Dynamical low rank approximation in hierarchical tensor format

R. Schneider (TUB Matheon)

John von Neumann Lecture – TU Munich, 2012

Motivation

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 $x = (x_1, \ldots, x_d)$, where usually, $d >> 3$!

Setting - Tensors

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

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

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

Main problem:

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

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:

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

Setting - Tensors

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

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

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

Main problem:

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

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:

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

Setting - Tensors

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

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

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

Main problem:

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

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

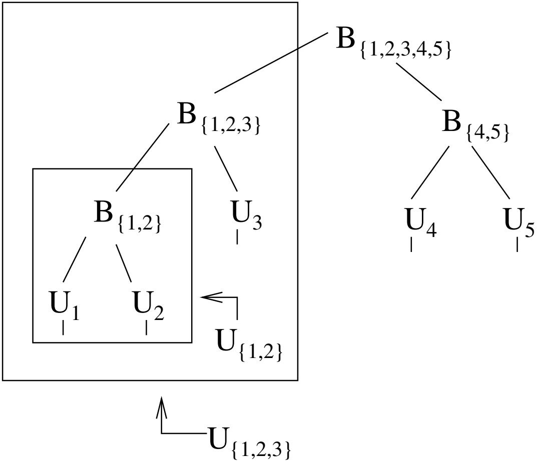
Approach: Some higher order tensors can be constructed (data-) sparsely from lower order quantities.

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

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

Tensor formats

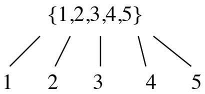
▷ Hierarchical Tucker format
(HT; Hackbusch/Kühn, Grasedyck, Kressner, Q: Tree-tensor networks)

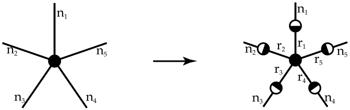


Tensor formats

- Hierarchical Tucker format
  (HT; Hackbusch/Kühn, Grasedyck, Kressner, Q: Tree-tensor networks)
- Tucker format (Q: MCTDH(F))

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

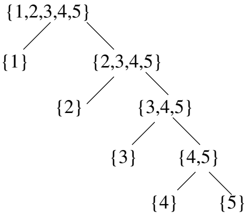


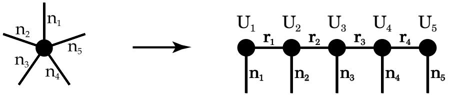


Tensor formats

- ▷ Hierarchical Tucker format
  (HT; Hackbusch/Kühn, Grasedyck, Kressner, Q: Tree-tensor networks)
- ▷ Tucker format (Q: MCTDH(F))
- ▷ Tensor Train (TT-)format
  (Oseledets/Tyrtyshnikov, $\simeq$ MPS-format of quantum physics)

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





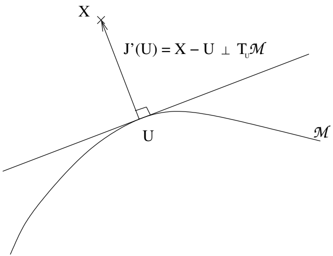
Approximation on low-rank manifold $\mathcal{M} \subseteq \mathcal{V}$

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

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

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

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



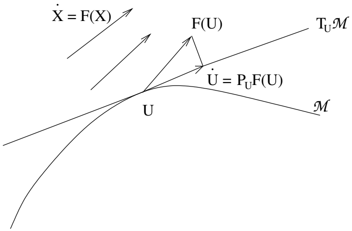
Approximation on low-rank manifold $\mathcal{M} \subseteq \mathcal{V}$

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

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

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

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



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

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

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 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

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

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

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

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 $\mathcal{M}_{\alpha_i}(\mathbf{v})$.

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/Vandereycken, 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/Vandereycken, in prep.)

Time-dependent equations:

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

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

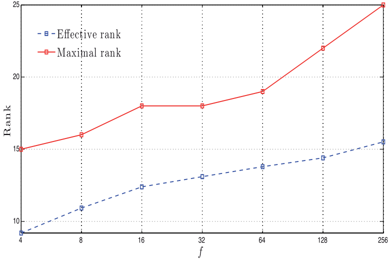
Practical example: EVP in Q-TT (Khoromkij/Oseledets, 2011)

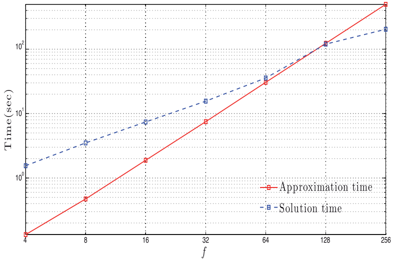
Problem from quantum molecular dynamics:

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

with potential energy surface given by Henon-Heiles potential

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





phys. dim. $f = 4, \ldots, 256$; one-direction grid size $n = 128 = 2^7$;
$\rightsquigarrow$ QTT-tensors $\in (\mathbb{R}^2)^{7f}$.

