R}^{r_{i-1} \times r_i} \ , \ r_0 = r_d := 1 \ .$$

Redundancy: $U[\mathbf{x}6 = \mathbf{U}_1[x_1] \mathbf{G}\mathbf{G}^{-1} \mathbf{U}_2[x_2] \cdots \mathbf{U}_i[x_i] \cdots \mathbf{U}_d[x_d]$.

HSVD - hierarchical (and high order) SVD

- Vidal (2003), Oseledets (2009), Grasedyck (2009), Kühn (2012)

Matricisation or unfolding

$$(x_1, \ldots, x_d) \mapsto \mathbf{A}_{(x_1),(x_2,\ldots,x_d)} = U[\mathbf{x}] \quad \in V_1 \otimes V_2^* \otimes \cdots V_d^*$$

The tensor $\mathbf{x} \to U[\mathbf{x}]$

$$U[x_1, \ldots, x_d] = \mathbf{U}_1[x_1] \cdots \mathbf{U}_i[x_i] \cdots \mathbf{U}_d[x_d]$$

$$= \sum_{k_1=1}^{r_1} \ldots \sum_{k_{d-1}=1}^{r_{d-1}} U_1[x_1, k_1] U_2[k_1, x_2, k_2] \ldots U_{d-1}[k_{d-2} x_{d-1}, k_{d-1}] U_d[k_{d-1}, x_d]$$

with matrices or component functions

$$\mathbf{U}_i[x_i] = (U_i[k_{i-1}, x_i, k_i]) \in \mathbb{R}^{r_{i-1} \times r_i} \ , \ r_0 = r_d := 1 \ .$$

Hard thresholding $\mathbf{H_s}(U)$: $s_1 \leq r_1$; truncate the above sums after s_1.

HSVD - hierarchicalSVD

- Vidal (2003), Oseledets (2009), Grasedyck (2009), Kühn (2012)

Matricisation or unfolding

$$(x_1, \ldots, x_d) \mapsto \mathbf{A}_{(x_1,x_2),(x_3,\ldots,x_d)} = U[\mathbf{x}] \quad \in V_1 \otimes V_2 \otimes V_3^* \otimes \cdots V_d^*$$

The tensor $\mathbf{x} \to U[\mathbf{x}]$

$$U[x_1, \ldots, x_d] = \mathbf{U}_1[x_1] \cdots \mathbf{U}_i[x_i] \cdots \mathbf{U}_d[x_d]$$

$$= \sum_{k_1=1}^{r_1} \cdots \sum_{k_{d-1}=1}^{r_{d-1}} U_1[x_1, k_1] U_2[k_1, x_2, k_2] \ldots U_{d-1}[k_{d-2} x_{d-1}, k_{d-1}] U_d[k_{d-1}, x_d]$$

with matrices or component functions

$$\mathbf{U}_i[x_i] = (U_i[k_{i-1}, x_i, k_i]) \in \mathbb{R}^{r_{i-1} \times r_i} \ , \ r_0 = r_d := 1 \ .$$

Hard thresholding $\mathbf{H_s}(U)$: $s_i \leq r_i$; truncate the above sums after s_i, $i = 1, \ldots, d-1$.

HSVD - hierarchical (and high order) SVD

- Vidal (2003), Oseledets (2009), Grasedyck (2009), Kühn (2012)

Matricisation or unfolding

$$(x_1, \ldots, x_d) \mapsto \mathbf{A}_{(x_1 \ldots, x_{d-1}),(x_d)} = U[\mathbf{x}] \quad \in V_1 \otimes \cdots V_{d-1} \otimes V_d^*$$

The tensor $\mathbf{x} \to U[\mathbf{x}]$

$$U[x_1, \ldots, x_d] = \mathbf{U}_1[x_1] \cdots \mathbf{U}_i[x_i] \cdots \mathbf{U}_d[x_d]$$

$$= \sum_{k_1=1}^{r_1} \ldots \sum_{k_{d-1}=1}^{r_{d-1}} U_1[x_1, k_1] U_2[k_1, x_2, k_2] \ldots U_{d-1}[k_{d-2} x_{d-1}, k_{d-1}] U_d[k_{d-1}, x_d]$$

with matrices or component functions

$$\mathbf{U}_i[x_i] = (U_i[k_{i-1}, x_i, k_i]) \in \mathbb{R}^{r_{i-1} \times r_i} \ , \ r_0 = r_d := 1 \ .$$

Data Complexity: $\mathcal{O}\big(ndr^2\big)$, $r = \max\{r_i : i = 1, \ldots, d-1\}$,

Complexity of HSVD

Let us assume that

$$
\begin{aligned}
& U[x_1, \ldots, x_d] \\
= \; & \sum_{k_1=1}^{R_1} \ldots \sum_{k_{d-1}=1}^{R_{d-1}} \tilde{U}_1[x_1, k_1] \cdots \tilde{U}_{d-1}[k_{d-2}, x_{d-1}, k_{d-1}] U_d[k_{d-1}, x_d]
\end{aligned}
$$

For $i = 1, \ldots, d-1$ compute,

$$
\overline{U}_i[k_{i-1}, x_i, \tilde{k}_i] := \sum_{\tilde{k}_{i-1}=1}^{R_i} V_{i-1}[k_{i-1}, \tilde{k}_{i-1}] U_i[\tilde{k}_{i-1}, x_i, k_i]
$$

we decompose

$$
\overline{U}_i[k_{i-1}, n_i, \tilde{k}_i] \quad = \quad \sum_{k_i=1}^{r_i} U_i[k_{i-1}, n_i, k_i] V_i[k_i, \tilde{k}_i]
$$

Computational costs are $\mathcal{O}(dn^2 r^2 R^2)$

Example

Any canonical representation with r terms

$$\sum_{k=1}^{r} U_1(x_1, k) \cdots U_d(x_d, k)$$

is also TT with ranks $r_i \leq r$, $i = 1, \ldots, d-1$.
But conversely canonical r term representation is bounded by $r_1 \times \cdots \times r_{d-1} = \mathcal{O}(r^{d-1})$
Hierarchical ranks could be much smaller than canonical rank.
Example $x_i \in [-1, 1]$, $i = 1, \ldots, d$, i.e $r = d$,

$$U(x_1, \ldots, x_d) = \sum_{i=1}^{d} x_d = x_1 \otimes I \cdots + I \otimes x_2 \otimes I \otimes \cdots \ ,$$

but

$$U(x_1, \ldots, x_d) = (1, x_1) \begin{pmatrix} 1 & x_2 \\ 0 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & x_{d-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_d \\ 1 \end{pmatrix}$$

here $r_1 = \ldots = r_{d-1} = 2$.

Fundamental properties of HT

Redundancy: we explain TT as model example

$$U[\mathbf{x}] = \mathbf{U}_1[x_1]\mathbf{G}_1\mathbf{G}_1^{-1}\mathbf{U}_2[x_2]\mathbf{G}_2\mathbf{G}_2^{-1}\cdots\mathbf{U}_i[x_i]\cdots\mathbf{U}_d[x_d] \; .$$

Given a linear parameter space X and groups G_i

$$X := \times_{i=1}^{d} X_i = \times_{i=1}^{d} (\mathbb{R}^{r_{i-1}n_i r_i}) \;\;,\;\; \mathcal{G}_{\mathbf{r}} := \times_{i=1}^{d-1} G_i = \times_{i=1}^{d-1} GL(\mathbb{R}^{r_i})$$

Lie group action

$$G_i U_i := \mathbf{G}_{i-1}^{-1}\mathbf{U}_i(x_i)\mathbf{G}_i \; , \; i = 1, \ldots, d, \;\; U_i \in X_i \; .$$

$$\underline{U} \sim \underline{V} \;\; \Leftrightarrow \underline{U} = \underline{G}\,\underline{V} \; , \underline{G} \in \mathcal{G}_{\underline{r}}$$

defines a manifold

$$\boxed{\mathcal{M}_{\underline{r}} \approx \left(\times_{i=1}^{d} X_i \right) / \mathcal{G}_{\underline{r}}}$$

Then tangent space $\mathcal{T}_U$ at U is given by

$$\begin{aligned} \delta U &= \delta U_1 + \ldots + \delta U_d \\ &= \delta\mathbf{U}_1 \circ \mathbf{U}_2 \cdots \mathbf{U}_d + \ldots + \mathbf{U}_1 \cdots \circ \delta\mathbf{U}_d \end{aligned}$$

where $\delta\mathbf{U}_i \;\; \perp \;\; \text{span}\, \mathbf{U}_i$.

