Hierarchical tensor representation and Tensor Networks -

R. Schneider (TUB Matheon), A. Uschmajew (U Bonn)

FOCM Montevideo 2014

Acknowledgment

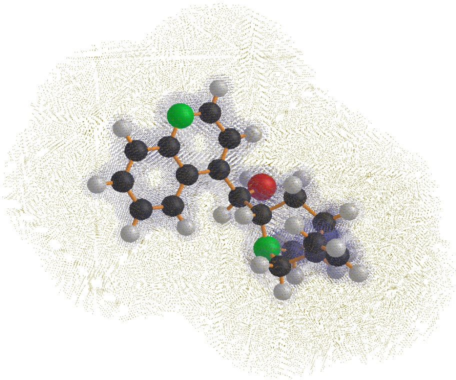
DFG Priority program SPP 1324

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



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

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

- 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,

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

  describing mechanical systems in stochastic environment,

- chemical master equations ...

Parametric PDEs - uncertainty quantification (UQ)

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

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

arise from Schwab et al., Tempone , Nobile, ...

$$\nabla_x a(x, \omega) \nabla_x u(x, \omega) = f(x) \quad + \quad b.c$$

by using Wiener- Ito chaos polynomials, Hermite polynomials h_k

$$u(x, \omega) := \sum_{i=1}^{\infty} \sum_{k_i=1}^{\infty} U[x, k_1, \ldots, k_d] \otimes h_{k_i}(y_i(\omega)) \ , \ x \in D$$

Simple case, by Karhunen Loéve expansion

$$a(x, \omega) \quad = \quad a_0(x) + \sum_{k=1}^{\infty} \lambda_k a_k(x) y_i(\omega)$$

$$\Rightarrow \ \tilde{a}(x, \mathbf{y}) \quad = \quad a_0(x) + \sum_{k=1}^{\infty} \lambda_k \tilde{a}_k(x) y_i \ , \ y_i \in \mathbb{R}$$

$$\text{Hilbert space } \ \mathcal{H} = H_0^1(D) \otimes L_2(\mathbb{R}^d, \mu) \ , \ d \to \infty$$

Stochastic PDEs (numerical Malliavin calculus)Karniadakis et al.

Quantum physics - Fermions

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

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

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

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

The stationary (discrete) Schrödinger equation (neutral syst.)

$$\boxed{\mathbf{HU} = E_0 \mathbf{U} \ , \quad \mathbf{U} \in \bigotimes_{j=1}^{d} \mathbb{C}^2 \simeq \mathbb{C}^{(2^d)} \ , \quad (d \to \infty)}$$

includes electronic Schrödinger equation as well as Hubbard, Heisenberg model etc.

Curse of dimensions

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

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

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

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

$$U = \Big(U[x_1, x_2, \ldots, x_d] \Big)_{x_i = 1, \ldots, n_i \;,\; i=1,\ldots,d} \in \mathcal{H}_d \;\;,$$

$d = 1$: n-tuples $(U_x)_{x=1}^{n}$, or $x \mapsto U[x]$,

or $d = 2$: matrices $\big(U[x, y] \big)$ or $(x, y) \mapsto U[x, y]$.

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

$\dim \mathcal{H}_d = \mathcal{O}(n^d)$ $\;\; --$ 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,

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

Notation: $\mathbf{x} = (x_1, \ldots, x_d) \mapsto U = U[x_1, \ldots, x_d] \in \mathcal{H}_d$ For simplicity let us consider the Hilbert spaces $\ell_2(\mathcal{I})$, $\mathcal{I} = \{1, \ldots n\}$

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,

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

Notation: $\mathbf{x} = (x_1, \ldots, x_d) \mapsto U = U[x_1, \ldots, x_d] \in \mathcal{H}_d$ For simplicity let us consider the Hilbert spaces $\ell_2(\mathcal{I})$, $\mathcal{I} = \{1, \ldots n\}$

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,

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

Notation: $\mathbf{x} = (x_1, \dots, x_d) \mapsto U = U[x_1, \dots, x_d] \in \mathcal{H}_d$ For simplicity let us consider the Hilbert spaces $\ell_2(\mathcal{I})$, $\mathcal{I} = \{1, \dots n\}$

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.

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

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

Curse of non-linearity - non-convexity

Formally low order scaling:

- operators + right hand side admit low rank tensor representation, e.g.
- there might be a hope the the solution of also low rank
- promising to get rid of the curse of dimensions

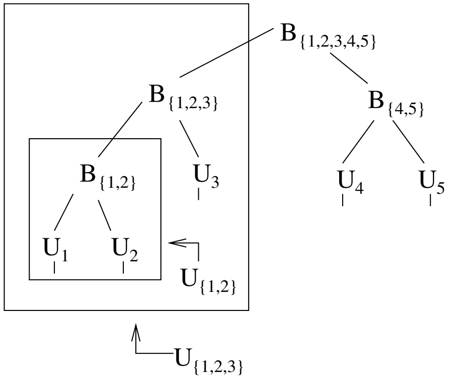
But curse of non-linearity or curse of non-convexity

- the multi-linear parametrization is not harmless
- counter example by Silva & Lim (non-closedness)
- even the computation of the best rank one approximation is NP hard Hilliar & Lim *Almost all tensor problems are NP hard*
- counter example by Landsberg & Ke Ye for tensor networks with closed loops

Can we circumvent also the problem with nonlinearity and non convexity?

II.

