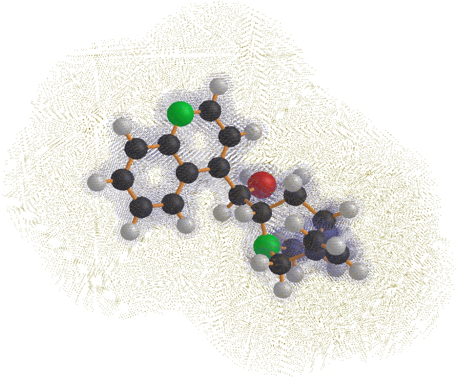
Dynamical Low Rank Approximation for Electronic Structure Calculation

R. Schneider (TUB Matheon)

IPAM 2013

I.

High-dimensional problems



Motivation 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,

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

  describing mechanical systems in stochastic environment,

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

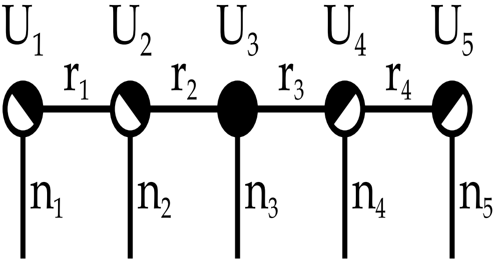
- parametric PDEs (stochastic PDEs)

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

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

I.

Tensor product approximation



Tensors

Goal: Generic perspective on methods for high-dimensional problems, i.e. problems posed on tensor spaces,

Notation: $\boxed{\mathbf{x} = (x_1, \ldots, x_d) \mapsto U(\mathbf{x}) = U(x_1, \ldots, x_d) \in \mathcal{H}}$,

$x_i = 1, \ldots, n_i$ - indices, discrete variables, $\mathcal{H} = \bigotimes_{i=1}^{d} \mathbb{R}^{n_i}$,

Main problem:

$$\boxed{\dim \mathcal{H} = n_1 \times \cdots \times n_d = \mathcal{O}(n^d) \quad -- \ \text{Curse of dimensionality!}}$$

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

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

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

As for matrices, incomplete SVD:

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

Tensors

Goal: Generic perspective on methods for high-dimensional problems, i.e. problems posed on tensor spaces,

Notation: $\boxed{\mathbf{x} = (x_1, \ldots, x_d) \mapsto U(\mathbf{x}) = U(x_1, \ldots, x_d) \in \mathcal{H}}$,

$x_i = 1, \ldots, n_i$ - indices, discrete variables, $\mathcal{H} = \bigotimes_{i=1}^d \mathbb{R}^{n_i}$,

Main problem:

$$\boxed{\dim \mathcal{H} = n_1 \times \cdots \times n_d = \mathcal{O}(n^d) \quad -- \quad \text{Curse of dimensionality!}}$$

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:

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

Tensors

Goal: Generic perspective on methods for high-dimensional problems, i.e. problems posed on tensor spaces,

Notation: $\boxed{\mathbf{x} = (x_1, \ldots, x_d) \mapsto U(\mathbf{x}) = U(x_1, \ldots, x_d) \in \mathcal{H}}$,

$x_i = 1, \ldots, n_i$ - indices, discrete variables, $\mathcal{H} = \bigotimes_{i=1}^{d} \mathbb{R}^{n_i}$,

Main problem:

$$\boxed{\dim \mathcal{H} = n_1 \times \cdots \times n_d = \mathcal{O}(n^d) \quad -- \quad \text{Curse of dimensionality!}}$$

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:

$$U(x_1, \ldots, x_d) \approx \sum_{k=1}^{r} u_{1,k}(x_1) \cdots u_{d,k}(x_d) = \sum_{k=1}^{r} \left(\otimes_{i=1}^{d} u_{i,k}(x_i) \right).$$

Sparsity versus rank sparsity

Example: Given $(x, y) \mapsto A(x, y)$ be an $n \times n$ matrix $\mathbf{A}$.
If $\mathbf{A}$ is rank one $r = 1$

$$A(x, y) = u(x)v(y) \quad , \quad \mathbf{A} = \mathbf{u} \otimes \mathbf{v} = \mathbf{u}^T\mathbf{v}$$

requires $2n$ coefficients $u(x), v(y), x, y = 1, \dots n$.
A sparse matrix is the linear combination of $\delta_{i,k} = \delta_i \otimes \delta_k$.

$$\mathbf{A} = \sum_{\mu=1}^{r} \alpha_\mu \delta_{i(\mu),k(\mu)} = \sum_{\mu=1}^{r} \alpha_\mu \delta_{i(\mu)} \otimes \delta_{k(\mu)}$$

r sparse matrix are also of rank r. But this is not optimal!
But example $\mathbf{u}^T = (1, \dots, 1), \mathbf{v} := 2\mathbf{u}$ then

$$\mathbf{u} \otimes \mathbf{v} = \begin{pmatrix} 2 & \dots & 2 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ 2 & \dots & 2 \end{pmatrix}$$

is not sparse!