Fundamental properties of HT (particularly TT)

Grouping indices at $t \in \mathbb{T}$, ($D \in \mathbb{T}$ is the root)

$$t := \{i_1, \dots, i_l\} \subset D := \{1, \dots, d\} \quad , \mathcal{I}_t = \{x_{i_1}, \dots, x_{i_l}\}$$

into row or column index of $\mathbf{U}_t = \mathbf{U}_t(U) = \left(\mathbf{U}_{\mathcal{I}_t, \mathcal{I}_D \setminus \mathcal{I}_t} \right) \Rightarrow$ matricisation or unfolding of

$$(x_1, \dots, x_d) \mapsto U[x_1, \dots, x_d] \simeq \mathbf{U}_{\mathcal{I}_t, \mathcal{I}_D \setminus \mathcal{I}_t} \quad \Rightarrow \quad r_t = \text{rank } \mathbf{U}_t(U)$$

e.g. TT format $r_i = \text{rank } \mathbf{U}_{x_1, \dots, x_i}^{x_{i+1}, \dots, x_d}$.

- There exist a well defined rank tuple $\mathbf{r} := (r_t)_{t \in \mathbb{T}}$, e.g. $\mathbf{r} = (r_1, \dots, r_{d-1})$ for TT
- $\mathcal{M}_{\mathbf{r}} = \{U \in \mathcal{H} : r_t = rank\ \mathbf{U}_t, t \in \mathbb{T}\}$ is analytic manifold
- $$\boxed{\mathcal{M}_{\underline{r}} \approx \left(\times_{i=1}^{d} X_i \right) / \mathcal{G}_{\underline{r}}}$$

  $$\boxed{\mathcal{M}_{\leq \underline{r}} = \bigcup_{s_i \leq r_i} \mathcal{M}_{\underline{s}} = \overline{\mathcal{M}_{\underline{r}}} \subset \mathcal{H} \text{ is (weakly) closed!}}$$ Hackbusch & Falco
- $$\mathcal{M}_{\leq \mathbf{r}} \text{ is a an algebraic variety.}$$

Table: Some comparison

	canonical	Tucker	HT
complexity	$\mathcal{O}(ndr)$	$\mathcal{O}(r^d + ndr)$	$\mathcal{O}(ndr + dr^3)$ TT- $\mathcal{O}(ndr^2)$
	++	$-$	+
rank	no	defined	defined
	$r_c \geq$	r_T	$r_T \leq r_{HT} \leq r_c$
(weak) closedness	no	yes	yes
essential redundancy	yes	no	no
embedded manifold	no	yes	yes
dyn. low rank approx.	no	yes	yes
recovery	??	yes	yes
quasi best approx.	no	yes	yes
best approx.	no	exist but NP hard	exist but NP hard

$\mathcal{M}_{\leq r}$ is an algebraic variety?! Not included here are general tensor

Convergence rates w.r.t. ranks for HT (TT)

$$\text{Let } \mathbf{A}_t = \mathbf{U}^T \Sigma \mathbf{V} \,, (SVD)\ \Sigma = \mathrm{diag}(\sigma_i)$$

For $0 < p \leq 2$, $s := \frac{1}{p} - \frac{1}{2}$, (e.g. Nuclear norm $p = 1$)

$$\|\mathbf{A}_t\|_{*,p} := \Big(\sum_i \sigma_{t,i}^p \Big)^{\frac{1}{p}} \,,$$

then the best rank k approximation satisfies

$$\inf_{\text{rank } \mathbf{V} \leq k} \|\mathbf{A}_t - \mathbf{V}\|_2 \lesssim k^{-s} \|\mathbf{A}_t\|_{*,p}$$

Theorem (Uschmajev & S. (2013))

Assume $\|\mathbf{A}\|_{*,p} := \max_t \|\mathbf{A}_t\|_{*,p} < \infty$, *and* $|\underline{r}| := \max\{r_t\}$, *then*

$$\inf_{\textit{rank } V \leq \underline{r}} \|U - V\|_2 \lesssim C(d) |\underline{r}|^{-s} \|\mathbf{A}\|_{*,p} \ \ \textit{with } C(d) \lesssim \sqrt{d} \,,$$

Mixed Sobolev spaces $H^{t,mix} \subset L_{*,p}$, $p = \frac{2}{4t+1}$, $\Rightarrow\ s = 2t$

Historical comparison of related topics

Principal ideas of hierarchical tensors have been invented several times:

1. Statistics: Hidden Markov Models (60s) ???
2. Condensed matter physics: Block Renormalization and renormalization group (70s)
3. Spin systems (AKLT 87)
4. Quantum lattice systems: DMRG White (91) and Ostlund & Rommer (94)
5. Finitely correlated states: Fannes, Nachtergale & Werner (92)
6. Molecular quantum dynamics: Meyer, (Cederbaum) et al. (2001)
7. Quantum computing: Vidal, Cirac, Verstraete (2003)
8. Hackbusch & Kühn (HT) (2009)
9. Oseledets & Tyrtyshnikov (TT) (2009)

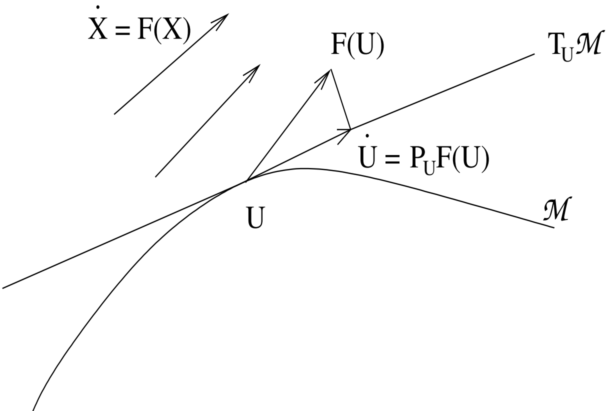
see e.g.

Contributions about hierarchical tensors

- HT - Hackbusch & Kühn (2009), TT - Oseledets & Tyrtyshnikov (2009)
- MPS- Affleck et al. AKLT (Affleck, Kennnedy, Lieb, Takesaki 1987), Fannes, Nachtergale & Werner (92), DMRG- S: White (91),
- HOSVD-Laathawer et.al. (2001), HSVD Vidal (2003), Oseledets (09), Grasedyck (2010), Kühn (2012)
- Riemannian optimization - Absil et al. (2008), Lubich, Koch, Conte, Rohwedder, S. Uschmajew, Vandereycken, Kressner, Steinlechner, Arnold & Jahnke, ...
- Oseledets, Khoromskij, Savostyanov, Dolgov, Kazeev, ...
- Grasedyck, Ballani, Bachmayr, Dahmen, ...
- Falco, Nouy, Ehrlacher ....
- Physics: Cirac, Verstraete, Schollwöck, Legeza, G. Chan, Eisert, ...

II.

How to compute with hierarchical tensors



Computation in hierarchical tensor format - HT arithmetics

Given a tree $\mathbb{T}$, all tensors $U, V \in \mathcal{M}_{\mathbf{s}}$ for some multilinear rank $\mathbf{s}$. Then

1. $U + V \in \mathcal{M}_{\leq 2\mathbf{s}}$
2. $\mathbf{x} \mapsto U[\mathbf{x}]V[\mathbf{x}] \in \mathcal{M}_{\leq \mathbf{s}^2}$ Hadamard product
3. $\langle U, V \rangle$ can be performed in $\mathcal{O}(ndr^2 + dr^4)$ resp. $\mathcal{O}(ndr^3)$ (TT) arithmetic operations
4. operators $\mathcal{A} : \mathcal{H} \to \mathcal{H}$ may be written in canonical, TT or HT format.
5. Assumption $\mathcal{A}U$ is accessible as a rank $\mathbf{S}$ HT tensor.

Remark: E.g. $A + U$ can be recovered into a standard form by HSVD, or approximated.