Subspace approximation, hierarchical tensors and tensor networks



(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

$$d_{r,\infty}(F) := \inf_{\{U : \dim U \leq r, U \subset V\}} \sup_{y\in\mathcal{K}} \inf_{f_y \in U} \|F_y - f_y\|$$

$$d_{r,2}(F) := \inf_{\{U : \dim U \leq r, U \subset V\}} \left(\int_{\mathcal{K}} \inf_{f_y \in U} \|F_y - f_y\|^2 dy \right)^{\frac{1}{2}}$$

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*

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

SVD as sub-space approximation

We are seeking subspaces $U_i \subset V_i$, $i = 1, 2$, fitting best to a given tensor $X \in V_1 \otimes V_2$, in the sense

$$\boxed{\|X - V^*\|^2 := \inf_{\{V \in U_1 \otimes U_2 \ : \ \dim\ U_i \leq r\}} \|X - V\|^2}$$

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

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

$$U_i = \text{ span } \{\mathbf{b}^i_{k_i} : k_i \leq r\} \subset V_i \quad , \quad \text{rank } r \ .$$

$$\Rightarrow \quad C[k_1, k_2] = \langle X, \mathbf{b}^1_{k_1} \otimes \mathbf{b}^2_{k_2} \rangle \ \text{ in SVD } C = \text{diag}[\sigma_k]$$

$$V^*[x_1, x_2] = \sum_{k_1=1}^{r} \sum_{k_2=1}^{r} C[k_1, k_2] \mathbf{b}^1_{k_1}[x_1] \otimes \mathbf{b}^2_{k_2}[x_2]$$

Tucker decomposition - sub-space approximation

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

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

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

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

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

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

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

Data complexity: $\mathcal{O}(ndr + r^d)$ curse of dimensions!

Subspace approximation

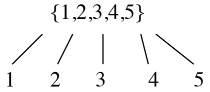
Subspace approximation

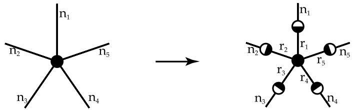
- ▷ Tucker format (MCSCF, MCTDH(F)) - robust
  But complexity $\mathcal{O}(r^d + ndr)$

  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.)

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





Subspace approximation

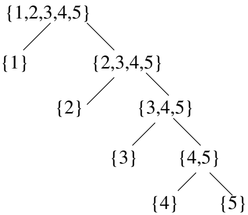
- Tucker format (MCSCF, MCTDH(F)) - robust
  But complexity $\mathcal{O}(r^d + ndr)$

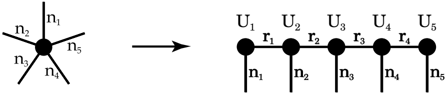
  Is there a robust tensor format, but polynomial in d?

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

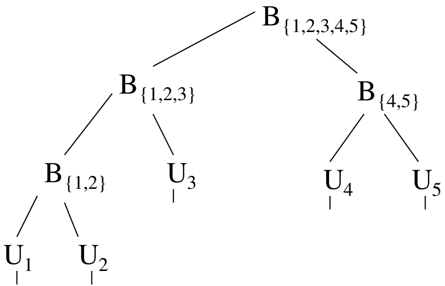
Hierarchical tensor (HT) format

- 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

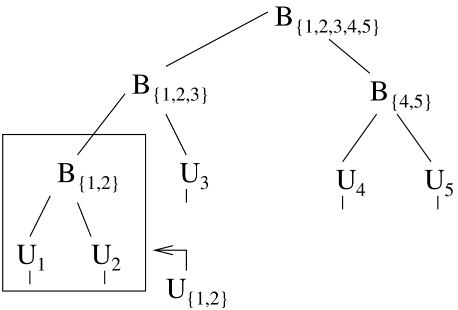
- Canonical decomposition   not closed, no embedded manifold!
- Subspace approach (Hackbusch/Kühn, 2009)



(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

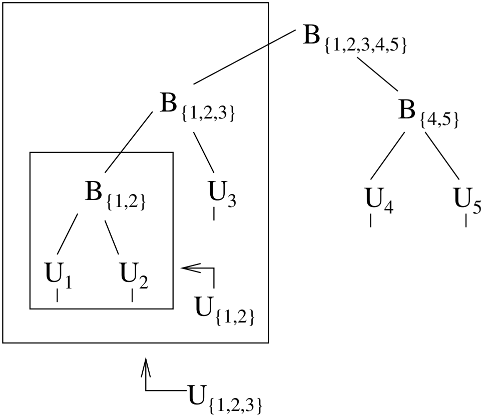
- Canonical decomposition   not closed, no embedded manifold!
- Subspace approach (Hackbusch/Kühn, 2009)



(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

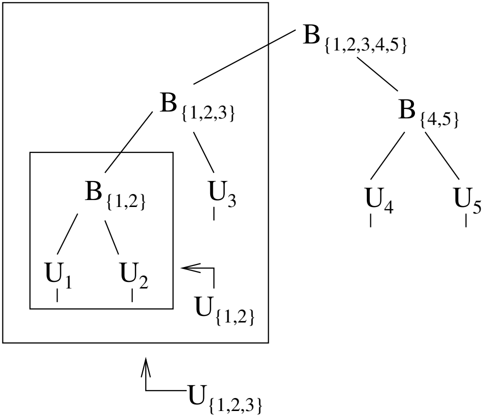
- Canonical decomposition   not closed, no embedded manifold!
- Subspace approach (Hackbusch/Kühn, 2009)



(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)

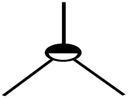


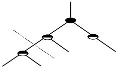
(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n \times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t_1} \times k_{t_2}}$)

Recursive definition by bases representations

$$U^{\alpha} \;=\; span\{\mathbf{b}_i^{(\alpha)} : 1 \leq ir_{\alpha}\}$$

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





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

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

(with obvious modifications for $\alpha = D$ or α is a leave.)
Parametrization τ is multi-linear
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

$$U[\mathbf{x}] = U_1 \circ \cdots U_d[\mathbf{x}] = \mathbf{U}_1[x_1] \cdots \mathbf{U}_i[x_i] \cdots \mathbf{U}_d[x_d]$$

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

with matrices or component functions

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

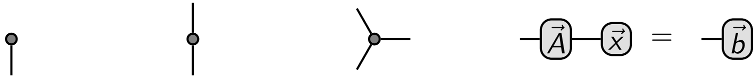
Redundancy: $U[\mathbf{x}] = \mathbf{U}_1[x_1] \mathbf{G} \mathbf{G}^{-1} \mathbf{U}_2[x_2] \cdots \mathbf{U}_i[x_i] \cdots \mathbf{U}_d[x_d]$.