Summary and some open questions

- Manifold approach, in particular tangent space representation, allows treatment and analysis of high-dimensional optimization tasks and PDEs on low-rank tensor manifolds.
- TT and HT formats overcome weaknesses of older formats (CANDECOMP, Tucker).
- Quantum chemistry applications quite well-established and blooming, Fokker-Planck eq.: see Kohromskij et al. Still mostly open.
- Algorithms for ordering of indices?
- Theory: Uniqueness of minimizers, function space setting?

Projection onto the tangent space - by recursion

Let $\mathbf{v} \in \mathbf{V}$ be an HT tensor, $\alpha \subset D$, we decompose

$$\mathbf{v} = \sum_{\ell=1}^{r_\alpha} \mathbf{b}_\ell^{(\alpha)} \otimes \mathbf{w}_\ell^{(\alpha^C)} ,$$

with $\mathbf{w}_\ell^{(\alpha^C)} \in \mathbf{V}_{\alpha^C}$, $\alpha^C = D \backslash \alpha$, $\ell = 1, \ldots, r_\alpha$.

cutting leaves algorithm

An element $\delta\mathbf{v} \in \mathcal{T}_\mathbf{v}$ is of the form

$$\delta\mathbf{v} = \sum_{\ell=1}^{r_\alpha} (\delta\mathbf{b}_\ell^{(\alpha)} \otimes \mathbf{w}_\ell^{(\alpha^C)} + \mathbf{b}_\ell^{(\alpha)} \otimes \delta\mathbf{w}_\ell^{(\alpha^C)}) ,$$

where $\delta\mathbf{b}_\ell^{(\alpha)} \perp \mathbf{U}_\alpha$ and $\mathbf{w}_\ell^{(\alpha^C)}, \delta\mathbf{w}_\ell^{(\alpha^C)} \in \mathbf{V}_{\alpha^C}$, $\ell = 1, \ldots, r_\alpha$.

Appendix: Matricisation

Considering the matricization $\mathbf{v} \approx \mathbf{v}_{\mathbf{x}^\alpha}^{\mathbf{y}^{\alpha^C}}$, $\mathbf{A} \approx \mathbf{A}_{\mathbf{x}^\alpha}^{\mathbf{y}^{\alpha^C}}$, we rewrite the decompostion (**??**) into matrix factorization form

$$\mathbf{v} = \mathbf{B}^{(\alpha)}\mathbf{W}^{(\alpha^C),H} =: \mathbf{B}\mathbf{W}^H \ , \tag{1}$$

and, consequently, the tangent tensor into

$$\delta\mathbf{v} = \delta\mathbf{B}\mathbf{W}^H + \mathbf{B}\delta\mathbf{W}^H \ . \tag{2}$$

Then the solution of equations (**??**), (**??**) are given by

$$\delta\mathbf{B} = \left(\mathbf{I} - \mathbf{B}\mathbf{B}^H\right)\mathbf{A}\mathbf{W}\left(\mathbf{W}^H\mathbf{W}\right)^{-1} \tag{3}$$

$$\delta\mathbf{W} = \mathbf{B}^H\mathbf{A} =: \widetilde{\mathbf{A}} \ . \tag{4}$$

We abbreviate the Moore Penrose (left) inverse of $\mathbf{W}$ by

$$\mathbf{W}^+ := \mathbf{W}\left(\mathbf{W}^H\mathbf{W}\right)^{-1} , \tag{5}$$

and $\mathbf{P}_\perp . = \left(\mathbf{I} - \mathbf{B}\mathbf{B}^H\right)$ then $\delta\mathbf{B} = \mathbf{B}_\perp\mathbf{A}\mathbf{W}^+$.

Appendix: Estimates by recursion

Idea: use the 2D *Koch & Lubich* result recursively:

$$\gamma^2 \leq \min \lambda(\mathbf{W}^H\mathbf{W}) \text{ and } \Gamma^2 \geq \max \lambda(\mathbf{W}^H\mathbf{W}) .$$

$$\|\delta\mathbf{B}\|_2 \leq \frac{1}{\gamma}\|\mathbf{A}\|_2 .$$

Assumption:

$$\|\mathbf{u}-\mathbf{v}\| \leq \frac{1}{2}\Gamma\delta , \ \delta < \gamma . \tag{6}$$

$$\|\mathbf{B}-\widetilde{\mathbf{B}}\|_2 \leq \frac{\Gamma\delta}{\gamma} , \ \|\mathbf{W}-\widetilde{\mathbf{W}}\|_2 \leq \Gamma\delta , \ \|\mathbf{P}-\widetilde{\mathbf{P}}\|_2 \leq 2\frac{\Gamma\delta}{\gamma} , \ \|\mathbf{W}^+ - \widetilde{\mathbf{W}}^+\| \leq \frac{3\Gamma\delta}{\gamma} \tag{7}$$

$$\begin{aligned} \|\delta\mathbf{B}(\mathbf{W}-\widetilde{\mathbf{W}}^H)\| &\leq \frac{\Gamma\delta}{\gamma}\|\mathbf{A}\| \\ \|(\delta\mathbf{B}-\widetilde{\delta\mathbf{B}})\widetilde{\mathbf{W}}^H\| &\leq 6\frac{\Gamma\delta}{\gamma}\|\mathbf{A}\| \\ \|(\mathbf{B}-\widetilde{\mathbf{B}})\widetilde{\delta\mathbf{W}}^H\| &\leq \frac{\Gamma\delta}{\gamma}\|\mathbf{A}\| \end{aligned}$$