Optimization Problems

Problem (Generic optimization problem (OP))

Given a cost functional $\mathcal{J} : \mathcal{H} \to \mathbb{R}$ *and an admissible set* $\mathcal{A} \subset \mathcal{H}$ *finding*

$$\mathit{argmin}\,\{\mathcal{J}(W) : W \in \mathcal{A}\}\,.$$

Working framework Fixed the model class - find the best or quasi-optimal approximate solution in this model class

Problem (Tensor product optimization problem (TOP))

$$U := \mathit{argmin}\,\{\mathcal{J}(W) : W \in \mathcal{M} = \mathcal{A} \cap \mathcal{M}_{\leq \mathbf{r}}\} \tag{1}$$

Admissible set is confined to $\mathcal{M}_{\leq \mathbf{r}}$ - tensors of rank at most $\mathbf{r}$.

WARNING: Hillar & Lim (2011):
Most tensor problems are NP hard if $d \geq 3$.
for example: best rank 1 approximation (multiple local minima).

Optimization Problems

Problem (Generic optimization problem (OP))

Given a cost functional $\mathcal{J} : \mathcal{H} \to \mathbb{R}$ *and an admissible set* $\mathcal{A} \subset \mathcal{H}$ *finding*

$$\textit{argmin}\,\{\mathcal{J}(W) : W \in \mathcal{A}\}\ .$$

Working framework Fixed the model class - find the best or quasi-optimal approximate solution in this model class

Problem (Tensor product optimization problem (TOP))

$$U := \textit{argmin}\,\{\mathcal{J}(W) : W \in \mathcal{M} = \mathcal{A} \cap \mathcal{M}_{\leq \mathbf{r}}\} \tag{2}$$

We have fixed our costs so far. But, in order to achieve a desired accuracy, we must enrich our model class (systematically).

Greedy techniques could be shown to provide to convergence to the exact solution

[Cances, Ehrlacher& Lelievere], [Falco & Nouy] and coworkers. Bachmayr & Dahmen

Example

Espig,Hackbusch, Rohwedder & Schneider (2010)

1. Approximation: for given $U \in \mathcal{H}$ minimize

$$\boxed{\mathcal{J}(W) = \|U - W\|^2 \ , \ W \in \mathcal{M}}$$

2. solving equations: where $A, g : \mathcal{V} \to \mathcal{H}$,

$$AU = B \ \ or \ \ g(U) = 0$$

here

$$\boxed{\mathcal{J}(W) := \|AW - B\|_*^2 \ \text{ resp. } \ F(W) := \|g(W)\|_*^2 \ .}$$

3. or, if $A : \mathcal{V} \to \mathcal{V}'$ is symmetric and $B \in \mathcal{V}'$, $\mathcal{V} \subset \mathcal{H} \subset \mathcal{V}'$,

$$\boxed{\mathcal{J}(W) := \frac{1}{2}\langle AW, W\rangle - \langle B, W\rangle}$$

4. computing the lowest eigenvalue of a symmetric operator $A : \mathcal{V} \to \mathcal{V}'$,

$$\boxed{U = \text{argmin}\,\{\mathcal{J}(W) = \langle AW, W\rangle : \langle W, W\rangle = 1\}\ .}$$

In many cases $\mathcal{A} \cap \mathcal{M}^{\leq \mathbf{r}} = \mathcal{M}^{\leq \mathbf{r}}$.

Hard Thresholding -

Projected Gradient Algorithms: E.G. inimize

$$\mathcal{J}(U) := \frac{1}{2}\langle U\mathcal{A}U\rangle - \langle U, Y\rangle \;\; \nabla J(X) = (\mathcal{A}U - Y)$$

w.r.t. low rank constraints

$$\begin{aligned} V^{n+1} &:= U^n \qquad - \alpha_n(\mathcal{C}^{-1}(\mathcal{A}U^n - Y)) \;\text{ gradient step} \\ U^{n+1} &:= \mathcal{R}_n(V^{n+1}) \;. \end{aligned}$$

$\mathcal{R}_n$ (nonlinear) projection to model class

$$\mathcal{R}_n : \mathbb{R}^{n_1 \times n_2} \to \mathcal{M}_{\mathbf{r}}$$

e.g HSVD $\sigma_s := \sigma_{s_t}$ singular values of $\mathbf{V}_t = \mathbf{V}_t(V^{n+1})$, $t \in \mathbb{T}$,

1. Hard thresholding, $\sigma_s := 0$, $s > r$, $\sigma_s \leftarrow \sigma_s$, $s \leq r$
2. Riemannian techniques including ALS:
3. Soft thresholding, $\sigma_s \leftarrow \max\{\sigma_s - \varepsilon, 0\}$

Hard Thresholding - Riemannian gradient iteration

$$\mathcal{J}(U) := \frac{1}{2}\langle U, \mathcal{A}U\rangle - \langle U, Y\rangle \ , \ \nabla\mathcal{J}(X) = (\mathcal{A}U - Y)$$

$$\begin{aligned} V^{n+1} &:= U^n - P_{\mathcal{T}_U}\alpha_n\big(\mathcal{C}^{-1}(\mathcal{A}U^n - Y)\big) \ \text{projected gradient step} \\ &= U^n + \boldsymbol{\xi}^n \ , \ \mathcal{M}_{\mathbf{r}} + \mathcal{T}_U \\ U^{n+1} &:= \mathcal{R}_n(V^{n+1}) := R(U^n, \boldsymbol{\xi}^n) \ . \end{aligned}$$

$P_{\mathcal{T}_U} : \mathcal{H} \to \mathcal{T}_U$ orthogonal projection onto tangent space at U
retraction (*Absil et al.*) $R(U, \xi) : T_{\mathcal{M}_{\mathbf{r}}} \to \mathcal{M}_{\mathbf{r}}$,

$$R(U, \boldsymbol{\xi}) = U + \boldsymbol{\xi} + \mathcal{O}(\|\boldsymbol{\xi}\|^2)$$

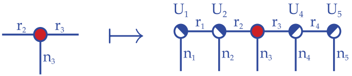
e.g. R is an approximate exponential map

Nonlinear Gauss Seidel local optimization for TT (HT) tensors

Alternating Linear Scheme - ALS

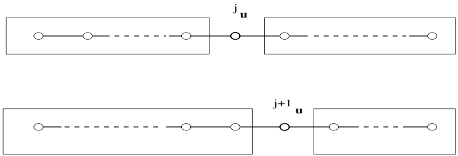
Relaxation (see e.g. Gauss-Seidel, ALS):
For $j = 1, \ldots, d$:

1. fix all component tensors $\mathbf{U}^{\nu}$, $\nu \in \{1, \ldots, d\} \setminus \{j\}$, except index j.

$\mathbf{P}_{i,1,U}$:


2. Optimize $\mathbf{U}^{j}[k_{j-1}, x_j, k_j]$, and orthogonalize left
3. Repeat with $\mathbf{U}^{j+1}$ (the tree is reorder to optimize alway the root!)

Repeat the relaxation procedure (in the opposite direction.)



S. Holtz, Rohwedder & Schneider (2010), Uschmajew & Rohwedder (2011).

ALS (single site DMRG) - Nonlinear Gauß Seidel

Solving: $\mathcal{A}U = B$, ($\mathcal{A}^T = \mathcal{A}$)

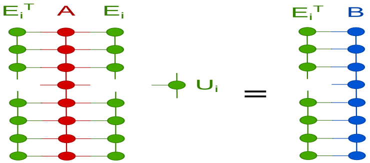
$$U = \operatorname{argmin}\{\frac{1}{2}\langle \mathcal{A}U, U\rangle - \langle B, U\rangle : U \in \mathcal{M}_{\underline{r}}\} \ ,$$

$$U[\mathbf{x}] = \mathbf{U}_1[x_1]\cdots\mathbf{U}_i[x_i]\cdots\mathbf{U}_d[x_d]$$

Optimizing $\mathbf{U}_i$, resp. $U_i[k_{i-1}, x_i, k_i]$ leads to a linear system

$\boxed{\widetilde{\mathcal{A}}_i U_i = \widetilde{B}_i}$, in the small (sub-) space $\mathbb{R}^{r_{i-1}\times n_i \times r_i}$

Renormalization group



$$\widetilde{\mathcal{A}}_i = \sum_\nu \mathbf{L}_{i,\nu}\otimes\mathbf{A}_{i,\nu}\otimes\mathbf{R}_{i,\nu}$$

$$\mathbf{L}_{i,\nu}\in\mathbb{R}^{r_{i-1}\times r_{i-1}}, \mathbf{A}_{i,\nu}\in\mathbb{R}^{n_i\times n_i}\mathbf{R}_{i,\nu}\in\mathbb{R}^{r_i\times r_i}$$