Tensor networks

Diagrammatic notation

Node $\sim$ dot , line $\sim$ index (variable) ,

edge connecting two nodes $\sim$ summation w.r.t this index



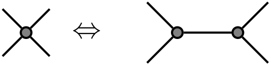
vector | matrix | tensor of order 3 | equation $\mathbf{Au} = \mathbf{b}$

$k \mapsto \mathbf{u}[k]$, $(x, y) \mapsto \mathbf{U}[x, y]$, $(x, y, z) \mapsto U[x, y, z]$, $\sum_{j=1}^{n} \mathbf{A}[x, j]\mathbf{u}[j] = \mathbf{b}[x]$

contraction and decontraction of a tensor of order 4

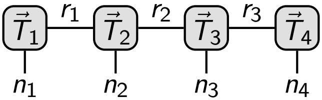
rank r decomposition, e.g. SVD

$$A[x_1, x_2, y_1, y_2] = \sum_{k=1}^{r} U[x_1, x_2, k] V[k, y_1, y_2]$$



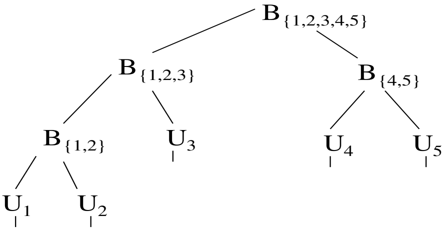
Tree tensor networks

Tensor trains - Oseledets & Tyrtyshnikov (2009) - simple tree tensor network



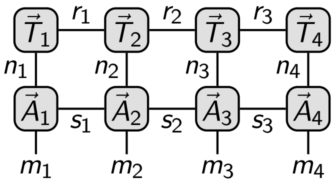
$$T[x_1, \ldots, x_4] \quad = \quad \sum_{k_1=1}^{r_1} \ldots \sum_{k_3=1}^{r_3} T_1[x_1, k_1] T_2[k_1, x_2, k_2] T_3[k_2, x_3, k_3] T_4[k_3, x_4]$$

Hierachical (Tucker) format Hackbusch & Kühn (2008)

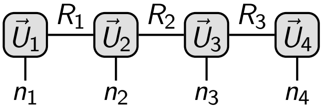


Operating on tensor networks

Application of an operator $U \mapsto AU$



this is again a TT tensor but with possibly larger R_i instead of r_i



Complexity of computing $T \mapsto \mathbf{A}T$ linear in d polynomial in n, r, s, m

Remark the cost of an SVD of $U[k_{i-1}, n_i; k_i]$with rank r is roughly $\mathcal{O}(rnR^2)$

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

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

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

Lie group action

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

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

defines a manifold

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

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 $\delta\mathbf{U}_i \quad \perp \quad \text{span } \mathbf{U}_i \ , \ \forall i < d$. compare with matrices of rank $\leq r$

Holtz & Rohwedder & S. , Uschmajev & Vandereycken, Arnold & Jahnke , Falco &

Fundamental properties of HT (particularly TT)

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

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

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

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

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

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

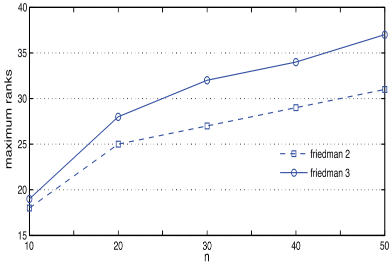
  $$\boxed{\mathcal{M}_{\leq \mathbf{r}} = \bigcup_{s_i \leq r_i} \mathcal{M}_\mathbf{s} = \overline{\mathcal{M}_\mathbf{r}} \subset \mathcal{H} \text{ is (weakly) closed!}}$$ Hackbusch & Falco
- $$\mathcal{M}_{\leq \mathbf{r}} \text{ is a an algebraic variety.}$$

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, $n \in \{3, \ldots, 50\}$.



full_to_tt

ALS (with A = I) (Holtz & Rohwedder & S.)

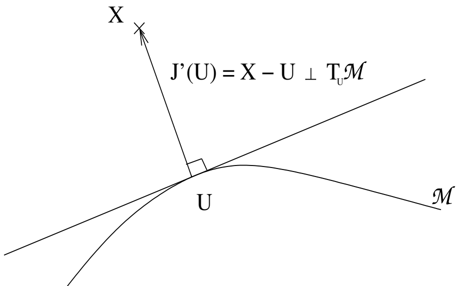
Summary - reduction to matrix analysis

- $d = 2$ is unique because tensors of order two can be considered as matrices or linear operators, where we have spectral theory
- removing an edge in a tree decompose it into 2 subtrees - low rank matrix factorisation
- the trick is that one can reduce the treatment of tree tensor networks to matrix analysis or to Hilbert-Schmidt operators
- but retaining a low order scaling (it is slightly worse than the canonical format)
- Subspace approximation and Grassmann manifold are the central concepts

We will see that with this approach the curse of non-linearity and non-convexity can be avoided to a large extend, and still with polynomial complexity

III.

Optimization with tensor networks and hierarchical tensors Dirac-Frenkel variational principle



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*

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

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

Problem (Tensor product optimization problem (TOP))

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

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

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

example

Espig,Hackbusch, Rohwedder & Schneider (2010)

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

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

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

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

here

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

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

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

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 $\mathcal{A} \cap \mathcal{M}_{\leq \mathbf{r}} = \mathcal{M}_{\leq \mathbf{r}}$.

