Low rank aproximation in traditional and novel tensor formats

R. Schneider (TUB Matheon)

MPI Munich, 2012

Our motivations

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

- stationary/instationary Schrödinger type equations

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

  describing quantum-mechanical many particle systems
- stochastic DEs and the Fokker-Planck equation,

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

  describing mechanical systems in stochastic environment,
- chemical master equations, parametric PDEs, machine learning, ...

Solutions depend on $x = (x_1, \dots, x_d)$, where usually, $d >> 3$!

Setting - Tensors of order d

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

$$\mathcal{V} := \bigotimes_{i=1}^{d} V_i, \qquad \text{today:} \quad \mathcal{V} = \bigotimes_{i=1}^{d} \mathbb{R}^n = \mathbb{R}^{(n^d)}$$

Notation: $(x_1, \ldots, x_d) \mapsto U = U(x_1, \ldots, x_d) \in \mathcal{V}$

Main problem:

$dim \mathrm{V} = \mathrm{O}\,(\mathrm{n}^d)$ – Curse of dimensionality!

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

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

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

As for matrices, incomplete SVD:

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

Canonical format

$$\mathcal{H} \simeq \{\mathbf{x} = (x_1, \ldots, x_d) \mapsto U(x_1, \ldots, x_d) \in \mathbb{R}, x_i = 1, \ldots, n_i\} \ .$$

Single tensor product

$$(x_1, \ldots, x_d) \mapsto U(x_1, \ldots, x_d) = \Pi_{i=1}^d u_{i,k}(x_i) = \Pi_{i=1}^d u_i(x_i, k) \ ,$$

$$U = \bigotimes_{i=1}^{d} \mathbf{u}_{k,i} \ .$$

Canonical (CP) format, PARAFAC or r-term expansion,

$$\boxed{U(x_1, \ldots, x_d) = \sum_{k=1}^{r_C} U_k(\mathbf{x}) = \sum_{k=1}^{r_C} \Pi_{i=1}^d u_{i,k}(x_i).}$$

or

$$\boxed{U = \sum_{k=1}^{r_C} U_k = \sum_{k=1}^{r_C} \bigotimes_{i=1}^{d} \mathbf{u}_{i,k}}$$

Canonical format - pros

Definition (Canonical format)

$$U(x_1, \ldots, x_d) = \sum_{k=1}^{r} \bigotimes_{i=1}^{d} u_{i,k}(x_i) = \sum_{i=1}^{r} \bigotimes_{\nu=1}^{d} u_i(x_i, k).$$

r - canonical rank (?)

Let $n := \max\{n_i : 1 \leq i \leq d\}$.

- Number of terms $r = r_c$ (canonical rank (?)) is invariant w.r.t. basis transformations
- canonical rank $r \leq \sharp$ DOF for a given tensor product basis - best N-term approximation *(super adaptivity)!*
- there is an additional cost storing the components $\mathbf{u}_{i,k_i}$
- degrees of freedom (DOF) or better storage complexity: $\boxed{\mathcal{O}(drn)}$
- complexity scaling is $\mathcal{O}(drn)$ instead of $\mathcal{O}(n^d)$ for the full tensor!

Counter example of Silva and Lim

Let $A = \bigotimes_{i=1}^{d} \mathbf{a}_i$, $B = \bigotimes_{i=1}^{d} \mathbf{b}_i \in \mathcal{H}$, (possibly $\langle \mathbf{a}_i, \mathbf{b}_i \rangle = 0 \ \forall i$).

$$
\begin{aligned}
\text{Let} \quad & U(x_1, \ldots, x_d) \\
:= \quad & U_1(x_1, \ldots, x_d) + \ldots + U_d(x_1, \ldots, x_d) \ , \ (\ U_i \perp U_j \ ) \\
:= \quad & b_1(x_1) a_2(x_2) \ldots a_d(x_d) + \cdots + a_1(x_1) \ldots a_{d-1}(x_{d-1}) b_d(x_d) \\
= \quad & \frac{1}{\epsilon} (a_1(x_1) + \epsilon b_1(x_1)) \cdots (a_d(x_d) + \epsilon b_d(x_d)) \\
& - \frac{1}{\epsilon} a_1(x_1) \cdots a_d(x_d) + \mathcal{O}(\epsilon) \\
=: \quad & V_\epsilon(x_1, \ldots, x_d) + \mathcal{O}(\epsilon) \ , \ \text{product -rule for } A' \ .
\end{aligned}
$$