Riemannian gradient iteration - Local Convergence

Theorem (Local convergence of Riemaniann gradient iteration)

Let $V^{n+1} := U^n + \mathcal{C}^{-1}(Y - \mathcal{A}U^n)$*, assume that* $\mathcal{A}$ *is SPD and* $U \in \mathcal{M}_{\mathbf{r}}$*. If*

$$\|U - U^m\| \leq \delta \ , \ \ \delta \sim dist(U, \partial\mathcal{M}_{\mathbf{r}})$$

sufficiently small, and $\delta \sim dist(U, \partial\mathcal{M}_{\mathbf{r}})$*, then, there exist* $0 < \rho < 1$ *s.t the series* $U^n \in \mathcal{M}_{\leq \mathbf{r}}$ *converges linearly to a unique solution* $U \in \mathcal{M}_{\leq \mathbf{r}}$ *with rate* ρ

$$\|U^{n+1} - U\| \leq \rho \|U^n - U\|$$

Remark: Suppose $\|U\| = 1$ then

$$\mathrm{dist}(U, \partial\mathcal{M}_{\mathbf{r}}) \leq \min_{t \in \mathbb{T}, 0 < k \leq r_t} \sigma_{t,k}$$

is smallest (non-zero) singular value of $\mathbf{U}_t(U)$!

Iterative Hard Thresholding - Local Convergence

Theorem (Global convergence of IHT)

Let $V^{n+1} := U^n + \mathcal{C}^{-1}(Y - \mathcal{A}U^n)$, *and* $U^{n+1} = \mathbf{H_r} V^{n+1}$ *assume that*

$$\gamma \|V\|^2 \leq \langle \mathcal{C}^{-1}\mathcal{A}V, V \rangle \leq \Gamma \|V\|^2$$

with e.g. $\frac{\Gamma}{\gamma} < C(d)$ *suff. small.*
Then, there exist $0 < \rho < 1$ *s.t the series* $U^n \in \mathcal{M}_{\leq \mathbf{r}}$ *convergences linearly to a unique tensor* $U_\epsilon \in \mathcal{M}_{\leq \mathbf{r}}$ *with rate* ρ

$$\|U^{n+1} - U_\epsilon\| \leq \rho \|U^n - U_\epsilon\|$$

and U_ϵ *is a quasi-optimal solution*

$$\|U - U_\epsilon\| \leq C inf_{V \in \mathcal{M}_{\mathbf{r}}} \|V - U_\epsilon\|$$

Iterative Hard Thresholding - Local Convergence

Theorem (Global convergence of Riemannian gradient iteration- (ongoing joint work with A. Uschmajew))

Let $V^{n+1} := U^n + \mathcal{C}^{-1}(Y - \mathcal{A}U^n)$, and $\mathcal{A}$ is SPD. Then, the series $U^n \in \mathcal{M}_{\leq \mathbf{r}}$ converges to a stationary point $U \in \mathcal{M}_{\leq \mathbf{r}}$. The same results holds for the Gauß Southwell variant of ALS (1site DMRG).

Lojasiewicz (-Kurtyka) inequality

$$\mathcal{J}(V)^\theta - \mathcal{J}(U)^\theta \leq \Gamma \|\text{grad } \mathcal{J}(V)\| \ , \ 0 < \theta \leq \frac{1}{2} \, , \|U_V\| \leq \delta \, .$$

LK inequality is valid on *algebraic sets*, *o-minimal structures* etc. [*Bolte et al.*]. It is a powerful mathematical tool for proving convergence.

1. $\theta = \frac{1}{2}$: linear convergence $\|U^n - U\| \lesssim q^n \|U^1 - U^0\|$, $q < 1$.
2. $0 < \theta < \frac{1}{2}$: $\|U^n - U\| \lesssim n^{-\frac{\theta}{2-\theta}}$

Low Rank Tensor Recovery - Tensor Completion

Sampling or inerpolation. Given p measurements

$\mathbf{y}[i] := (\mathcal{A}\mathbf{U})_i = U[\mathbf{k}_i]\ ,\ \mathbf{k}_i = (k_{i,1},\ldots,k_{i,d})\ \ i = 1,\ldots,p\,(<< n_1\ \cdots n_d)\ ,$

reconstruct the tensor $U \in \mathcal{H} := \otimes_{i=1}^d \mathbb{R}^{n_i}$

Tensor completion: given values

$$U[\mathbf{k}_i]\ \ ,\ \ i = 1,\ldots p << N = n^d\ .$$

at randomly chosen points $\mathbf{k}_i$,

Can one reconstruct $U \in \mathcal{M}_{\underline{r}}$?

Assumption: $U \in \mathcal{M}_{\mathbf{r}}$ with multi-linear rank $\mathbf{r} = (r_i)_{t\in\mathbb{T}}$. or $U \in \mathcal{M}_{\leq \mathbf{r}}$

E.g. as a prototype example TT-format in matrix product representation, oracle dimension

$$\dim \mathcal{M}_{\mathbf{r}} = \mathcal{O}(ndr^2)\ \Rightarrow\ p = \mathcal{O}(ndr^2 \log^a ndr)\ ?$$

$(n = \max_{i=1,\ldots,d} n_i\ ,\ r = \max_{t\in\mathbb{T}} r_t)$

1. random sampling: joint work with H. Rauhut and Z. Stojanac
2. adaptive sampling: based on max-volume strategies

Iterative Hard Thresholding

Projected Gradient Algorithms: Minimize residual

$$J(U) := \frac{1}{2}\langle \mathcal{A}U - \mathbf{y}, \mathcal{A}U - \mathbf{y}\rangle \quad \nabla J(X) = \mathcal{A}^T(\mathcal{A}U - \mathbf{y})$$

w.r.t. low rank constraints

$$\begin{aligned} Y^{n+1} &:= U^n - \alpha_n\big(\mathcal{A}^T(\mathcal{A}U^n - \mathbf{y})\big) \quad \text{gradient step} \\ U^{n+1} &:= \mathcal{R}_n(Y^{n+1}) \ . \end{aligned}$$

$\mathcal{R}_n$ (nonlinear) projection to model class

$$\mathcal{R}_n : \mathbb{R}^{n_1 \times n_2} \to \mathcal{M}_{\mathbf{r}}$$

e.g HSVD $\sigma_s := \sigma_{s_t}$ singular values of $\mathbf{Y}_t = \mathbf{Y}_t(Y^{n+1})$, $t \in \mathbb{T}$,

1. Hard thresholding, $\sigma_s := 0$, $s > r$, $\sigma_s \leftarrow \sigma_s$, $s \leq r$ compressive sensing: Blumensath et al. , matrix recovery : Tanner et al.
2. Riemannian techniques including ALS: e.g. Kressner et al. (2013), da Silva & Herrmann (2013)

We obtain first, similar convergence results based on Tensor RIP.

Convex framework for tensor product approximation - in preparation

or can we learn from *Compressive Sensing*?
We want to find

$$U_{\mathbf{r}} \in \{V \in \mathcal{H}_d : \|U - V\| \leq \epsilon\} \, , \text{ where } \mathcal{A}U - Y = 0 \, .$$

with minimal ranks (ℓ_0-norm) i.e. we are minimize your costs. (fixing our accuracy) In compressive sensing ℓ_0 is relaxed by ℓ_1-norm.
Soft Thresholding [Daubechies, Defrise, DelMol (2004)] (linear) convergence but only to the minimizer of, e.g. $d = 2$

$$\epsilon \|U\|_{*,1} + \|\nabla \mathcal{J}(U)\|^2$$

(Bachmayr & S.) work under construction

Iterative Hard Thresholding - Remarks

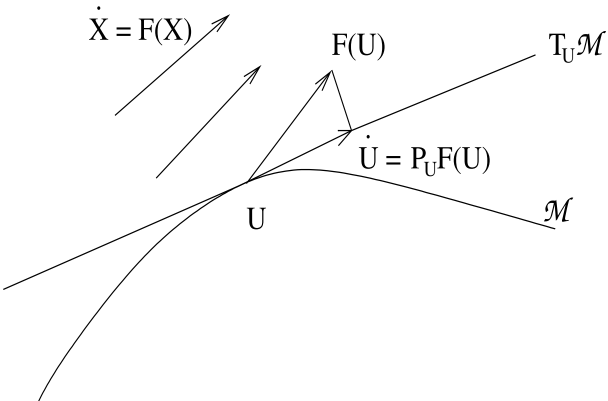
- IHT converges only if the pre-conditioner is sufficiently good. Convergence is linear.
- IHT can be easily combined with enrichment strategies ($\mathbf{r} \uparrow$ (see also Bachmayr / Damen)
- RGI is fast (avoiding large HSVD), but only to local minimizers.
- RGI requires special care at singular point (where $\mathbf{s} < \mathbf{r}$).
- Good preconditioners can speed up the convergence of RGI
- Subspace accelerations like CG, BFGS, DIIS, Anderson are powerful using an appropriate *vector transport* (i.e. transporting previous tangent vectors to the new tangential space) (Pfeffer 2014, Vandereycken, Haegemann et al. (CG))

Practically

- good initial guesses are important
- RGI must be combined with enrichment strategies, e.g. greedy techniques, two-site DMRG or AMEN (Sebastianov et al.)