Block coordinate search for TT (HT) tensors - ALS

E.g. Let $\mathcal{J}(U) := \langle \mathcal{A}U - f, \mathcal{A}U - f \rangle$ For $j = 1, \ldots, d$ do,

1) fix all component tensors U_ν, $\nu \in \{1, \ldots, d\} \backslash \{j\}$, except index j. Then the actual parametrization becomes linear,

2) Optimize $\mathbf{U}^j[k_{j-1}, x_j, k_j]$, $U_1 \circ \cdots U_{i-1} \otimes U_{i+1} \circ \cdots U_d$ spans a linear subspace $\simeq \mathbb{R}^{r_{i-1}} \otimes V_i \otimes \mathbb{R}^{r_i} \subset \mathcal{H} \Rightarrow$ coupled (linear) system of low dimensional PDEs (ODEs)
3) and orthogonalize left to define a basis for the next step
4) Repeat with $\mathbf{U}^{j+1}$ (tree hierarchy is reordered to optimize alway the root!)

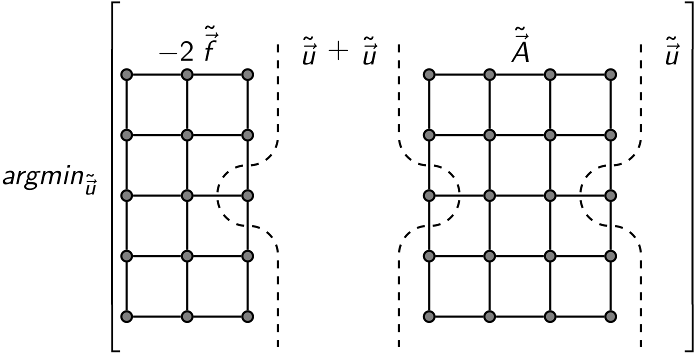
S. Holtz & Rohwedder & S. (2010), Uschmajew & Rohwedder (2011), rigorous analysis is difficult, see Uschmajew & S.

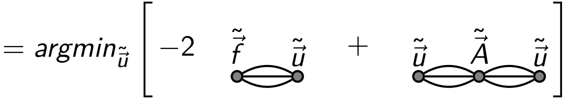
This is the single site DMRG /density matrix renormalization algorithm (S. White (1992))

This provides a model reduction often by a factor $10^3 - 10^{10}$ and more!

ALS - (single site) Density Matrix Renormalization Group

$$U = \operatorname{argmin}\{\langle \mathcal{A}U, \mathcal{A}U\rangle - 2\langle \mathcal{A}f, U\rangle : U \in \mathcal{M}_{\leq \mathbf{r}}\} \ ,$$





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

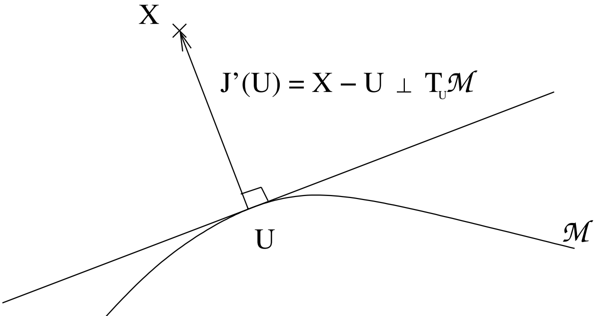
Dirac Frenkel principle $\mathcal{M}_{\mathbf{r}} \subseteq \mathcal{H}$

$\triangleright$ 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., Arias & Edelman &Lippert ...)



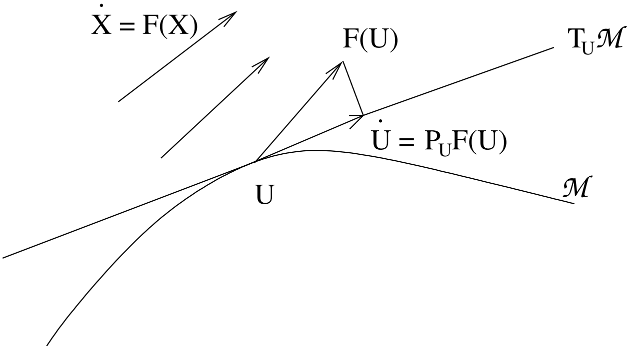
Dirac Frenkel principle $\mathcal{M}_{\mathbf{r}} \subseteq \mathcal{H}$

$\triangleright$ 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., Chemistry: TDMCH ...)



Riemannian gradient iteration

Along the manifold $M_{\mathbf{r}}$, minimize

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

by Riemannian gradient iteration Edelman et al. Absil et al. ...

1) $V^{n+1} := U^n \ - P_{\mathcal{T}_U}\alpha_n\Big(\mathcal{C}^{-1}(\mathcal{A}U^n - Y)\Big)$

$V^{n+1} = U^n + \boldsymbol{\xi}^n \quad , \ \mathcal{M}_{\mathbf{r}} + \mathcal{T}_U$

2) $U^{n+1} := \mathcal{R}_n(V^{n+1}) := R(U^n, \boldsymbol{\xi}^n)$.

daSilva & Herrman, Lubich & Rohwedder & S. & Vandereycken

$P_{\mathcal{T}_U} : \mathcal{H} \to \mathcal{T}_U$ orthogonal projection onto tangent space at U

retraction (Absil et al., M. Shub) $R(U, \boldsymbol{\xi}) : T_{\mathcal{M}_{\mathbf{r}}} \to \mathcal{M}_{\mathbf{r}}$,

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

e.g. R is an approximate exponential map

Riemann gradient iteration global convergence

Theorem (joint work with A. Uschmajew ($d = 2$))

Let $V^{n+1} := U^n + \lambda_n \mathcal{C}^{-1}(Y - \mathcal{A}U^n)$, and $\mathcal{A}$ is SPD ($n < \infty$). 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). The paper includes the treatment of singular points.