$V_\epsilon \to U$ as $\epsilon \to 0$, but rank $r_c(U) = d \neq r_c(V_\epsilon) = 2$ if $d \geq 3$!!!

$$\Rightarrow$$

- $\mathcal{K}^{\leq r} := \{ U \in \mathcal{H} : U = \sum_{k=1}^{r} U_k \}$ is not closed.
  (nor weakly closed)
- The notion of canonical rank is not well defined! Border rank problem.

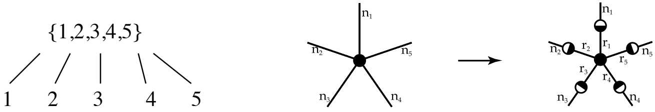
Remark: The above example shows the product rule for the directional derivative

Tensor formats

Tensor formats

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

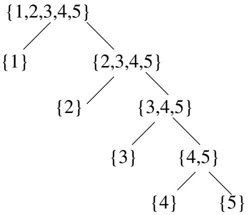
$$U(x_1, .., x_d) = \sum_{k_1=1}^{r_1} \dots \sum_{k_d=1}^{r_d} B(k_1, .., k_d) \bigotimes_{i=1}^{d} \mathbf{U}_i(k_i, x_i)$$

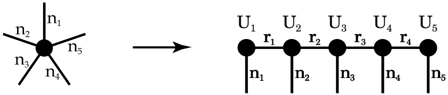


Tensor formats

- ▷ Hierarchical Tucker format
  (HT; Hackbusch/Kühn, Grasedyck, Kressner, Q: Tree-tensor networks)
- ▷ Tucker format (Q: MCTDH(F))
  But complexity $\mathcal{O}(r^d + ndr)$
- ▷ Tensor Train (TT-)format
  (Oseledets/Tyrtyshnikov, $\simeq$ MPS-format of quantum physics)

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





Tensor formats

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

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

Tensor formats

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

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

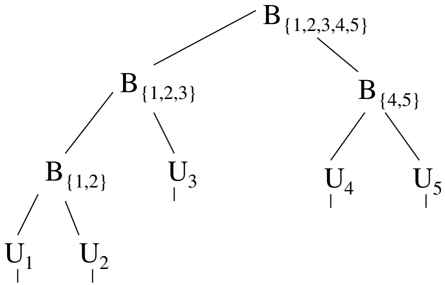
Tensor formats

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

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

Tensor formats

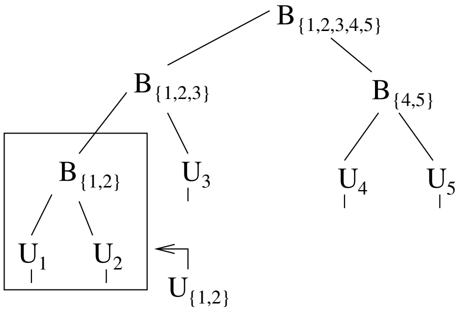
- Canonical decomposition — not closed, no embedded manifold!
- Subspace approach (Hackbusch/Kühn, 2009)



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

Tensor formats

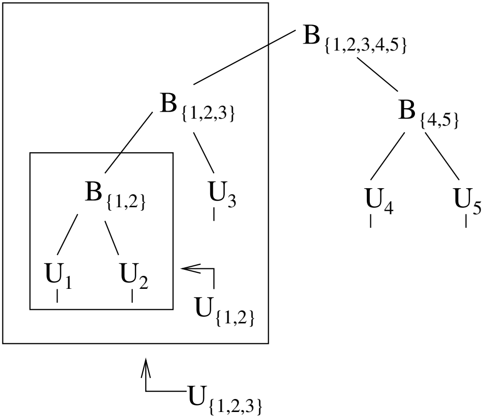
- ▷ Canonical decomposition   not closed, no embedded manifold!
- ▷ Subspace approach (Hackbusch/Kühn, 2009)