II.

Dynamical Low Rank Approximation - TT resp. HT Tensors



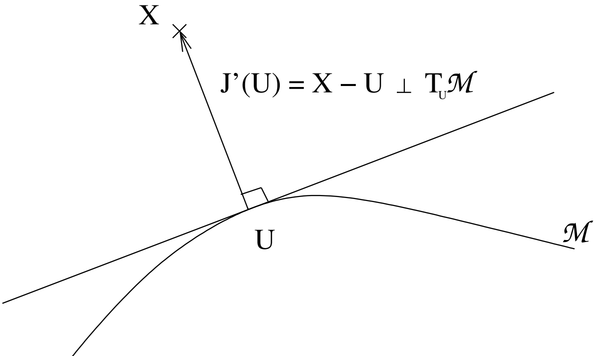
Dirac Frenkel principle $\mathcal{M} \subseteq \mathcal{V}$

$\triangleright$ for optimisation tasks $\mathcal{J}(U) \to \min$:

Solve first order condition $\mathcal{J}'(U) = 0$ on tangent space,

$$\langle \mathcal{J}'(U), V \rangle = 0 \quad \forall V \in \mathcal{T}_U.$$

(Dirac-Frenkel variational principle, Absil et al., Q.Chem.: MCSCF, ...)



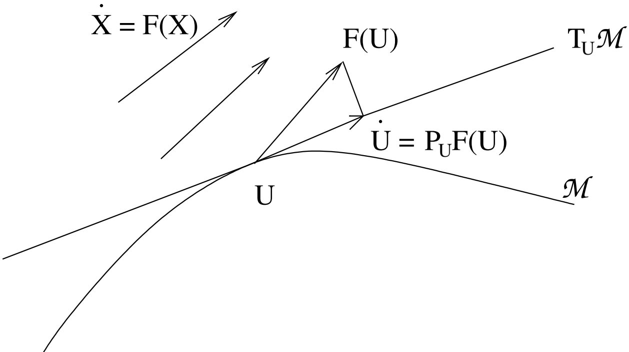
Dirac Frenkel principle $\mathcal{M} \subseteq \mathcal{V}$

$\triangleright$ for differential equations $\dot{X} = f(X), X(0) = X_0$:

Solve projected DE, $\dot{U} = P_U f(U), U(0) = X_0 \in \mathcal{M}$,

$$\langle \dot{U}(t), V \rangle = \langle f(U(t)), V \rangle \quad \forall V \in \mathcal{T}_{U(t)} .$$

(Dirac-Frenkel variational principle, Lubich et al., Q.Chem.: TDMCH . . .)



Convergence estimates

Time-dependent equations:

$$\frac{\partial}{\partial t} U = \mathcal{A} U + F(U) \ , \ U(0) = U_0 \in \mathcal{M}_{\underline{r}} \ ,$$

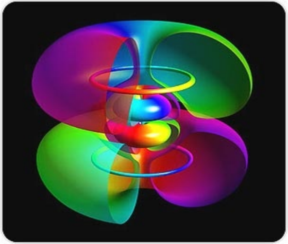
$\mathcal{A} = \sum_{i=1}^{d} I \otimes \cdots I \otimes A_i \otimes I \cdots$, $A_i = H_0^1(\Omega) \cap H^2(\Omega) \to L_2(\Omega)$.

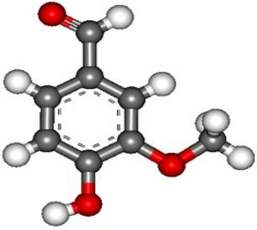
- ▷ Quasi-optimal error bounds
  (Lubich/Rohwedder/Schneider/Vandereycken)
  $\mathcal{A} = 0$, $0 \leq t < T$ solution $X(t)$ with approx. $U(t) \in \mathcal{M}_{\mathbf{r}}$, $X(0) = U(0)$,

$$\begin{aligned} & \| U(t) - U_{\text{best}}(t) \| \\ \lesssim \quad & \| \Psi(t) - V(t) \| + tL \int_0^t \Big(\inf_{V(s) \in \mathcal{M}_{\mathbf{r}}} \| \Psi(s) - V(s) \| + \varepsilon \Big) ds \end{aligned}$$

IV.

Appendix: Tensorization (and second quantization)





Vector-tensorization - e.g. binary coding

1D example: vector, e.g. signal or function $g : [0,1] \to \mathbb{R}$,

$$k \to f(k) \,, \ \left(\text{or } \ g(\frac{k}{2^d}) \right) \,, \ k = 0, \ldots, 2^d - 1 \,.$$

Labeling of indices $k \simeq \mu \in \mathcal{I}$ by an binary string of length d,

$$\boldsymbol{\mu} = \mu(k) \quad = \quad (0,0,1,1,0,\ldots) \simeq \sum_{j=0}^{d-1} \mu_i 2^j = k(\boldsymbol{\mu}) \,, \ \mu_i = 0,1 \,.$$

Tensorization

$$\boldsymbol{\mu} \mapsto U(\boldsymbol{\mu}) := f(k(\boldsymbol{\mu})) \in \bigotimes_{j=0}^{d-1} \mathbb{R}^2 \quad , \quad \text{or } \bigotimes_{j=0}^{d-1} \mathbb{C}^2 .$$

This provides an isomorphm $T : \mathbb{R}^{2^d} \leftrightarrow \bigotimes_{j=0}^{d-1} \mathbb{R}^2$ by $Tf := U$.
So far no information is lost, $N = 2^d$ or $d = \log_2 N$.

Binary coding - signal compression - 1 D functions

Quantized TT - Oseledets (2009), Khoromskij (2009) :

TT approximation of U

- Storage complexity N is reduced to $2r^2 \log_2 N$! (linear in $d = \log_2 N$)
- Allow e.g extreme fine grid size $h = o(\epsilon) = 2^{-d} = \frac{1}{N}$.

Examples:

1. For Kronecker $\delta_{i,j}$ (Dirac function) is $r = 1$.
2. For plane wave (fixed $k = \sum_{j=1}^{d} \nu_j 2^{j-1}$)

$$e^{2\pi ik} = e^{2\pi i \sum_{j=1}^{d} \nu_j 2^{j-1}} = \Pi_{j=1}^{d} e^{2\pi i \nu_l 2^{j-1}} \ , \ \nu_j = 0, 1,$$

again (complex) $r = 1$, or (real $r = 2$) .