Case $d = 2$ Singular value decomposition (SVD)

(other names: Schmidt decomposition, proper orthogonal dec. (POD), principle component analysis (PCA), Karhunen-Loeve transformation)

The case $d = 2$ is an exceptional, since $\mathcal{V} \otimes \mathcal{W} \simeq \mathcal{L}(\mathcal{V}, \mathcal{W})$, and spectral theory could be applied

Theorem (Eckhardt-Young (36), E. Schmidt (07))

$$\boxed{U(x_1, x_2) = \sum_{k=1}^{r} \sigma_k \big(u_k(x_1) \otimes v_k(x_2)\big) = \sum_{k,j=1}^{r} \big(\tilde{u}(x_1, k) s(k, j) \tilde{v}(x_2, j)\big)}$$

The best rank r approximation of U in $\mathcal{V}, \|.\|_2$ is given by setting $\sigma_{r+1} = \cdots = \sigma_r = 0$.

A multi-linear analog for $d < 2$ to the Schmidt theorem does not really exist! The canonical format leads to deep mathematical difficulties and problems. See recent book of J. Landsberg.

Tensors in Electronic Structure Theory

If $V_\nu := \mathbb{R}^{n_\nu}$ a tensor is a multi-indexed vector e.g.

$$U(k_1, \dots, k_d) = U_{k_1,\dots,k_d} \quad , \quad k_i = 1 \dots, n_\nu \ .$$

Examples:

- vector $(u_k) \simeq u(k)$,
- matrix $\left(u_{k_1,k_2}\right) \simeq u(k_1, k_2)$,
- electron integrals h_p^r, $g_{p,q}^{r,s}$; or amplitudes $t_{a,b}^{i,j}$, $t_{a,b,c,d}^{i,j}$; etc.
- reduced density matrices d_p^a, $d_{p,q}^{a,b}$, $d_{p,q,r}^{a,b,c}$
- or in field representation e.g. $D(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \mathbf{r}'_1, s_1, \mathbf{r}'_1, s_1)$
- dielectric tensor and linear response
- (single particle) Greens function $d = 7$

Low rank techniques in quantum chemistry: density fitting, Cholesky decomposition, (see also recent work of T. Martinez et al., Auer & Espig & Hackbusch),
mean field theory (rank one!), e.g. Hartree Fock, Hartree approximation

Anti-symmetry - Second quantization

Basis set $\varphi_k : k = 1, \ldots, d\}$ single particle ON basis, e.g. MO

Let $\Psi_\mu := \Psi_{SL}[\varphi_{k_1}, \ldots, \varphi_{k_N}] = \Psi[k_1, \ldots, k_N]$ be Slater determinants.

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

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

- $\mu_i = 1$ 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 $\dim \mathcal{F}_d = 2^d$, $(\mathbb{K} := \mathbb{C}, \mathbb{R})$

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

$$\boxed{\mathcal{F}_d \simeq \{U : \mu \mapsto U(\mu_1, \ldots, \mu_d) = U_\mu \in \mathbb{K} \ , \ \mu_i = 0,1 \ \} = \bigotimes_{i=1}^{d} \mathbb{K}^2 =: \mathcal{H}}$$

Second Quantization is a basis dependent formalism $\Rightarrow$:

Fermions in Fock space - second quantization

$$|u\rangle = U \in \mathcal{H} = \bigotimes_{j=1}^{d} \mathbb{C}^2 \simeq \mathbb{C}^{(2^d)} \quad , \quad \big(\text{number of particles } N \leq d \to \infty\big)$$

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

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 (discrete) creation operators

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

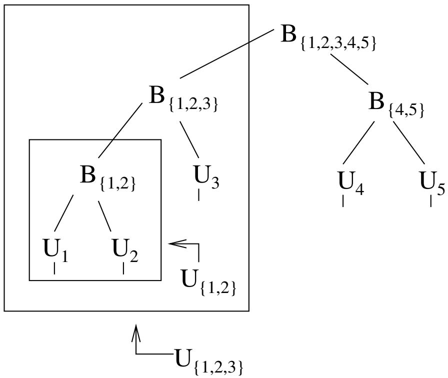
The (discrete) Schrödinger equation is

$$\mathbf{H}U = EU \;\text{ or }\; \frac{dU}{dt} = \frac{1}{i}\mathbf{H}U \;,\; U \in \bigotimes_{j=1}^{d} \mathbb{C}^2 \simeq \mathbb{C}^{2^d} .$$

$$\mathbf{H} = \sum_{p,q=1}^{d} h_p^q \mathbf{a}_p^T \mathbf{a}_q + \sum_{p,q,r,s=1}^{d} g_{r,s}^{p,q} \mathbf{a}_r^T \mathbf{a}_s^T \mathbf{a}_p \mathbf{a}_q \; (\in \mathbb{C}^{2^d \times 2^d}) .$$

I.