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

Tensor formats

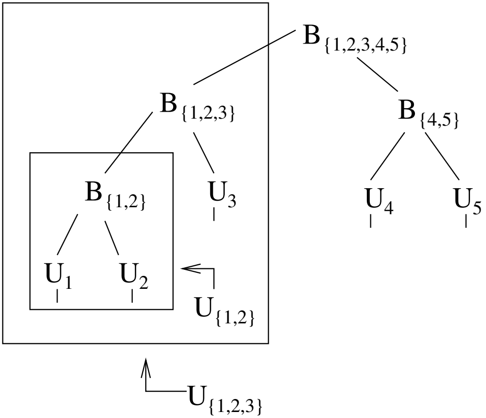
- ▷ Canonical decomposition not closed, no embedded manifold!
- ▷ Subspace approach (Hackbusch/Kühn, 2009)



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

Tensor formats

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



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

Optimization Problems

Problem (Generic optimization problem (OP))

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

$$\mathit{argmin} \left\{ \mathcal{J}(W) : W \in \mathcal{M} \right\} .$$

Problem (Tensor product optimization problem (TOP))

$$U := \mathit{argmin} \left\{ \mathcal{J}(W) : W \in \mathcal{M} \cap \mathcal{K}^{\leq r} \right\} \tag{1}$$

Here we consider a modified optimization problem where the original admissible set is confined to tensors of rank at most r. Most problems can cast into this form.

Example

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

$$\boxed{F(W) = \|U - W\|^2_{\mathcal{H}} == \|U - W\|^2 \ , \ W \in \mathcal{K}^r}$$

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

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

here

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

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

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

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

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

In many cases $\mathcal{M} \cap \mathcal{K}^{\leq r} = \mathcal{K}^{\leq r}$. and most F are quadratic.

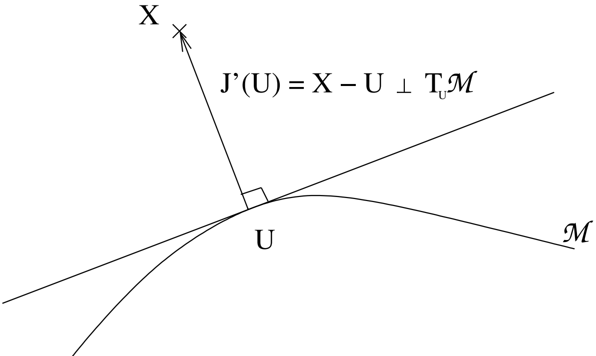
Approximation on low-rank manifold $\mathcal{M} \subseteq \mathcal{V}$

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

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, ...)



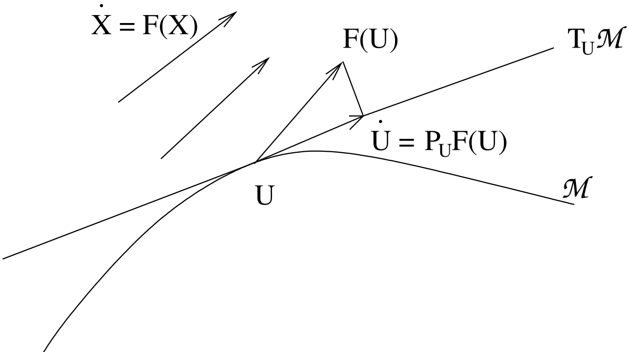
Approximation on low-rank manifold $\mathcal{M} \subseteq \mathcal{V}$

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

Solve projected DE, $\dot{U} = P_U f(U), U(0) = U_0 \in \mathcal{M}$,

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

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



Unique representation of the tangent space of $\mathbb{T}^{\mathbf{r}}$

Theorem (Holtz, R. Schneider, 2010)