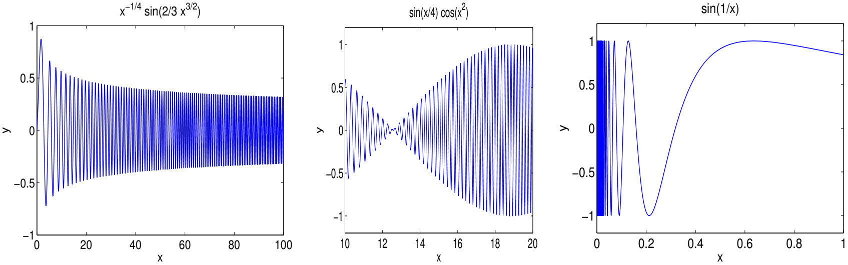
Theorem (Grasedyck)

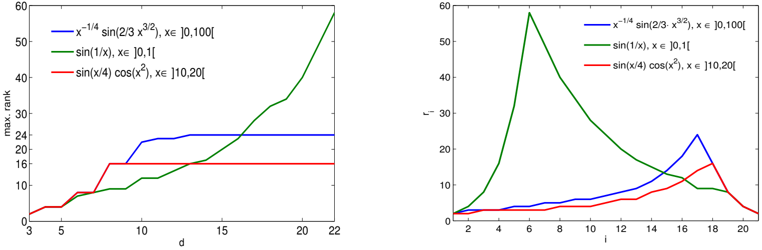
Let $\epsilon > 0$. If $g : [0, 1] \to \mathbb{R}$ is piecewise analytic, for $r_i \geq -\log^{\alpha} \epsilon$, $N \sim \epsilon^{-\tau}$, there is a TT tensor U_ϵ of rank $\leq \underline{r}$ s.t.

$$\|U - U_\epsilon\| \lesssim \epsilon \ , \ dr^2 \lesssim \log^{2\alpha+1} N \ .$$

Examples: TT approximation of tensorized functions

Airy function: $f(x) = x^{1/4} \sin \frac{2x^{2/3}}{3}$, chirp: $f(x) = \sin \frac{x}{4} \cos(x^2)$ and $f(x) = \sin \frac{1}{x}$





Anti-Symmetric Functions - Fermions

Consider a univariate (complete) orthonormal basis

$$V_i := \text{span}\,\{\varphi_i : i = 1, \ldots, d\}\ ,\ \mathcal{H} = \bigotimes_{i=1}^{N} V_i$$

ONB of antisymmetric functions by *Slater determinants*

$$\begin{aligned}\Psi_{SL}[k_1, \ldots, k_N](\mathbf{x}_1; \ldots; \mathbf{x}_N) &:= \varphi_{k_1}(\mathbf{x}_1) \wedge \ldots \wedge \varphi_{k_N}(\mathbf{x}_N) \\ &= \frac{1}{\sqrt{N!}} \det(\varphi_{k_i}(\mathbf{x}_j, s_j))_{i,j=1}^{N}\end{aligned}$$

$$\boxed{\mathcal{V}_{FCI}^N = \bigwedge_{i=1}^{N} V_i = \text{span}\{\Psi_{SL} = \Psi[k_1, \ldots, k_N] : k_1 < \ldots < k_N \le d\} \subset \mathcal{H}}$$

Curse of dimensionality $\dim \mathcal{V}_{FCI}^N = \begin{pmatrix} d \\ N \end{pmatrix}$!

Fock space

Let $\Psi_\mu := \Psi_{SL}[\varphi_{k_1}, \ldots, \varphi_{k_N}] = \Psi[k_1, \ldots, k_N]$ basis Slater det.
Labeling of indices $\mu \in \mathcal{I}$ by an binary string of length d

$$\text{e.g.: } \mu \quad = \quad (0,0,1,1,0,\ldots) =: \sum_{i=0}^{d-1} \mu_i 2^i \ , \ \mu_i = 0,1 \ ,$$

- $\mu_i = 1$ means φ_i is (occupied) in $\Psi[\ldots]$.
- $\mu_i = 0$ means φ_i is absend (not occupied) in $\Psi[\ldots]$.

(discrete) Fock space $\mathcal{F}_d$ is of $\dim \mathcal{F}_d = 2^d$, $(\mathbb{K} := \mathbb{C}, \mathbb{R})$

$$\mathcal{F}_d := \bigoplus_{N=0}^{d} \mathcal{V}_{FCI}^N = \{\Psi : \Psi = \sum_\mu c_\mu \Psi_\mu\}$$

$$\boxed{\mathcal{F}_d \simeq \{\mathbf{c} : \mu \mapsto \mathbf{c}(\mu_0, \ldots, \mu_{d-1}) = c_\mu \in \mathbb{K} \ , \ \mu_i = 0,1 \ \} = \bigotimes_{i=1}^{d} \mathbb{K}^2}$$

This is a basis depent formalism $\Rightarrow$: Second Quantization

Discrete annihilation and creation operators

$$A := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \; A^T = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

In order to obtain the correct phase factor, we define

$$S := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and the discrete annihilation operator

$$\boxed{a_p \simeq \mathbf{a}_p := S \otimes \ldots \otimes S \otimes A_{(p)} \otimes I \otimes \ldots \otimes I}$$

where $A_{(p)}$ means that A appears on the p-th position in the product.
The creation operator

$$\boxed{a_p^\dagger \simeq \mathbf{a}_p^T := S \otimes \ldots \otimes S \otimes A_{(p)}^T \otimes I \otimes \ldots \otimes I}$$

Hamilton operator in second quantization

For a Hamilton operator $H : \mathcal{V}^N_{FCI} =: \mathcal{V} \to \mathcal{V}'$

$$H\Psi = \sum_{\nu',\nu} \langle \Psi_{\nu'}, H\Psi_\nu \rangle c_\nu \Psi_{\nu'} = \sum_{\nu'} (\mathbf{Hc})_{\nu'} \Psi_{\nu'} .$$

Theorem (Slater -Condon)

The Galerkin matrix $\mathbf{H}$ *of a two particle Hamilton operator acting on Fermions is sparse and can be represented by*

$$\boxed{\mathbf{H} = \sum_{p,q=1}^{d} h_p^q \mathbf{a}_p^T \mathbf{a}_q + \sum_{p,q,r,s=1}^{d} g_{r,s}^{p,q} \mathbf{a}_r^T \mathbf{a}_s^T \mathbf{a}_p \mathbf{a}_q \ .}$$

Remark: In case of spin systems the 2×2 matrices $\mathbf{A}, \mathbf{A}^T, \mathbf{S}, \mathbf{I}$ are most replaced by Pauli matrices - e.g. Heisenberg model etc. (-origin of matrix product states (MPS) and DMRG)

See also quantum information theory and quantum computing!

Particle number operator and Schrödinger eqn.

$$P := \sum_{p=1}^{d} a_p^\dagger a_p \ , \quad \simeq \quad \mathbf{P} := \sum_{p=1}^{d} \mathbf{A}_p^T \mathbf{A}_p \ .$$

The space of N-particle states is given by

$$\mathcal{V}^N := \{\mathbf{c} \in \bigotimes_{i=1}^{d} \mathbb{K}^2 : \mathbf{Pc} = N\mathbf{c}\} \ .$$

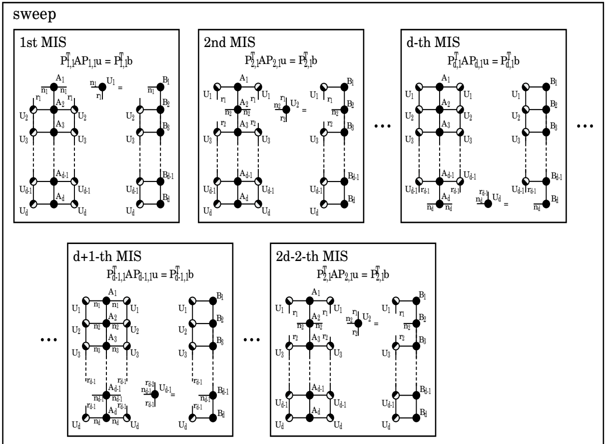
Variational formulation of the Schrödinger equation

$$\boxed{\mathbf{c} = (c(\mu)) = \mathrm{argmin}\{\langle \mathbf{Hc}, \mathbf{c}\rangle : \langle \mathbf{c}, \mathbf{c}\rangle = 1 \ , \ \mathbf{Pc} - N\mathbf{c} = 0\} \ .}$$

Table: New paradigm - discretization

traditional	vs.	new
d fixed , $n \to \infty$ $\mathbb{K}^{nd}$		n fixed, e.g. $n = 2$, $d \to \infty$ $\bigotimes_{j=1}^{d} \mathbb{K}^2$

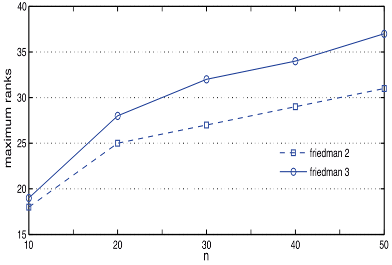
II.
Numerical experiments

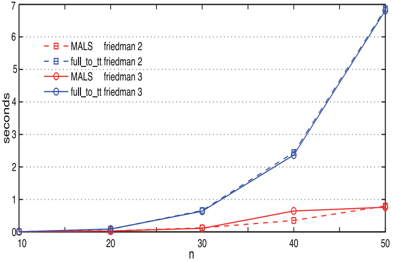


TT approximations of Friedman data sets

$$f_2(x_1,x_2,x_3,x_4) = \sqrt{(x_1^2 + (x_2 x_3 - \frac{1}{x_2 x_4})^2},$$
$$f_3(x_1,x_2,x_3,x_4) = \tan^{-1}\left(\frac{x_2 x_3 - (x_2 x_4)^{-1}}{x_1}\right)$$

on $4 - D$ grid, n points per dim. $\rightsquigarrow n^4$ tensor, $\quad n \in \{3,\ldots,50\}$.