Lojasiewicz (-Kurtyka) inequality

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

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

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

$0 < \theta < \frac{1}{2}$: $\|U^n - U\| \lesssim n^{-\frac{\theta}{2-\theta}}$

There remains the curse of non-convexity, i.d. convergence only to stationary points - one is often trapped by local minima

Convergence estimates

Time-dependent equations:

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

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

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

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

Some numerical results - e.g. Parabolic PDEs

joint work with B. Khoromskij, I. Oseledets

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

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

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

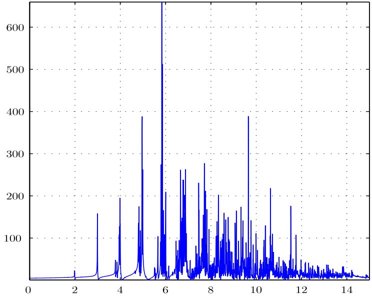
τ	Error
1.000e-01	3.137e-03
5.000e-02	7.969e-04
2.500e-02	2.000e-04
1.250e-02	5.001e-05
6.250e-03	1.247e-05
3.125e-03	3.081e-06
1.563e-03	7.335e-07

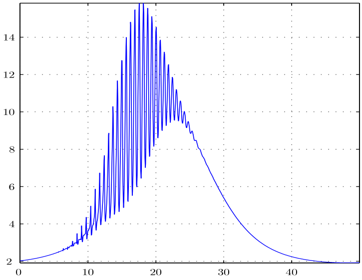
Some numerical results - e.g. time dependent Schrödinger

joint work with B. Khoromskij, I. Oseledets

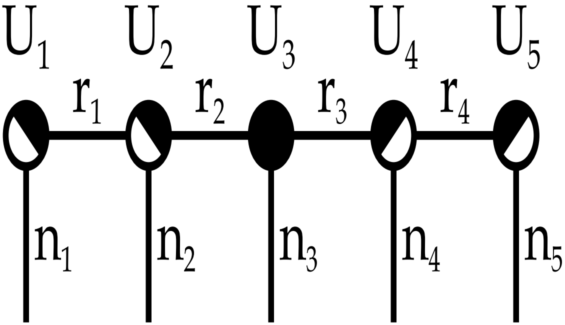
$$i\frac{\partial}{\partial t}\Psi = H\Psi = (-\frac{1}{2}\Delta + V)\Psi \quad , \quad \Psi(0) = \Psi_0 \ .$$

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





IV.
The hierarchical SVD (HSVD)



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

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

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

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

HSVD - hierarchicalSVD

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

Matricisation or unfolding

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

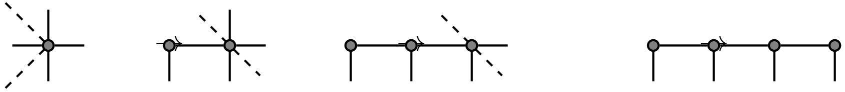
The tensor $\mathbf{x} \to U[\mathbf{x}]$

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

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

with matrices or component functions

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



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

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

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

Error estimate of truncated HSVD

and quasi-optimal approximation Grasedyck

$$\| U - \mathbf{H}_{\mathbf{s}}(U) \| \leq \sum_t \sum_{k_t=1}^{s_t} \sigma_{k_t}^2$$

$$\| U - \mathbf{H}_{\mathbf{s}}(U) \| \leq \sqrt{d-1} \, \mathrm{inf}_{V \in \mathcal{M}_{\mathbf{r}}} \| U - V \|$$

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

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

Iterative Hard Thresholding

Minimize residual

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

w.r.t. low rank constraints by

1) $Y^{n+1} \quad := \quad U^n \qquad - \mu\Big(\mathcal{C}_n^{-1}(\mathcal{A}U^n - f)\Big)$ prec. gradient step
2) $U^{n+1} \quad := \quad \mathcal{R}_n(Y^{n+1})$.

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

$$\mathcal{R}_n : \mathcal{H} \to \mathcal{M}_{\leq \mathbf{r}}$$

e.g HSVD σ_{s_t} singular values of $\mathbf{Y}_t = \mathbf{Y}_t(Y^{n+1})$, $\sigma_s := 0$, $s_t > r_t$,

Iterative Hard Thresholding

Minimize residual

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

w.r.t. low rank constraints by

$$\begin{aligned} 1)\ Y^{n+1} &:= U^n - \mu\Big(\mathcal{C}_n^{-1}(\mathcal{A}U^n - f)\Big) \quad \text{prec. gradient step} \\ 2)\ U^{n+1} &:= \mathcal{R}_n(Y^{n+1}) \ . \end{aligned}$$

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

$$\mathcal{R}_n : \mathcal{H} \to \mathcal{M}_{\leq \mathbf{r}}$$

e.g HSVD σ_{s_t} singular values of $\mathbf{Y}_t = \mathbf{Y}_t(Y^{n+1})$, $\sigma_s := 0$, $s_t > r_t$,
Theorem With good preconditioning, i.e.spectrum $C_n^{-1} \in [\gamma, \Gamma]$ s.t.

$$\mu := \frac{\Gamma + \gamma}{2}\ ,\ \kappa := \frac{\Gamma}{\gamma}, \rho := \frac{1-\kappa}{1+\kappa} \leq \frac{1}{2\sqrt{d-1}}$$

one obtains linear convergence $\rho < 1$ to a quasi best approximation

$$\|U^n - U\| \lesssim \frac{\rho^n}{1-\rho}\|U_1 - U_0\| + \frac{1}{1-\rho}\mathrm{dist}(U, \mathcal{M}_{\mathbf{r}})$$

Iterative Hard Thresholding - adaptive algorithm

Projected Gradient Algorithms: Minimize residual

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

w.r.t. weakend low rank constraints $(r_t, s_t \to \infty, n \to \infty)$

1) $Y^{n+1} := U^n - \mu_n \mathcal{S}_n\Big(\mathcal{C}_n^{-1}(\mathcal{A}U^n - \mathbf{y})\Big)$ gradient step