For all $U \in \mathbb{T}$, and for

$$\delta U \in \mathcal{T}_U\mathbb{T} \simeq \Big\{\gamma'(t)|_{t=0} \mid \gamma \in C^1(]-\delta,\delta[,\mathbb{T}), \quad \gamma(t) = \mathbf{U}_1(x_1,t)\cdot\ldots\cdot\mathbf{U}_d(x_d,t),\ \gamma(0) = U(\underline{x})\Big\}$$

there is a unique vector $(\mathbf{W}_1 \ldots, \mathbf{W}_d)$ of component functions $\mathbf{W}_i(\cdot) : \mathcal{I}_i \to \mathbb{R}^{r_{i-1}\times r_i}$, such that

$$\boxed{\delta U = \delta U_1 + \ldots + \delta U_d}$$

with

$$\boxed{\delta U_i := \mathbf{U}_1(x_1)\ldots\mathbf{U}_{i-1}(x_{i-1})\mathbf{W}_i(x_i)\mathbf{U}_{i+1}(x_{i+1})\ldots\mathbf{U}_d(x_d),}$$

and s.t. $\mathbf{W}_i(\cdot)$ fulfil the (left-orthogonality) gauge conditions

$$\boxed{\mathbf{L}(\mathbf{U}_i)^T\mathbf{L}(\mathbf{W}_i) = \mathbf{0} \in \mathbb{R}^{r_i\times r_i} \quad \text{for } i = 1,..,d-1.}$$

Sketch of proof

- Existence:

$$\delta U \simeq \mathbf{U}_1'(x_1,0)\mathbf{U}_2(x_2)\cdot .. \cdot \mathbf{U}_d(x_d) + \mathbf{U}_1(x_1)\mathbf{U}_2'(x_2,0)\cdot .. \cdot \mathbf{U}_d(x_d) + \ldots + \mathbf{U}_1(x_1)\cdot .. \cdot \mathbf{U}_{d-1}(x_{d-1})\mathbf{U}_d'(x_d,0).$$

  Left orthogonal decomposition: there ex. unique Λ_1 s.t.

$$\mathbf{U}'(x_1,0) = \mathbf{U}_1(x_1)\Lambda_1 + \mathbf{W}_1(x_1),$$

  iterate.

- Uniqueness: (Idea from Lubich et al., Tucker format)
  - Testing δU with

$$V_i(x) := \mathbf{U}_1(x_1)\cdot .. \cdot \mathbf{U}_{i-1}(x_{i-1})\mathbf{V}_i(x_i)\mathbf{U}_{i+1}(x_{i+1}) .. \cdot \mathbf{U}_d(x_d),$$

    for $i = 1, .., d$, gauge condition (in other inner product) gives upper block-Δ-system with SPD matrices.
  - $\mathbf{L}(\mathbf{W}_i)$, $i = d, d-1, .., 1$ can uniquely be computed recursively.

Parametrization of $\mathcal{T}_U\mathbb{T}$

- C_i spaces of component functions $\mathbf{U}_i$, $(i = 1, .., d)$
- Left-orthonormal spaces of $\mathbf{U}_i$:

$$U_i^\ell := \{\mathbf{W}_i(x_i) \in C_i, \;\; \mathbf{L}(\mathbf{U}_i)^T \mathbf{L}(\mathbf{W}_i) = \mathbf{0}\}.$$

- Parameter space $X := U_1^\ell \times \ldots \times U_{d-1}^\ell \times C_d$.

Corollary (Holtz, R., Schneider, 2010)

The mapping $\tau : X \to \mathcal{T}_U\mathbb{T}$,

$$\tau(\mathbf{W}_1, \ldots, \mathbf{W}_d) \;\; = \;\; \sum_{i=1}^{d} \mathbf{U}_1 \cdot \ldots \cdot \mathbf{U}_{i-1} \mathbf{W}_i \mathbf{U}_{i+1} \cdot \ldots \cdot \mathbf{U}_d$$

is a linear bijection between X *and* $\mathcal{T}_U\mathbb{T}$*. In particular,*

$$dim\, \mathbb{T} \;\; = \;\; \sum_{i=1}^{d} r_{i-1} n_i r_i - \sum_{i=1}^{d-1} r_i^2.$$