Novel tensor formats



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

Subspace approximation for model reduction

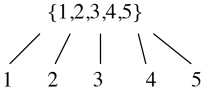
Subspace approximation for model reduction

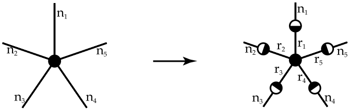
- Tucker format (Q: MCTDH(F)) - robust and stable But complexity $\mathcal{O}(r^d + ndr)$

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

Univariate bases $x_i \mapsto \left(U_i(k_i, x_i)\right)_{k_i=1}^{r_i}$

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





Subspace approximation for model reduction

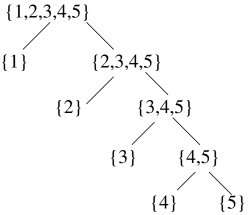
- ▷ Tucker format (Q: MCTDH(F)) - robust and stable But complexity $\mathcal{O}(r^d + ndr)$

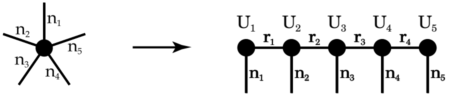
  Is there a similar tensor format, but polynomial in d?

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

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

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





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, Chan, 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

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

TT Tensors, matrix product states - reconstruction

Matricisation or unfolding

$$\mathbf{A}_{(x_1),(x_2,\ldots,x_d)}$$

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

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

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

Matrix product states - reconstruction

Matricisation or unfolding

$\mathbf{A}_{(x_1,x_2),(x_3,\ldots,x_d)}$

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

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

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

TT Tensors, matrix product states - reconstruction

Matricisation or unfolding

$$\mathbf{A}_{(x_1 \dots, x_{d-1}),(x_d)}$$

The tensor $\mathbf{x} \to U(\mathbf{x})$

$$U(x_1, \dots, x_d) = \mathbf{U}_1(x_1) \cdots \mathbf{U}_i(x_i) \cdots \mathbf{U}_d(x_d)$$

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

with matrices or component functions

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

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

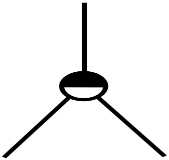
Recursive definition by bases representations

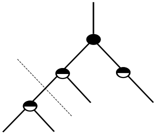
Constructing a basis for

$$V_\alpha \subset V_{\alpha_1} \otimes V_{\alpha_2}$$

(hopefully $\dim V_\alpha << \dim V_{\alpha_1} \times \dim V_{\alpha_2}$)
by contracting tensor product bases

$$\begin{aligned} V_\alpha &= span\{\mathbf{b}_i^{(\alpha)} : 1 \leq i \leq r_\alpha\} \\ \mathbf{b}_\ell^{(\alpha)} &= \sum_{i=1}^{r_{\alpha_1}} \sum_{j=1}^{r_{\alpha_2}} c_{i,j}^{(\alpha,\ell)} \mathbf{b}_i^{(\alpha_1)} \otimes \mathbf{b}_j^{(\alpha_2)} \qquad (\alpha_1, \alpha_2 \text{ sons of } \alpha \in T_D). \end{aligned}$$





like e.g. geminals contracted from gaussians

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

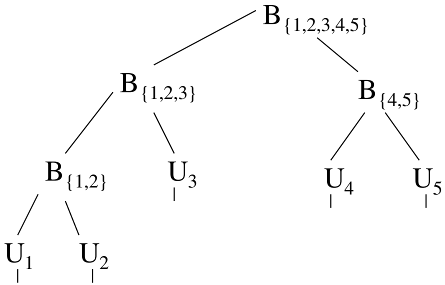
Hierarchical tensor (HT) format

- 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

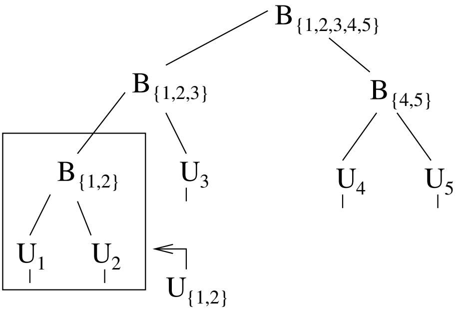
- ▷ Canonical decomposition — not closed, no embedded manifold!
- ▷ Subspace approach (Hackbusch/Kühn, 2009)



(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)

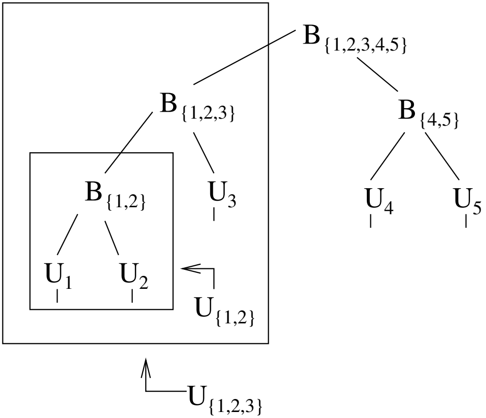


(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n\times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t\times k_{t_1}\times k_{t_2}}$)

Hierarchical tensor (HT) format

- ▷ <u>Canonical decomposition</u> not closed, no embedded manifold!