2) $U^{n+1} := \mathcal{R}_n(Y^{n+1})$.

with (nonlinear) projections to model class

$$\mathcal{R}_n : \mathcal{H} \to \mathcal{M}_{\leq \mathbf{r}_n} \ , \ \mathcal{S}_n : \mathcal{H} \to \mathcal{M}_{\leq \mathbf{s}_n}$$

AMEn Savostyanov & Dolgov, Adaptive Algorithm Dahmen & Bachmayr

A posteriori control of the residual by HSVD

With appropriate choice of $\mathbf{r}_n, \mathbf{s}_n$ the iterate $U_n \in \mathcal{M}_{\leq \mathbf{r}_n}$ converges to the exact solution $U \in \mathcal{H}$

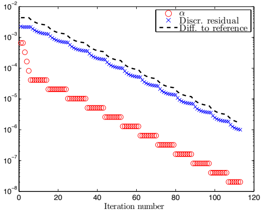
$$\|U - U_n\| \lesssim \frac{\rho^n}{1-\rho}\|U_1 - U_1\| \to 0 \ , \ n \to \infty$$

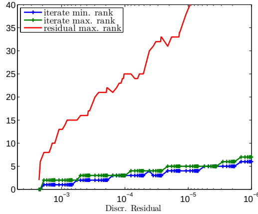
no curse of non convexity - but more expensive!

Iterative Soft Thresholding - adaptive algorithm

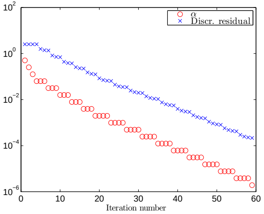
Ongoing joint work with M. Bachmayr (in preparation)

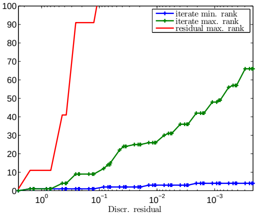
Iteration history and effective ranks, α threshold, for $\Delta u = f$





for a parametric PDE





Preconditioning and Sobolev-norms

Remark: $\mathcal{H}$ is always assumed to be Hilbert spaces with cross norm, i.e. $\|U\| = \|U_1\|_{V_1} \cdots \|U\|_{V_d}$ for all rank 1 tensors $U \in \mathcal{H}$. and $\|U\| = \|U\|_{\ell^2}$ for $U \in \mathcal{H}_d$. E.g. $\|U\| = \|U\|_{L^2}$.Let

$$\langle \mathcal{A}U, U \rangle \sim \|U\|_{H^1(\mathcal{I}^d)}$$

$$\Delta = \sum_{i=1}^{d} \Delta_i = \sum_{i=1}^{d} I \otimes \cdots \Delta_i \otimes \cdots I \ .$$

There is no isomorphic operator $\mathcal{C} : L^2(\mathcal{I}^d) \to H^1(\mathcal{I}^d)$ with finite (multi-linear) rank. But on finite dimensional subspaces $\otimes_{i=1}^d V_{i,J}$, $n_i \sim 2^J$. one can use the d-dim BPX

$$\mathcal{C}_J = \sum_{j=0}^{J} 2^j P_j \otimes P_j \otimes \cdots \otimes P_j$$

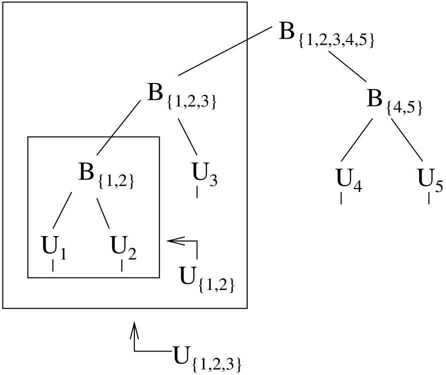
Or exponential sums: e.g for $\Delta^{-1} : H^{-1}(\mathcal{I}^d)L \to H_0^1(\mathcal{I}^d)$

$$\Delta^{-1} = \int_0^\infty e^{-t\Delta_1} \otimes \cdots \otimes e^{-t\Delta_d} dt \approx \sum_{k=1}^{n} \omega_k e^{-t_l \Delta_1} \cdots e^{-t_k \Delta_d}$$

(Beylkin et al., Hackbusch & Braess, Bachmayr & Dahmen)

Both are exponential converging. For details, see Bachmayr & Dahmen.

V.
Regularity and a priori error estimate



Tractablility, regularity and a priori error estimates

- The previous problems are only formally polynomially tractable, e.g TT is of data complexity $\mathcal{O}(dnr^2)$, but for a required accuracy ϵ,

$$dnr^2 = d(\epsilon)n(\epsilon)\Big(r(\epsilon)\Big)^2$$

- fundamental question: what is the dependence of $d(\epsilon), n(\epsilon), r(\epsilon)$ on ϵ?
- Counter examples, i.e. intractable systems, are known from physics spin systems
- Theory is still widely incomplete!!!

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

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

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

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

then the best rank k approximation satisfies (talk Petruchev (Wed))

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

Theorem (Uschmajev & S. (2013))

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

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

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

Estimating the ranks for HT (TT) work in progress

this is ongoing work with A. Uschmajew & D. Kressner , see talk Uschmajew this afternoon

$$\mathcal{A} = \mathcal{A}_1 + \cdots - A_d \ ,$$

$\mathcal{A}_i$ with nearest neighbor interaction

Using ideas from Arad & Kitaev & Landau & Verzirani. M. Hastings - proof for the area law in physics has simplified by them to a purely combinatorial argument

under construction ...