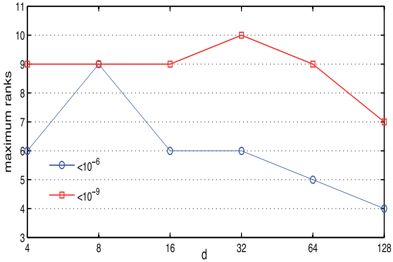


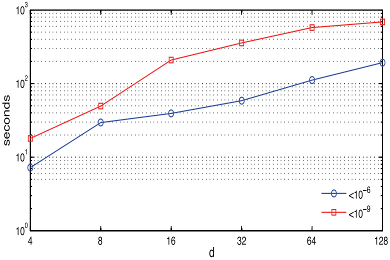
`full_to_tt` (Oseledets, successive SVDs)

and `MALS` (with A = I) (Holtz & Rohwedder & S.)

Solution of $-\Delta U = b$ using MALS/DMRG

- Dimension $d = 4, \ldots, 128$ varying
- Gridsize $n = 10$
- Right-hand-side b of rank 1
- Solution U has rank 13





By now, we are able to solve Fokker Planck, chemical master equations parametric PDE'S, for moderate $r < 100$ and $d \sim 10 - 100$ with Matlab on a laptop. In QTT $n = 2$, $d \sim 1000$. See B. Khoromskij (MPI Leipzig)

Some numerical results - e.g. Parabolic PDEs

joint work with B. Khoromskij, I. Oseledets

$$\frac{\partial}{\partial t}\Psi = H\Psi = (-\frac{1}{2}\Delta + V)\Psi \quad , \quad \Psi(0) = \Psi_0 \ .$$

$$V(x_1, \dots, x_d) = \frac{1}{2}\sum_{k=1}^{f} x_k^2 + \sum_{k=1}^{d-1}\left(x_k^2 x_{k+1} - \frac{1}{3}x_k^3\right) .$$

Timings and error dependence for the modified heat equation (imaginary time) with a Henon-Heiles potential

time interval $[0, 1], \tau = 10^{-2}$, the manifold has ranks 10

Table: Time

Dimension	Time (sec)
2	2.77
4	21.39
8	64.82
16	142.2
32	346.9
64	832.31

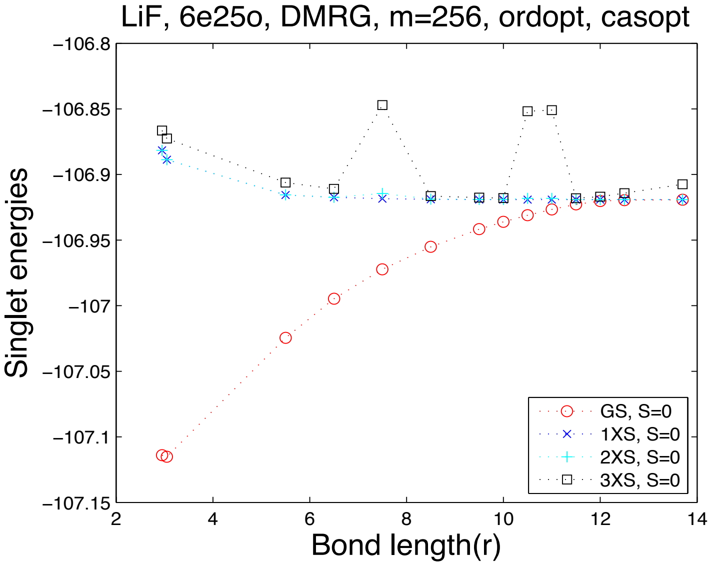
Table: Errror

τ	Error
1.000e-01	3.137e-03
5.000e-02	7.969e-04
2.500e-02	2.000e-04
1.250e-02	5.001e-05
6.250e-03	1.247e-05
3.125e-03	3.081e-06
1.563e-03	7.335e-07

QC-DMRG for HT - tree tensor networks

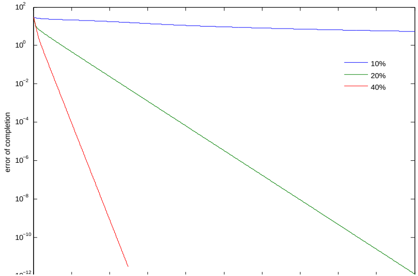
recent joint paper with Legeza, Murg, Nagy, Verstraete (in preparation)

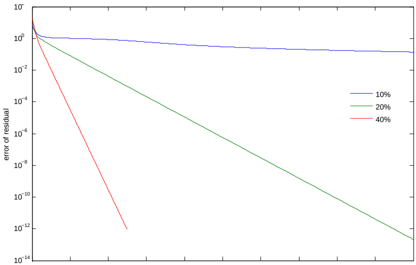
dissoziation of a diatomic molecule *LiF* - first eigenvalues - tree tensor networks (HT)

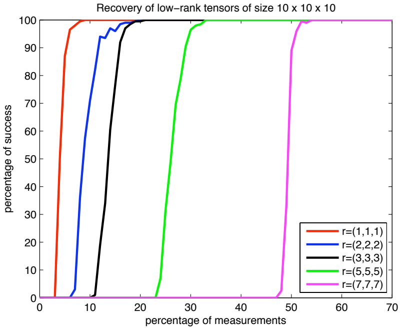


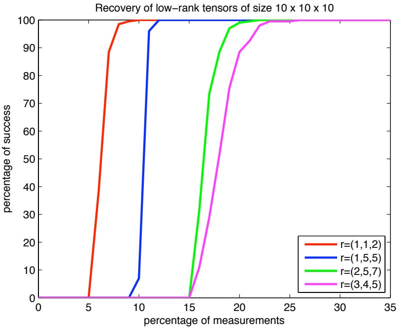
First numerical examples

J.M. Claros -Bachelor thesis, *M. Pfeffer*, TT $d = 4$, $r = 1, 3$, *Stojanac*-Tucker $d = 3$





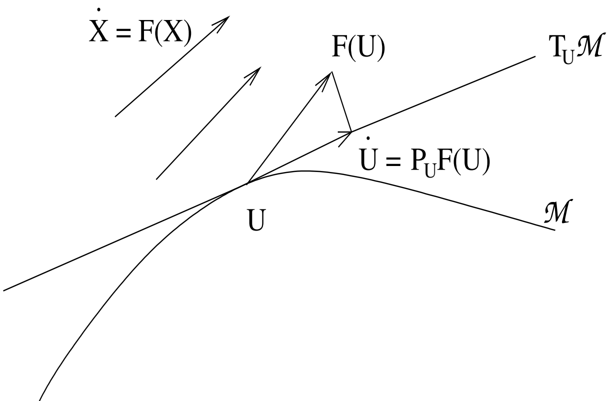




Thank you
for your attention.

II.

Dynamical Low Rank Approximation - Manifolds and Gauge Conditions



Appendix: Manifolds and gauge conditions

Koch&Lubich (2009), Holtz/Rohwedder/Schneider (2011a), Uschmajew/Vandereycken (2012), Arnold& Jahnke (2012)
Lubich/Rohwedder/Schneider/Vandereycken (2012)

- ▷ The sets of above tree (HT, TT or Tucker) tensors of fixed rank $\underline{r}$ each provide embedded submanifolds $\mathcal{M}_{\mathbf{r}}$ of $\mathbb{R}^{(n^d)}$.
- ▷ Canonical tangent space parametrization via component functions $\mathbf{W}_t \in \mathcal{C}_t$ is redundant, but unique via gauge conditions for nodes $t \neq t_r$, e.g.

$$G_t = \left\{ \mathbf{W}_t \in \mathcal{C}_t \mid \langle \mathbf{W}_t^T, \mathbf{B}_t \rangle \ \text{resp.} \ \langle \mathbf{W}_t^T, \mathbf{U}_t \rangle = \mathbf{0} \in \mathbb{R}^{k_t \times k_t} \right\}$$

- ▷ Linear isomorphism

$$E : \times_{t \in T} G_t \to \mathcal{T}_U \mathcal{M}, \qquad E = \sum_{t \in T} E_t$$

E_t: "node-t embedding operators", defined via current iterate $(\mathbf{U}_t, \mathbf{B}_t)$.