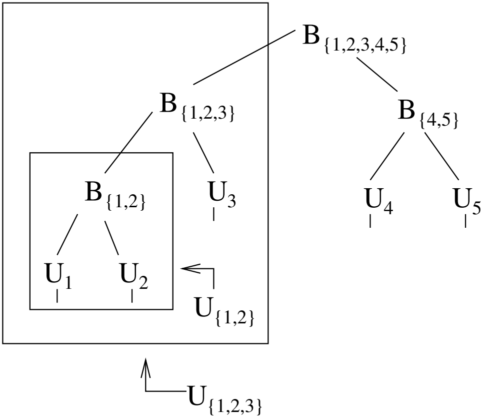
- ▷ Subspace approach (Hackbusch/Kühn, 2009)



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

Hierarchical tensor (HT) format

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



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

Ranks and error control

Grouping indices $\mathcal{I} = \{x_{i_1}, \ldots, x_{i_l}\} \subset \mathcal{I}_d = \{x_1, \ldots, x_d\}$ into row or column index of $\mathbf{A}_{\mathcal{I}} \Rightarrow$ matricisation or unfolding of

$$(x_1, \ldots, x_d) \mapsto U(x_1, \ldots, x_d) \simeq A_{\mathcal{I}}^{\mathcal{I}_d \setminus \mathcal{I}} \ ,$$

Ranks are multi-integer $\underline{r} = (r_1, \ldots, r_d)$

- Tucker : $r_i = \text{rank}\ A_{x_i}^{\mathcal{I}_d \setminus \{x_i\}}$
- TT format $r_i = \text{rank}\ A_{x_1,\ldots,x_i}^{x_{i+1},\ldots,x_d}$, *entanglement* $e_i := r_i - 1$,
- HT format $r_{\mathcal{I}} = \text{rank}\ A_{\mathcal{I}}^{\mathcal{I}_d \setminus \mathcal{I}}$,

Tool: HOSVD (Higher order SVD) - error control by truncating SVDs quasi-optimal error estimates

$$\|U - U_{best}\| \leq \|U - U_{\epsilon(HOSVD)}\| \leq C\sqrt{d}\|U - U_{best}\|$$

Entanglement- size consistancy

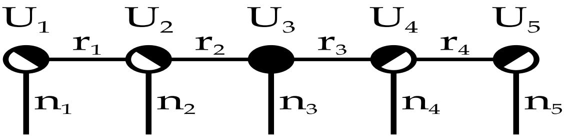
$\mathcal{F}_{A_i}$ = generated by $\{\varphi_1 \ldots \varphi_i\}, \mathcal{F}_{B_i}$ = generated by $\{\varphi_{i+1} \ldots \varphi_d\}$

Theorem (Separation Theorem)

E.g for matrix product states (TT), $x_i = 0, 1$, the rank r_i, $i = 1, \ldots, d-1$, is the rank of $\left(U_{(x_1,\ldots,x_i);(x_{i+1},\ldots,x_d)} \right)$, i.e.

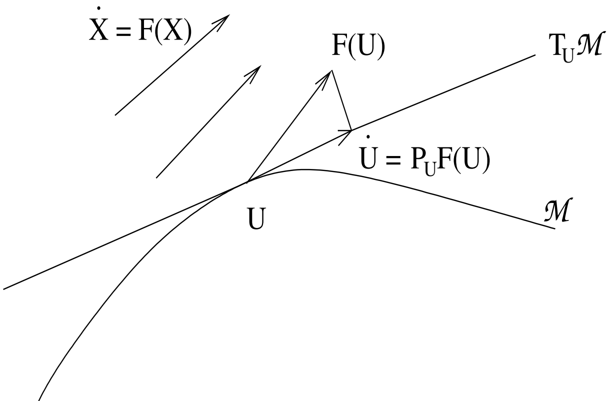
$$U(x_1, \ldots, x_d) = \sum_{k_i=1}^{r_i} A(x_1, \ldots, x_i, k_i) B(k_i, x_{i+1}, \ldots, x_d)$$

Sep. into sub-systems $(x_1, \ldots, x_i), (x_{i+1}, \ldots, x_d)$, $\mathcal{H} = \mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}}$.
If A_i, B_i are independent $\rightarrow$ $r_i = 1$ - size consistency.



II.