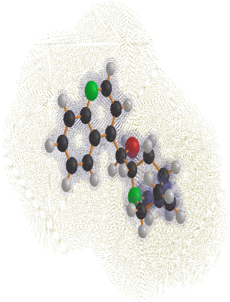
I have not reported about

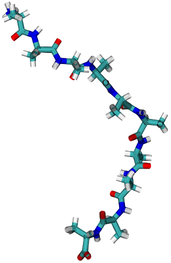
- tensor completion or tensor recovery (Kressner et al., Hermann & daSilva, Rauhhut & Stojanac & S. , Cichocki et al., ...),
- adaptive sampling techniques (Oseledets, Khoromskij et al., Grasedyck, & Kluge)
- vector tensorization e.g. QTT (Oseledets, Khoromskij et al.) - related to wavelets.
- tree optimization (Ballani & Grasedyck)
- greedy methods (Ehrlacher & Cances & Lelievre & Nouy

Comparison and summary

	canonical	Tucker	HT
complexity	$\mathcal{O}(ndr)$ $++$	$\mathcal{O}(r^d + ndr)$ $-$	$\mathcal{O}(ndr + dr^3)$ TT- $\mathcal{O}(ndr^2)$ $+$
rank	no $r_c \geq$	defined r_T	defined $r_T \leq r_{HT} \leq r_c$
(weak) closedness	no	yes	yes
ALS (1site DMRG)	yes - but slow	yes	yes
H (O) SVD	no	yes	yes
embedded manifold	no	yes	yes
Dirac Frenkel	no	yes	yes
algebraic var. $\mathcal{M}_{\leq \mathbf{r}}$	no	yes	yes
recovery	??	yes	yes
quasi best approx.	no	yes	yes
best approx.	no	exist but NP hard	exist but NP hard

VI.
Numerical examples

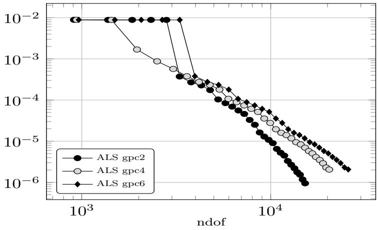




Uncertainty quantification - Stochastic Galerkin

D: L-shaped domain, joint work with M. Eigel (WIAS), M. Pfeffer (TUB)

- u_{TT} - ALS approximation, fixed grid,
- $\|r_{TT}\|_{L_2}$ norm of residual of tensor approximation, is computable by HSVD
- error in $e_{TT} := \|u_h - u_{TT}\|_{H^1} \lesssim \|r_{TT}\|_{L_2}$
- *ndof* number of total degrees of freedom (FEM + TT)
- , $n_i = 2, 4, 6$, $d = 25$, $r = 1, \ldots, 25$



Uncertainty quantification, with a posteriori error bounds

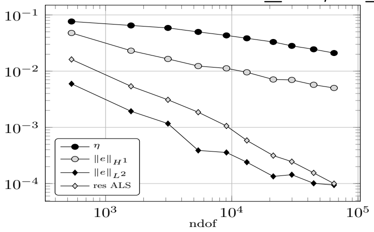
each step increase rank and refine mesh.

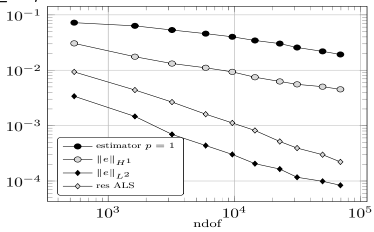
- spatial error of $e_{H^1} := \|\overline{u} - \overline{u_h}\|_{H^1} \approx \|\overline{u} - \overline{u_{TT}}\|_{H^1}$,
- spatial error estimator $\eta := \|r_h\|_{H^{-1} \otimes L_2}$ (FEM)
- tensor approximation error in

$$e_{TT} := \|u_h - u_{TT}\|_{H^1} \leq \|r_{TT}\|_{L_2} \leq e_{H^1}$$

u_{TT} - TT solution, $\|r_{TT}\|_{L_2}$ is computable by *hierarchical SVD (HSVD)*

$r \leq 10,\ n \leq 5,\ d = 25$





$\sigma_\ell^{KL} \sim \ell^{-4}$ (Fast decay) versus $\sigma_\ell^{KL} \sim \ell^{-2}$ (Slow decay)

Transfer operator for MD simulation

Transfer operator at time $\tau > 0$

$$Tp(\mathbf{x}, \tau) = \int_{\mathbb{R}^d} P(\mathbf{x}, \mathbf{y}, \tau) p(\mathbf{y}, \tau) \pi(\mathbf{y}), \; x_i \in \mathcal{I} = [0, 2\pi]$$

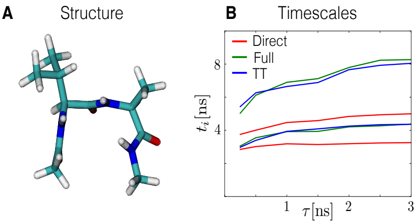
$P(\mathbf{x}, \mathbf{y}, \tau)$ transition probability at time τ, $\pi \sim e^{-\frac{k}{T} V(\mathbf{x})}$, V potential (classical molecular dynamics)

$$T : L^2(\mathcal{I}^d, \pi) \to L^2(\mathcal{I}^d, \pi) \;, \; Tp_i = \lambda_i \pi_i$$

symmetric generalized eigenvalue problem

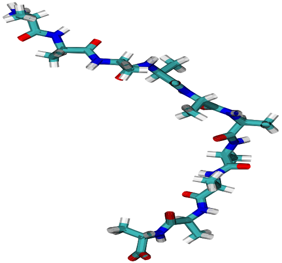
ALS with Kressner & Steinlechner & Uschmajew

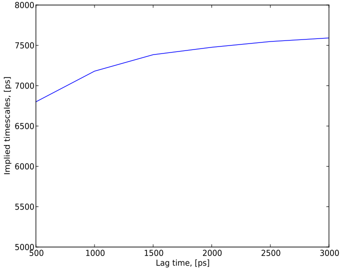
$D = 4, \; N = 3, \; r_{\max} = 4$ (first 3 (2) eigenfunctions)



Transfer operator for MD simulation

Ongoing joint work with Feliks Nüsken & Frank Noe (FU Berlin), in preparation,
We look for the first 3(2) eigenfunctions corresponding to the largest N eigenvectors (the largest $\lambda_0 = 1$ is known)
ALS variant Kressner & Steinlechner & Uschmajew Dimension $d = 18$ rotational vibrations, $N = 2$, $r_{max} = 2$!