Projector onto $\mathcal{T}_U \mathcal{M}$: $\quad P = EE^+$.

Appendix: Manifolds and gauge conditions

Koch&Lubich (2009), Holtz/Rohwedder/Schneider (2011a), Uschmajew/Vandereycken (2012), Arnold& Jahnke (2012)
Lubich/Rohwedder/Schneider/Vandereycken (2012)

- The sets of above tree (HT, TT or Tucker) tensors of fixed rank $\underline{r}$ each provide embedded submanifolds $\mathcal{M}_{\mathbf{r}}$ of $\mathbb{R}^{(n^d)}$.
- Canonical tangent space parametrization via component functions $\mathbf{W}_t \in \mathcal{C}_t$ is redundant, but unique via gauge conditions for nodes $t \neq t_r$, e.g.

$$G_t = \left\{ \mathbf{W}_t \in \mathcal{C}_t \mid \langle \mathbf{W}_t^T, \mathbf{B}_t \rangle \ \text{ resp. } \ \langle \mathbf{W}_t^T, \mathbf{U}_t \rangle = \mathbf{0} \in \mathbb{R}^{k_t \times k_t} \ \right\}$$

- Linear isomorphism

$$E : \times_{t \in T} G_t \ \rightarrow \mathcal{T}_U \mathcal{M}, \qquad E = \sum_{t \in T} E_t$$

E_t: "node-t embedding operators", defined via current iterate $(\mathbf{U}_t, \mathbf{B}_t)$.

Projector onto $\mathcal{T}_U \mathcal{M}$: $\quad P = EE^+$.

Appendix: Manifolds and gauge conditions

Koch&Lubich (2009), Holtz/Rohwedder/Schneider (2011a), Uschmajew/Vandereycken (2012), Arnold& Jahnke (2012)
Lubich/Rohwedder/Schneider/Vandereycken (2012)

- The sets of above tree (HT, TT or Tucker) tensors of fixed rank $\underline{r}$ each provide embedded submanifolds $\mathcal{M}_{\mathbf{r}}$ of $\mathbb{R}^{(n^d)}$.
- Canonical tangent space parametrization via component functions $\mathbf{W}_t \in \mathcal{C}_t$ is redundant, but unique via gauge conditions for nodes $t \neq t_r$, e.g.

$$G_t = \left\{ \mathbf{W}_t \in \mathcal{C}_t \mid \langle \mathbf{W}_t^T, \mathbf{B}_t \rangle \text{ resp. } \langle \mathbf{W}_t^T, \mathbf{U}_t \rangle = \mathbf{0} \in \mathbb{R}^{k_t \times k_t} \right\}$$

- Linear isomorphism

$$E : \times_{t \in T} G_t \to \mathcal{T}_U \mathcal{M}, \qquad E = \sum_{t \in T} E_t$$

E_t: "node-t embedding operators", defined via current iterate $(\mathbf{U}_t, \mathbf{B}_t)$.

Projector onto $\mathcal{T}_U \mathcal{M}$: $\quad P = EE^+$.

Manifolds and gauge conditions

Lubich et al. (2009), Holtz/Rohwedder/Schneider (2011a), Uschmajew/Vandereycken (2012),
Lubich/Rohwedder/Schneider/Vandereycken (2012), Arnold/Jahnke (2012)

- ▷ The sets of above tree (HT, TT or Tucker) tensors of fixed rank $\underline{r}$ each provide embedded submanifolds $\mathcal{M}_{\underline{r}}$ of $\mathbb{R}^{(n^d)}$.
- ▷ Canonical tangent space parametrization via component functions $\mathbf{W}_t \in \mathcal{C}_t$ is redundant, but unique via gauge conditions for nodes $t \neq t_r$, e.g.

$$G_t = \left\{ \mathbf{W}_t \in \mathcal{C}_t \mid \langle \mathbf{W}_t^T, \mathbf{B}_t \rangle \ \text{resp.}\ \langle \mathbf{W}_t^T, \mathbf{U}_t \rangle = \mathbf{0} \in \mathbb{R}^{k_t \times k_t} \right\}$$

- ▷ Linear isomorphism

$$E : \times_{t \in T} G_t \ \to \mathcal{T}_U \mathcal{M}, \qquad E = \sum_{t \in T} E_t$$

E_t: "node-t embedding operators", defined via current iterate $(\mathbf{U}_t, \mathbf{B}_t)$.

Projector onto $\mathcal{T}_U \mathcal{M}$: $\quad P = EE^+$.

Manifolds and gauge conditions

Linear isomorphism

$$E = E(U) : \times_{t \in T} G_t \to \mathcal{T}_U \mathcal{M}, \qquad E(U) = \sum_{t \in T} E_t(U)$$

E^+ Moore Penrose inverse of E

$$\text{Projector onto } \mathcal{T}_U \mathcal{M}: \quad P(U) = EE^+.$$

Theorem (Lubich/Rohwedder/Schneider/Vandereycken, Arnold/Jahnke (2012))

For tensor B, U, V; $\|U - V\| \le c\rho$; there exists C depending only on n, d, such that there holds

$$\begin{aligned} \|\big(P(U) - P(V)\big)B\| &\le C\rho^{-1}\|U - V\|\|B\| \\ \|\big(I - P(U)\big)(U - V)\| &\le C\rho^{-1}\|U - V\|^2 . \end{aligned}$$

These are estimates for the curvature of $\mathcal{M}_{\underline{r}}$ at U.

Optimization problems/differential flow

The problems

$$\langle \mathcal{J}'(U), V\rangle = 0 \quad \text{resp.} \quad \langle \dot{U}, V\rangle = \langle f(U), V\rangle \quad \forall V \in \mathcal{T}_U$$

on $\mathcal{M}$ can now be re-cast into equations for components $(\mathbf{U}_t, \mathbf{B}_t)$ representing low-rank tensor

$$U = \tau(\mathbf{U}_t, \mathbf{B}_t) :$$

With $\mathbf{P}_t^{\perp}$ projector to G_t, embedding operator $E_t = E_t^{\mathbf{U}}$ as above, solve

$$\mathbf{P}_t^{\perp} E_t^T \mathcal{J}'(U) \; = \; \mathbf{0} \quad \text{resp.} \;\; \dot{\mathbf{U}}_t \; = \; \mathbf{P}_t^{\perp} E_t^{+} f(U),$$

for $t \neq t_r$, and

$$E_{t_r}^T \mathcal{J}'(U) = \mathbf{0} \quad \text{resp.} \quad \dot{\mathbf{U}}_t \; = \; E_t^{+} f(U).$$

for the “root” (e.g. by standard methods for nonlinear eqs.)

Convergence estimates

Time-dependent equations:

$$\frac{\partial}{\partial t} U = \mathcal{A} U + F(U) \; , \; U(0) = U_0 \in \mathcal{M}_{\underline{r}} \; ,$$

$\mathcal{A} = \sum_{i=1}^{d} I \otimes \cdots I \otimes A_i \otimes I \cdots$, $A_i = H_0^1(\Omega) \cap H^2(\Omega) \to L_2(\Omega)$.

- Quasi-optimal error bounds
  (Lubich/Rohwedder/Schneider/Vandereycken)
  $\mathcal{A} = 0$, $0 \leq t < T$ solution $X(t)$ with approx. $U(t) \in \mathcal{M}_{\mathbf{r}}$, $X(0) = U(0)$,

$$\begin{aligned} & \|U(t) - U_{\text{best}}(t)\| \\ \lesssim \quad & \|\Psi(t) - V(t)\| + tL \int_0^t \big(\inf_{V(s) \in \mathcal{M}_{\mathbf{r}}} \|\Psi(s) - V(s)\| + \varepsilon \big) ds \end{aligned}$$