Dynamical Low Rank Approximation - TT resp. HT Tensors



HT: Geometry and closedness

Redundancy: we explain TT as model example *Lubich et al., Rohwedder & Sch.*
, Uschmajev, Vandereycken, Haegemann, Verstraete & Cirac

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

(with ranks $\mathbf{r} = (r_1, \ldots, r_{d-1})$) .
Identifying those representations which provides the same tensor, i.e.

$$\mathbf{U}_i(x_i) \sim \mathbf{G}_1^{i-1}\mathbf{U}_2(x_i)\mathbf{G}_i$$

defines an embedded Riemannian manifold

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

$$\boxed{\mathcal{M}_{\leq\mathbf{r}} = \bigcup_{s_i \leq r_i} \mathcal{M}_{\mathbf{s}} = \overline{\mathcal{M}_{\mathbf{r}}} \subset \mathcal{H} \text{ is (weakly) closed!}}$$

HT: Geometry and closedness

Redundancy: we explain TT as model example *Lubich et al., Rohwedder & Sch., Uschmajev, Vandereycken, Haegemann, Verstraete & Cirac*

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

(with ranks $\mathbf{r} = (r_1, \ldots, r_{d-1})$) .
Identifying those representations which provides the same tensor, i.e.

$$\mathbf{U}_i(x_i) \sim \mathbf{G}_1^{i-1}\mathbf{U}_2(x_i)\mathbf{G}_i$$

defines an embedded Riemannian manifold

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

$$\boxed{\mathcal{M}_{\leq\mathbf{r}} = \bigcup_{s_i\leq r_i} \mathcal{M}_{\mathbf{s}} = \overline{\mathcal{M}_{\mathbf{r}}} \subset \mathcal{H} \text{ is (weakly) closed!}}$$

Variational framework

Example

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

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

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

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

   here

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

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 $A : \mathcal{V} \to \mathcal{V}'$,

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

Variational framework

Problem (Original optimization problem (OP))
Given a convex cost functional $\mathcal{J} : \mathcal{H} \to \mathbb{R}$ finding

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

Problem (Tensor product optimization problem (TOP))
Given a model class $\mathcal{M}_{\mathbf{r}}$ - tensors of rank at most $\mathbf{r}$, find the (quasi-) optimal tensor within this class

$$U := \textit{argmin}\, \{\mathcal{J}(W) : W \in \mathcal{M}_{\mathbf{r}}\} \qquad (1)$$

Espig, Hackbusch & Sch.,

Variational framework

Problem (Original optimization problem (OP))

Given a convex cost functional $\mathcal{J} : \mathcal{H} \to \mathbb{R}$ *finding*

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

Problem (Tensor product optimization problem (TOP))

Given a model class $\mathcal{M}_{\mathbf{r}}$ *- tensors of rank at most* $\mathbf{r}$*, find the (quasi-) optimal tensor within this class*

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

Espig, Hackbusch & Sch.,

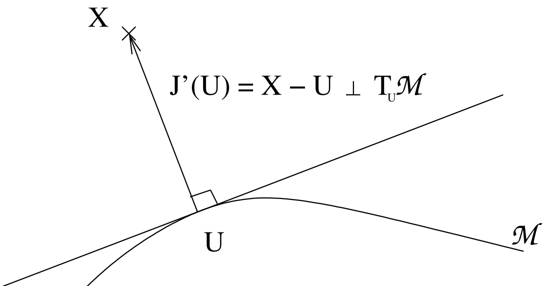
Approximation on low-rank manifold $\mathcal{M}_{\underline{r}} \subseteq \mathcal{H}$

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

Solve first order condition $\mathcal{J}'(U) = 0$ on tangent space $\mathcal{T}_U$, using the Riemannian gradient $\mathrm{Grad}_{\mathcal{M}}$

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

(Dirac-Frenkel variational principle, Khoromskij, Oseledets & Sch. , Dolgov & Sebastianov , Uschmajew & Sch., Vandereycken ... . . .)



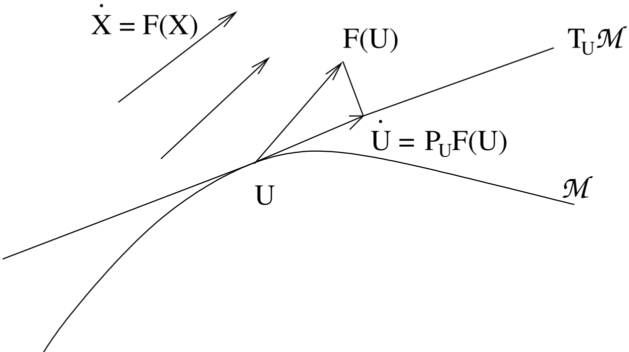
Approximation on low-rank manifold $\mathcal{M}_{\underline{r}} \subseteq \mathcal{H}$

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

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

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

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



Dynamical low rank approximation

Iteration scheme - time stepping

$$U_{n+1} := U_n + F(U_n) \in \mathcal{H}$$

For computation inside $\mathcal{M}_{\mathbf{r}}$

$$\begin{aligned} Y_{n+1} &:= U_n + F(U_n) \in \mathcal{H} \\ U_{n+1} &:= \mathcal{P}_{\mathcal{M}_{\mathbf{r}}}(Y_{n+1}) \,. \end{aligned}$$

