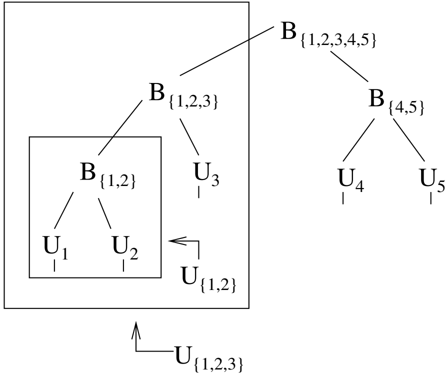
Tensor completion with hierarchical tensors

R. Schneider (TUB Matheon),
joint work with H. Rauhut and Z. Stojanac

Berlin December 2015

I.

Classical and novel tensor formats



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

Setting - Tensors of order d - hyper matrices

high-order tensors - multi-indexed arrays (hyper matrices)

$$\mathbf{x} = (x_1, \ldots, x_d) \mapsto U = U[x_1, \ldots, x_d] \in \mathcal{H}$$

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

Main problem: Let e.g. $\mathcal{V} = \mathbb{R}^{n^d}$

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

e.g.

$n = 10, d = 23, \ldots, 100, 200 \rightsquigarrow dim\mathcal{H} \sim 10^{23}, \ldots 10^{100}, 10^{200}$,
$6,1 \cdot 10^{23}$ Avogadro number, 10^{200} is a number much larger than the estimated number of all atoms in the universe!

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

As for matrices, incomplete SVD: reduces only to

$\sharp$DOFs $\geq Cn^{\frac{d}{2}} = C\sqrt{N}$ curse of dimensionality!

$$A[x_1, x_2] \approx \sum_{k=1}^{r} \left(u_k[x_1] \otimes v_k[x_2]\right) = \sum_{k=1}^{r} \tilde{u}[x_1, k] \cdot \tilde{v}[x_2, k]$$

Setting - Tensors of order d - hyper matrices

high-order tensors - multi-indexed arrays (hyper matrices)

$$\mathbf{x} = (x_1, \ldots, x_d) \mapsto U = U[x_1, \ldots, x_d] \in \mathcal{H}$$

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

Main problem: Let e.g. $\mathcal{V} = \mathbb{R}^{n^d}$

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

e.g.

$n = 10, d = 23, \ldots, 100, 200 \rightsquigarrow dim\mathcal{H} \sim 10^{23}, \ldots 10^{100}, 10^{200}$,
$6,1 \cdot 10^{23}$ Avogadro number, 10^{200} is a number much larger than the estimated number of all atoms in the universe!

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

As for matrices, incomplete SVD: reduces only to

$\sharp$DOFs $\geq Cn^{\frac{d}{2}} = C\sqrt{N}$ curse of dimensionality!

$$A[x_1, x_2] \approx \sum_{k=1}^{r} \left(u_k[x_1] \otimes v_k[x_2] \right) = \sum_{k=1}^{r} \tilde{u}[x_1, k] \cdot \tilde{v}[x_2, k]$$

Setting - Tensors of order d - hyper matrices

high-order tensors - multi-indexed arrays (hyper matrices)

$$\mathbf{x} = (x_1, \dots, x_d) \mapsto U = U[x_1, \dots, x_d] \in \mathcal{H}$$

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

Main problem: Let e.g. $\mathcal{V} = \mathbb{R}^{n^d}$

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

e.g.

$n = 10, d = 23, \dots, 100, 200 \rightsquigarrow dim\mathcal{H} \sim 10^{23}, \dots 10^{100}, 10^{200}$, $6,1 \cdot 10^{23}$ Avogadro number, 10^{200} is a number much larger than the estimated number of all atoms in the universe!

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

We do NOT use: **Canonical decomposition** for order-d-tensors:

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

Low Rank Matrix Approximation

$$U[x,y] = \sum_{k=1}^{r} U_1[x,k]U_2[y,k] \ , \ \sharp = rn_1 + rn_2 << n_1 \times n_2$$

Compressive sensing techniques - matrix completion by Candes, Recht & ....

Various ways to reshape $U[x_1, \ldots, x_d]$ into a matrix. Let $t \subset \{1, \ldots, d\}$, $\sharp t =: j$

$$\mathcal{M}_t(U) = (A_{\mathbf{x},\mathbf{y}}) \ , \ \mathbf{x} = (x_{t_1}, \ldots, x_{t_j})$$

example $\mathbf{x} := (x_1, \ldots, x_j), \mathbf{x} := (x_{j+1}, \ldots, x_d)$, $t = \{1, \ldots, j\}$

Basic Assumption Low dimensional subspace assumption

$$\mathcal{M}_t(U) \approx \mathcal{M}_t^{\epsilon}(U)$$

where

$$r_t := \text{rank}\mathcal{M}_t^{\epsilon}(U) = \mathcal{O}(d) = \mathcal{O}(f(\epsilon) \log n^d))$$

(e.g. $f(\epsilon) = \frac{1}{\epsilon^2}$ motivated by Johnson-Lindenstrauß Lemma.)