Local parametrization for $\mathbb{T}$

Theorem (Holtz, R., Schneider, 2010)

Let $U \in \mathbb{T}$, $\Psi : X \to \mathcal{H}$ defined by

$$\boxed{\Psi(\mathbf{W}_1, \ldots, \mathbf{W}_d) \;=\; (\mathbf{U}_1 + \mathbf{W}_1)(x_1) \cdot .. \cdot (\mathbf{U}_d + \mathbf{W}_d)(x_d).}$$

- *There exists open $N_\delta(0) \subseteq \mathbb{R}^{\dim X}$ such that*

$$\Psi|_{N_\delta} : \quad N_\delta(0) \mapsto \Psi(N_\delta(0)) \overset{open}{\subseteq} \mathbb{T}$$

  is an embedding (i.e. an immersion that is a homeomorphism onto its image), that is, $N_U \cap \mathbb{T}$ is a regular submanifold of $\mathcal{H}$.

- *There exists $N_\delta \subset \mathcal{H}$ open, a constraint function $g = g_U : N_\delta \to \mathbb{R}^c$, $c = \sum_{i=1}^{d-1} r_i^2$, such that*

$$N_\delta \cap \mathbb{T} \;=\; \{U \in \mathbb{R}^{n^d} : \; g(U) = 0\} \;=\; \psi(N_\delta(0))$$

<u>Proof:</u> Inv. mapping theorem for manifolds, tang. space par. τ

Manifolds and gauge conditions

Lubich et al. (2009), Holtz/Rohwedder/S. (2011a), Uschmajew/Vandereycken (2012), Lubich/Rohwedder/S./Vandereycken (in prep.)

- 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

Lubich et al. (2009), Holtz/Rohwedder/S. (2011a), Uschmajew/Vandereycken (2012), Lubich/Rohwedder/S./Vandereycken (in prep.)

- 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

Lubich et al. (2009), Holtz/Rohwedder/S. (2011a), Uschmajew/Vandereycken (2012), Lubich/Rohwedder/S./Vandereycken (in prep.)

- 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

Lubich et al. (2009), Holtz/Rohwedder/S. (2011a), Uschmajew/Vandereycken (2012), Lubich/Rohwedder/S./Vandereycken (in prep.)

- 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

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

Theorem (Lubich/Rohwedder/S./Vandereycken (in prep.))

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

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

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.).

Differential equations for components under gradient flow

Let $U(t) = \mathbf{U}_1(t)\cdots\mathbf{U}_i(t)\cdots\mathbf{U}_d(t) \in \mathbb{T}^{\mathbf{r}}$ be fixed.
Then

$$\boxed{\mathbf{U}_d'(t) = -\mathbf{E}_d^T(t)\big(f(U(t))\big) \ \in \mathbb{R}^{r_{d-1}\times n_d}\ .}$$

provided that $\mathbf{U}_i(t)$, $i = 1, \ldots, d-1$, are left orthogonal, since $\mathbf{D}_d(t) = \mathbf{I}$.
The other components $\mathbf{U}_i'(t)$, , $i = 1, \ldots, d-1$, are given by

$$\boxed{\mathbf{U}_i'(t) = \Big((\mathbf{I} - \mathbf{P}_i(t)) \otimes \mathbf{D}_i{}^{-1}(t)\Big)\mathbf{E}_i^T(t)\big(f(U(t))\big)\ .}$$

with the orthogonal projection $\mathbf{P}_i(t)$ onto the parameter space,

$$\mathbf{P}_i(t)\mathbf{W}(k_{i-1}, x_i, k_i) =$$
$$= \sum_{k_{i_1}', x_i', k_i'} \mathbf{U}_i(t, k_{i-1}', x_i', k_i')\mathbf{W}(k_{i-1}', x_i', k_i)\mathbf{U}_i(t, k_{i-1}, x_i, k_i')\ .$$