$\mathcal{P}_{\mathcal{M}_{\mathbf{r}}}(Y_{n+1})$ can be realized in two steps

$$\begin{aligned} Z_{n+1} &:= U_n + P_{\mathcal{T}_{U_n}} F(U_n) = U_n + \xi_n \\ U_{n+1} &:= \mathcal{R}(U_n, \xi_n) \,. \end{aligned}$$

$P_{\mathcal{T}_{U_n}}$ is the linear projection onto the tangent space at U_n.
$\mathcal{R}$ - retraction onto the manifold, simple to realize.
In order not to get trapped into local minima apply greedy techniques, e.g. (E. Cances & V. Ehrlacher & T. Lelievre)

Dynamical low rank approximation

Iteration scheme - time stepping

$$U_{n+1} := U_n + F(U_n) \in \mathcal{H}$$

For computation inside $\mathcal{M}_{\mathbf{r}}$

$$\begin{aligned} Y_{n+1} &:= U_n + F(U_n) \in \mathcal{H} \\ U_{n+1} &:= \mathcal{P}_{\mathcal{M}_{\mathbf{r}}}(Y_{n+1}) \,. \end{aligned}$$

$\mathcal{P}_{\mathcal{M}_{\mathbf{r}}}(Y_{n+1})$ can be realized in two steps

$$\begin{aligned} Z_{n+1} &:= U_n + P_{\mathcal{T}_{U_n}} F(U_n) = U_n + \xi_n \\ U_{n+1} &:= \mathcal{R}(U_n, \xi_n) \,. \end{aligned}$$

$P_{\mathcal{T}_{U_n}}$ is the linear projection onto the tangent space at U_n.
$\mathcal{R}$ - retraction onto the manifold, simple to realize.
In order not to get trapped into local minima apply greedy techniques, e.g. (E. Cances & V. Ehrlacher & T. Lelievre)

Dynamical low rank approximation

Iteration scheme - time stepping

$$U_{n+1} := U_n + F(U_n) \in \mathcal{H}$$

For computation inside $\mathcal{M}_{\mathbf{r}}$

$$\begin{aligned} Y_{n+1} &:= U_n + F(U_n) \in \mathcal{H} \\ U_{n+1} &:= \mathcal{P}_{\mathcal{M}_{\mathbf{r}}}(Y_{n+1}) \, . \end{aligned}$$

$\mathcal{P}_{\mathcal{M}_{\mathbf{r}}}(Y_{n+1})$ can be realized in two steps

$$\begin{aligned} Z_{n+1} &:= U_n + P_{\mathcal{T}_{U_n}} F(U_n) = U_n + \boldsymbol{\xi}_n \\ U_{n+1} &:= \mathcal{R}(U_n, \boldsymbol{\xi}_n) \, . \end{aligned}$$

$P_{\mathcal{T}_{U_n}}$ is the linear projection onto the tangent space at U_n.
$\mathcal{R}$ - retraction onto the manifold, simple to realize.
In order not to get trapped into local minima apply greedy techniques, e.g. (E. Cances & V. Ehrlacher & T. Lelievre)

Simple local optimization for TT (HT) tensors

Alternating directional search - ALS

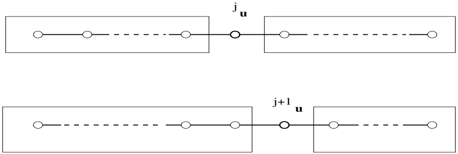
Relaxation (see e.g. Gauss-Seidel, ALS):

For $j = 1, \ldots, d$:

1. fix all component tensors U_ν, $\nu \in \{1, \ldots, d\} \backslash \{j\}$, except index j, (to be the root!).

2. Optimize $U_j(k_{j-1}, x_j, k_j)$, and orthogonalize left

Repeat the relaxation procedure (in the opposite direction.)



S. Holtz, Rohwedder & Schneider (SISC 2012), Uschmajew & Rohwedder, Uschmajew & Schneider (in prep.)

This is the single site DMRG (S. White (1991)) algorithm!

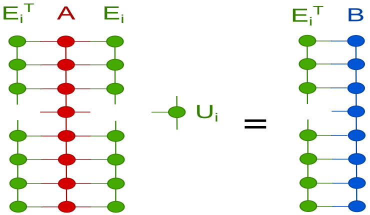
ALS (single site DMRG)

$$U = \text{argmin}\{\frac{1}{2}\langle HU, U\rangle : \langle U, U\rangle = 1 \ , \ U \in \mathcal{M}_{(\leq)\underline{r}}\} \ ,$$

$$U(\mathbf{x}) = U_1(x_1) \cdots U_i(x_i) \cdots U_d(x_d)$$

Optimizing U_i, resp. $U_i(k_{i-1}, x_i, k_i) \Rightarrow$ eigenvalue problem

$\boxed{\widetilde{H}_i U_i = E \widetilde{U}_i}$, in the small space $\mathbb{K}^{r_{i-1} \times n_i \times r_i}$



Example: $AU = F$

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 $r_c \geq$	defined r_T	defined $r_T \leq r_{HT} \leq r_c$
(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