Projection estimates by recursion

This treatment is differs from Lubich, Rohwedder, S., Vandereycken! be aware of typos!

$$
\begin{aligned}
(P_{\mathbf{v}} - P_{\mathbf{u}})\widetilde{\mathbf{a}} = \delta\mathbf{v} - \delta\mathbf{u} &= && (8)\\
\delta\mathbf{B}\mathbf{W}^H + \mathbf{B}\delta\mathbf{W}^H - (\widetilde{\delta\mathbf{B}}\widetilde{\mathbf{W}}^H - \widetilde{\mathbf{B}}\widetilde{\delta\mathbf{W}}^H) &= \delta\mathbf{B}(\mathbf{W}^H - \widetilde{\mathbf{W}}^H) + (\delta\mathbf{B} - \widetilde{\delta\mathbf{B}})\widetilde{\mathbf{W}}^H && (9)\\
&+ (\mathbf{B} - \widetilde{\mathbf{B}})\widetilde{\delta\mathbf{W}}^H + \mathbf{B}(\delta\mathbf{W}^H - \widetilde{\delta\mathbf{W}}^H) && (10)
\end{aligned}
$$

The first three terms $\leq 8\frac{\Gamma\delta}{\gamma}\|\mathbf{A}\|$ by induction we show

$$
\|\mathbf{B}(\delta\mathbf{W}^H - \widetilde{\delta\mathbf{W}}^H)\| \leq \frac{\Gamma\delta}{\gamma}\|\widetilde{\mathbf{A}}\| \leq \frac{\Gamma\delta}{\gamma}\|\mathbf{A}\| , \qquad (11)
$$

Theorem

Let $\mathbf{u}, \mathbf{v}, \in \mathbf{V}$ *HT tensors with tangent spaces* $\mathcal{T}_{\mathbf{u}}, \mathcal{T}_{\mathbf{v}}$ *and* $\mathbf{a}, \in \mathbf{V}$*, with* $\|\mathbf{u} - \mathbf{v}\| = \frac{1}{2}\Gamma\delta$*. Assume* $\gamma \leq \min\{\sigma_i^{(\alpha)}\}$*,* $\Gamma \geq \max\{\sigma_i^{(\alpha)}\}$*, then*

$$
\|(P_{\mathbf{u}} - P_{\mathbf{v}})\mathbf{a}\| \leq 8(2d-3)\frac{\|\mathbf{v} - \mathbf{u}\|}{\gamma}\|\mathbf{a}\| . \qquad (12)
$$

$$
\|P_{\mathbf{v}})(\mathbf{v} - \mathbf{u})\| \leq 8(2d-3)\frac{1}{\gamma}\|\mathbf{v} - \mathbf{u}\|^2
$$

Thank you
for your attention.

References:

D. Conte and C. Lubich, *An error analysis of the multi-configuration time-dependent Hartree method of quantum dynamics*, M2AN, 44 (2010), pp. 759–780.

Chr. Lubich, Th. Rohwedder, R. Schneider, B. Vandereycken, *Dynamical approximation of tensors in the hierarchical Tucker and Tensor train formats (working title)*, in preparation, 2012.

B. Khoromskij, I. Oseledets, *DMRG+QTT approach to computation of the ground state of the molecular Schrödinger operator*, submitted to Num. Math., 2011.

S. Holtz, Th. Rohwedder, R. Schneider, *On manifolds of tensors with fixed TT rank*, Num. Math. DOI: 10.1007/s00211-011-0419-7, 2011.

O. Koch, C. Lubich, *Dynamical tensor approximation*, SIAM J. Matrix Anal. Appl. 31, p. 2360, 2010.

I. Oseledets, B.N. Khoromskij, and R. Schneider. *Efficient time-stepping scheme for dynamics on TT manifolds.* Preprint 24/2012 MPI MiS Leipzig, 2012.

A. Uschmajew and B. Vandereycken, *The geometry of algorithms using hierarchical tensors*, SPP 1324 preprint, 2012,

Remark: there are many more interesting papers of B. Khoromskij, to all previous lectures, only few are cited here, for a collection see e.g.
http://personal-homepages.mis.mpg.de/bokh/reference1.html