Stabilization and preconditioning

In time step $t \to t + \Delta t$ compute the components
$\mathbf{V}_i(\tau) \approx (\mathbf{I} \otimes \mathbf{D}_i(t))\mathbf{U}_i(\tau), \, , \; i = 1, \ldots, d-1, \, t \leq \tau < t + \Delta t$ by

$$\mathbf{V}_i'(\tau) = (\mathbf{I} \otimes \mathbf{D}_i(t))\mathbf{U}_i'(\tau) = \Big((\mathbf{I} - \mathbf{P}_i(\tau)) \otimes \mathbf{I}\Big)\mathbf{E}_i^T(\tau)\big(f(U(\tau))\big) \; .$$

and $\mathbf{U}_i(t + \delta) = \text{left-orth}\big(\mathbf{V}_i(t + \Delta t)\big)$ I. Oseledets et al. : Strang splitting and alternating directions ALS, (compare TD DMRG) Generalization of HOSVD bases of Hackbusch.

For non-leaf vertices $\alpha \in T_D$, $\alpha \neq D$, we have

$$\sum_{\ell=1}^{r_\alpha} (\sigma_\ell^{(\alpha)})^2 \, C^{(\alpha,\ell)} C^{(\alpha,\ell)\mathsf{H}} = \Sigma_{\alpha_1}^2 ,$$
$$\sum_{\ell=1}^{r_\alpha} (\sigma_\ell^{(\alpha)})^2 \, C^{(\alpha,\ell)\mathsf{T}} \overline{C^{(\alpha,\ell)}} = \Sigma_{\alpha_2}^2 ,$$

where α_1, α_2 are the first and second son of $\alpha \in T_D$ and Σ_{α_i} the diagonal of the singular values of $\mathcal{M}_{\alpha_i}(\mathbf{v})$.

Closedness

Let $\mathbf{A}_i$ with ranks rank$\mathbf{A}_i \leq r$.
If $\lim_{i\to\infty} \|\mathbf{A}_i - \mathbf{A}\|_2 = 0$ then rank$\mathbf{A} \leq r$:
$\Rightarrow$ closedness of Tucker and HT tensor in $\mathbb{T}^{\leq \underline{r}}$ (Falco & Hackbsuch).

$$\boxed{\mathbb{T}^{\leq \underline{r}} = \bigcup_{s \leq r} \mathbb{T}_{\underline{s}} \subset \mathcal{H} \text{ is closed!}}$$

due to Hackbusch & Falco

(Weak) closedness implies the existence of minimers of convex optimization problems constraint to $\mathbb{T}^{\leq \underline{r}}$.
Landsberg & Ye: If a tensor network has not a tree structure, the set of all tensor of this form need not to be closed!

Summary

For Tucker and HT redundancy can be removed (see next talk)

Table: 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_{HT} , $r_T \leq r_{HT}$
closedness	no	yes	yes
essential redundancy	yes	no	no
recovery quasi best approx. best approx.	?? no no	yes yes exist but NP hard	yes yes exist but NP hard

Some current results and trends

Optimization:

- Alternating optimization of components for TT robust practical algorithm (ALS/MALS, Holtz/Rohwedder/S., SISC 2012)
- DMRG $\hat{=}$ MALS sees boost of interest in quantum physics/quantum chemistry community
- Gradient methods — gradient flow (see below)
- (Quasi-) Newton methods on $\mathcal{M}$ (Rohwedder et al. , in prep.)

Time-dependent equations:

- Quasi-optimal error bounds
  (Lubich/Rohwedder/S./Vandereycken, in prep.)
  solution $X(t)$ with approx. $U(t) \in \mathcal{M}_{\underline{r}}$, $X(0) = U(0)$,

$$\|U(t) - U_{\text{best}}(t)\| \lesssim t \cdot \max_{s \in [0,t]} \|U_{\text{best}}(s) - X(s)\|.$$

Some current results and trends

Optimization:

- ▷ Alternating optimization of components for TT robust practical algorithm (ALS/MALS, Holtz/Rohwedder/S., SISC 2012)
- ▷ DMRG $\hat{=}$ MALS sees boost of interest in quantum physics/quantum chemistry community
- ▶ Gradient methods — gradient flow (see below)
- ▷ (Quasi-) Newton methods on $\mathcal{M}$ (Rohwedder et al. , in prep.)