QC (quantum chemistry)- DMRG

MPS (TT) approximation in 2nd quantization:

1. approximates the FCI wave function (in a controllable way), it permeates the full FCI space
2. complexity (storage) : $2\sum_{i_1}^{d} r_{i-1} r_i = \mathcal{O}(r^2 d)$
3. electron density and reduced density matrices can be computed $\Rightarrow$ gradients, forces, ...
4. subspaces, mixed states, (e.g. multiple eigenvalues, degeneracy and near degeneracy) can be computed
5. time dynamics and thermal problems can be treaten
6. size consistency and black box (yes and no)

- S. White (1991) spin chains
- S. Östlund, S. Rommer (1994), matrix product states
- S. White & Martin (1999) quantum chemistry
- O. Legeza, J. Röder and B. Hess (2003),
- G. Chan & Head-Gordon (2002) electronic structure computation
- G. Moritz, B. Hess and M. Reiher (2005)
- T. Yania, G. Chan with others (2008)
- D. Zgid, M. Noijen (2008)
- G. Chan et al. (–2013)

DMRG provides an (approximate) FCI algorithm in polynomial complexity

dynamical correlation (short range - ee-cusp)	statical or strong correlation (multi-configurational)
Coupled Cluster (CCSD-)	CASPT-MRSCF - DMRG

Historical comparison of related topics

Table: quantum physics - numerical mathematics

topic	mathematics	physics
MPS	AKLT (1988) , Östlund &Rommer (1994)	TT (Oseledets & T. 2009)
DMRG	S. White (1992)	ALS - Rohwedder, S. (2010)
		Uschmajev - Rohw. (2012)
HT	Mayer et al. (2000), Wang & Thoss (2004)	Hackbusch & Kühn (2009)
TNS	Conference Leyden (2004)	
	Vidal, Verstraete, I. Cirac, U. Schollwöck	
	Block renormalization (60')	hidden Markov models (TT)
Renorm. th.	Wilson (Nobel pr. 82) , I. Peschel ...	
QC DMRG	G. Chan , O. Legeza (2003)	S. (2012)
HOSVD	Vidal (2003)	Grasedyck (2010)
manifolds	Haegeman, Cirac, Verstraete (2010)	Rohwedder & S. et al. (2011)
		Ushmajev et al. (2012)
conv. rates	Renyi entropies (2008)	Uschmajev, S. (2013)
regularity	Area law -	

These development have been independent, and only related in content!

I. Cirac received the Wolfe price in physics 2013!

Enforcing symmetries

Example: Quantum number N, particles
Let us consider matricisation $\mathbf{A}_j$ separating system
$A \simeq \mu_A = (x_1, \ldots, x_j)$ and $B \simeq \mu_B = (x_{j+1}, \ldots, x_d)$
Any eigen state C of $\mathbf{P} = \sum_{i=1}^{d} \mathbf{a}_j^\dagger \mathbf{a}$ has form $C = C_A \otimes C_B$

$$C(\mu_A, \mu_B) = C_A(\mu_A) C_B(\mu_B) \text{ and } N_A + N_B = N_{total} \ .$$

Grouping $k_{j,i} : i \in \mathcal{I}_{N_A}$, $N_A = 0, \ldots, N \ \Rightarrow \ N_B = N - N_A$
$\Rightarrow$

$$U(k_{j-1}, 0, k_j) \neq 0 \ \text{ iff } \ k_{j-1}, k_j \in \mathcal{I}_{N_A}$$

$$U(k_{j-1}, 1, k_j) \neq 0 \ \text{ iff } \ k_{j-1} \in \mathcal{I}_{N_A} \ , k_j \in \mathcal{I}_{N_A+1}$$

Works also for spin S_z, quantum numbers m. S^2 is more difficult but possible.

$$\sharp \ nnz \ U(k_{j-1}, x_j, k_j) \leq c r_{j-1} r_j \ \text{ but } \ c << 2 \ .$$

Comparison with Coupled Cluster

The multiplicative representation of the (single reference) CC ansatz, e.g. CCD,

$$\begin{aligned} U &= e^{T_2} U_0 = e^{\sum_{a<b,i<j} t_{i,j}^{a,b} \mathbf{a}_a^\dagger \mathbf{a}_b^\dagger \mathbf{a}_j \mathbf{a}_i} U_0 \\ &= \Pi_{i<j,a<b} (\mathbf{I} + t_{i,j}^{a,b} \mathbf{a}_a^\dagger \mathbf{a}_b^\dagger \mathbf{a}_j \mathbf{a}_i) U_0 \ , \end{aligned}$$

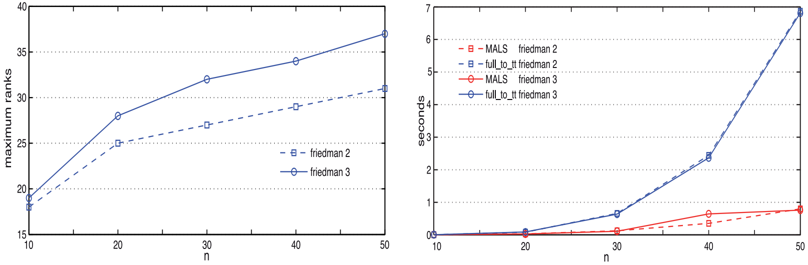
shows that CCD can even represent some highly entangled states in a data sparse way.
Notice that, an m fold product of rank two operators can have rank 2^m!.
Similar conclusion holds for *CCS* which corresponds to a basis transformation!