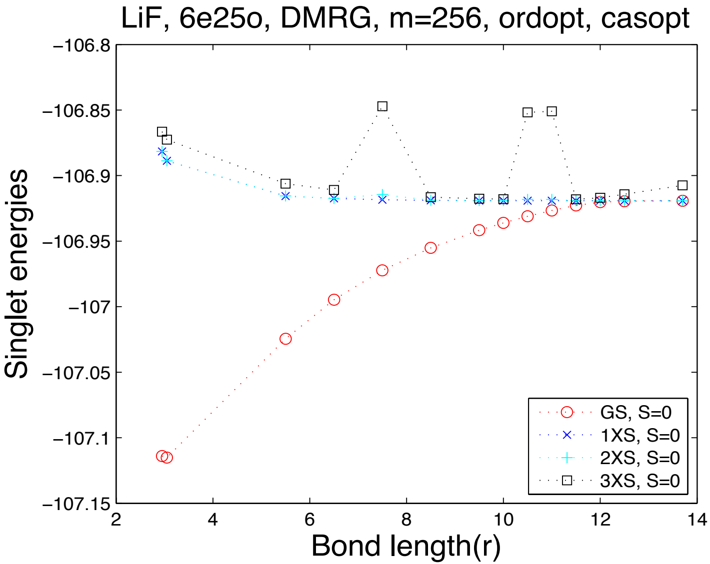


We can compute only $\langle \Psi_i, T\Psi_j \rangle$ by Monte-Carlo sampling over all trajectories from MD simulation.

QC-DMRG for HT - tree tensor networks

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

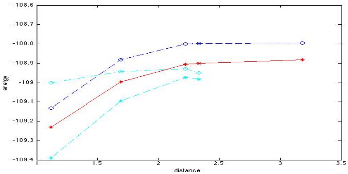
dissoziation of a diatomic molecule *LiF* - first eigenvalues - tree tensor networks (HT)

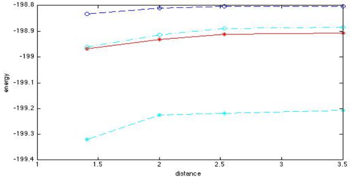


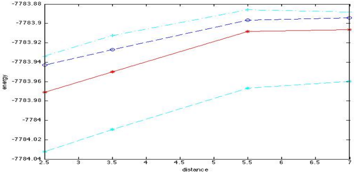
Dissoziation of a diatomic molecules

Dissoziation of a diatomic molecules N_2, F_2 , CsH,

- Budapest DMRG program of O Legeza, computations performed in the group of M. Reiher (ETH Zurich)
- Basis ets H: (6s3p2d)! [4s3p2d]; N and F: (11s6p3d2f)! [5s4p3d2f]).
- For the Cs atom, QZP ANO-RCC basis set ((26s22p15d4f2g)! [9s8p7d3f2g])
- 4 positions: equilibrium position, 2 intermediate distances, and far distance (≈ diss. limit)
- blue - MCSCF(CAS); cyan: CCSD(red) CCSD (all), red - DMRG Ê ,
- CC - Coupled Cluster method fails except for F_2 due to presence of strong correlation!







Conclusions

- tree tensor networks - hierarchical tensors provides a stable parametrization with $\mathcal{O}(d(nr^2 + r^3))$ parameters
- curse of dimension -curse of nonlinearity - curse of non convexity can be widely circumvented
- tree structure reduces most tasks to matrix analysis
- geometric optimization techniques can be applied
- HSVD allows a posteriori control
- model order reduction by a factor (compression rate) $10^3 - 10^{10}$ Khoromskij & Oseledets et al. have model computation with a factor up to 10^{20} and more!
- formally the complexity is reduced $\mathcal{O}(N) \Rightarrow \mathcal{O}(\log^\alpha N)$
- we need either a priory error analysis or a posteriori error control to justify the (somehow astronomical) rates

Historical comparison of related topics

- statistics: Hidden Markov Models (60s) ???
- condensed matter physics: renormalization group: Wilson (70s) Nobel price
- Spin systems (AKLT 87)
- quantum lattice (spin) systems: DMRG White (91) and Ostlund & Rommer (94)
- finitely correlated states: Fannes, Nachtergale & Werner (92)
- molecular quantum dynamics: Meyer, (Cederbaum) et al., Thoss & Wang (2003)
- quantum computing: Vidal, Cirac, Verstraete, (2003)
- hierarchical Tucker format: Hackbusch & Kühn (HT) (2009)
- tensor trains: Oseledets & Tyrtyshnikov (TT) (2009)

High dimensional PDEs and signal analysis

- spin systems and quantum information theory: Cirac, Verstraete, Schollwöck, ...
- Fokker Planck, chemical master equation: Oseledets & Khoromskij & Dolgov , Kazeev & Schwab,
- quantum-chemistry: G. Chan (Princeton), Reiher & Legeza & Verstraete (& S.) , Yanai ...
- uncertainty quantification: Falco & Nouy, Ehrlacher ....
- machine learning etc.: Laathawer, Kolda, Cichocki et al. ...

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, Kressner, ...
- Falco, Nouy, Ehrlacher ....
- Physics: Cirac, Verstraete, Schollwöck, Legeza, G. Chan, Eisert, Hastings, Kitaev, ...

Thank you
for your attention.