Time-dependent equations:

- ▷ Quasi-optimal error bounds
  (Lubich/Rohwedder/S./Vandereycken, in prep.)
  solution $X(t)$ with approx. $U(t) \in \mathcal{M}_{\underline{r}}$, $X(0) = U(0)$,

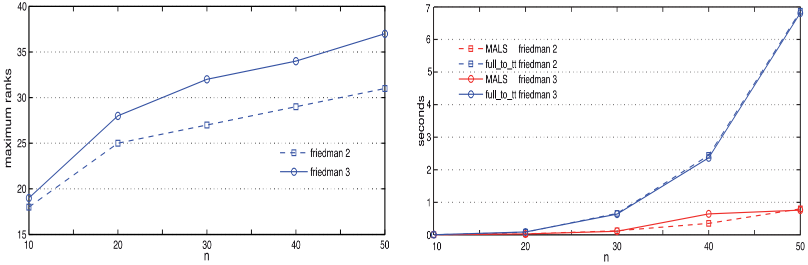
$$\|U(t) - U_{\text{best}}(t)\| \lesssim t \cdot \max_{s \in [0,t]} \|U_{\text{best}}(s) - X(s)\|.$$

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

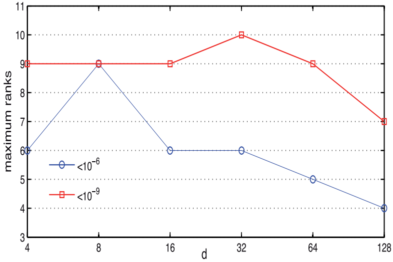


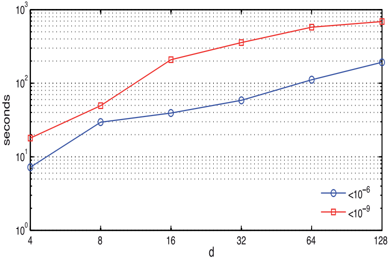
`full_to_tt` (Oseledets, successive SVDs)

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





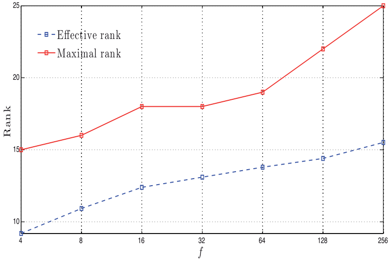
Example: Eigenvalue problem in QTT

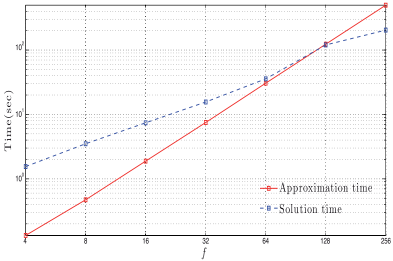
in courtesy of B. Khoromskij, I. Oseledets, *QTT: Toward bridging high-dimensional quantum molecular dynamics and DMRG methods*,

$$H\Psi = (-\frac{1}{2}\Delta + V)\Psi = E\Psi$$

with potential energy surface given by Henon-Heiles potential

$$V(q_1, \ldots, q_f) = \frac{1}{2}\sum_{k=1}^{f} q_k^2 + \lambda \sum_{k=1}^{f-1}\left(q_k^2 q_{k+1} - \frac{1}{3}q_k^3\right).$$





Dimensions $f = 4, \ldots, 256$; 1-D grid size $n = 128 = 2^7 = 2^d$; $\rightsquigarrow$

$$\text{QTT-tensors} \in \bigotimes_{i=1}^{7f} \mathbb{R}^2 = \mathbb{R}^{\overbrace{2\times..\times 2}^{7f=1792}}.$$

QC-DMRG and TT resp, MPS approximations

In courtesy of O Legeza (Hess & Legeza & ..)
LiF dissoziation, 1st + 2nd eigenvalue