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



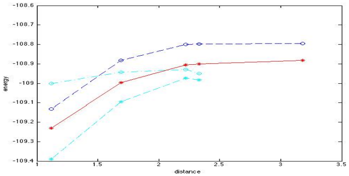
`full_to_tt` (Oseledets, successive SVDs)

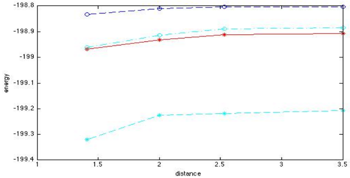
and `MALS` (with A = I) (Holtz & Rohwedder & S.)

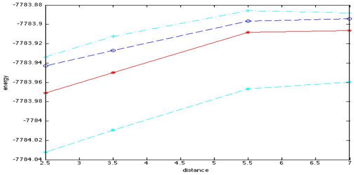
Dissoziation of a diatomic molecules

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 ($\approx$ diss. limit)
- blue - MCSCF(CAS); cyan: CCSD(red) CCSD (all), red - DMRG



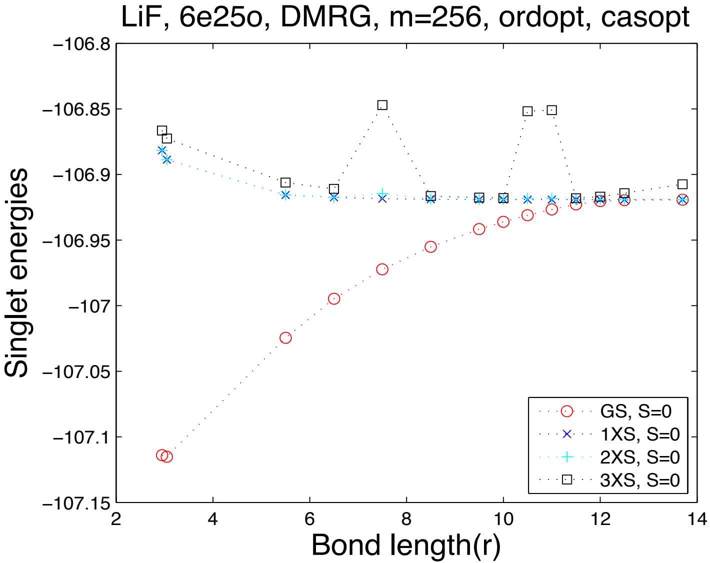




QC-DMRG for HT - tree tensor netowrks

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

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



Parabolic PDEs

joint paper with B. Khoromskij, I. Oseledets

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

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

Next slide: Schrödinger type equation :

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

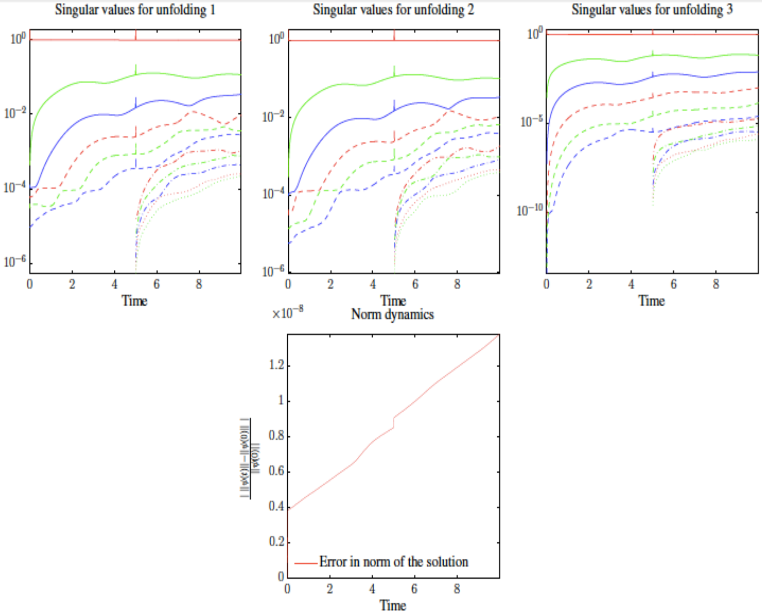


Figure 4: The dynamics of the singular values of unfoldings with a bump from $r = 5$ to $r = 10$ at $t = 5$. ($d = 4$, $\tau = 0.025$).

Thank you
for your attention.