Low Rank Matrix Approximation

$$\sharp \mathcal{M}_t(U) = O(rn^{d-j} + rn^j)$$ curse of dimensions!!!

A single low rank matrix factorization cannot circumvent the curse of dimensions!

Can we benefit from various matricisation $\mathcal{M}_{t_1}(U), \mathcal{M}_{t_2}(U), \ldots$? Yes, we can!

Idea replicate low rank matrix factorization (HT)

$$U[x_1, \ldots, x_j, x_{j+1}, \ldots, x_d] = \sum_k U_L[x_1, \ldots, x_j, k] U_R[k, x_{j+1}, \ldots, x_d]$$

$$U_L[k, x_1, \ldots, \ldots, x_j] = \sum_{k'} U_{LL}[k', k, x_1, \ldots] U_{LR}[\ldots, x_j, k'] \quad \text{etc.}$$

Prototype example. TT tensor trains

$$U[x_1, x_2, \ldots, x_d] = \sum_{k_1=1}^{r_1} U_1[x_1, k_1] V_1[k_1, x_2, \ldots, x_d]$$

$$V_1[k_1, x_2, x_3, \ldots, x_d] = \sum_{k_2=1}^{r_2} U_2[k_1, x_2, k_2] V_2[k_2, x_3, \ldots, x_d] \quad \text{etc.}$$

$$\rightsquigarrow U[x_1, \ldots, x_d] = \sum_{k_1, \ldots, k_{d-1}} U_1[x_1, k_1] U_2[k_1, x_2, k_2] \cdots U_i[k_{i-1}, x_i, k_i] \cdots U_d[k_{d-1}, x_d]$$

Low Rank Matrix Approximation

$$\sharp\mathcal{M}_t(U) = O(rn^{d-j} + rn^j)$$ curse of dimensions!!!

A single low rank matrix factorization cannot circumvent the curse of dimensions!

Can we benefit from various matricisation $\mathcal{M}_{t_1}(U), \mathcal{M}_{t_2}(U), \ldots$? Yes, we can!

Idea replicate low rank matrix factorization (HT)

$$U[x_1, \ldots, x_j, x_{j+1}, \ldots, x_d] = \sum_k U_L[x_1, \ldots, x_j, k] U_R[k, x_{j+1}, \ldots, x_d]$$

$$U_L[k, x_1, \ldots, \ldots, x_j] = \sum_{k'} U_{LL}[k', k, x_1, \ldots] U_{LR}[\ldots, x_j, k'] \quad \text{etc.}$$

Prototype example. TT tensor trains

$$U[x_1, x_2, \ldots, x_d] = \sum_{k_1=1}^{r_1} U_1[x_1, k_1] V_1[k_1, x_2, \ldots, x_d]$$

$$V_1[k_1, x_2, x_3, \ldots, x_d] = \sum_{k_2=1}^{r_2} U_2[k_1, x_2, k_2] V_2[k_2, x_3, \ldots, x_d] \quad \text{etc.}$$

$$\rightsquigarrow U[x_1, \ldots, x_d] = \sum_{k_1, \ldots, k_{d-1}} U_1[x_1, k_1] U_2[k_1, x_2, k_2] \cdots U_i[k_{i-1}, x_i, k_i] \cdots U_d[k_{d-1}, x_d]$$

Low Rank Matrix Approximation

$$\sharp \mathcal{M}_t(U) = O(rn^{d-j} + rn^j)$$ curse of dimensions!!!

A single low rank matrix factorization cannot circumvent the curse of dimensions!

Can we benefit from various matricisation $\mathcal{M}_{t_1}(U), \mathcal{M}_{t_2}(U), \ldots$? Yes, we can!

Idea replicate low rank matrix factorization (HT)

$$U[x_1, \ldots, x_j, x_{j+1}, \ldots, x_d] = \sum_k U_L[x_1, \ldots, x_j, k] U_R[k, x_{j+1}, \ldots, x_d]$$

$$U_L[k, x_1, \ldots, \ldots, x_j] = \sum_{k'} U_{LL}[k', k, x_1, \ldots] U_{LR}[\ldots, x_j, k'] \text{ etc.}$$

Prototype example. TT tensor trains

$$U[x_1, x_2, \ldots, x_d] = \sum_{k_1=1}^{r_1} U_1[x_1, k_1] V_1[k_1, x_2, \ldots, x_d]$$

$$V_1[k_1, x_2, x_3, \ldots, x_d] = \sum_{k_2=1}^{r_2} U_2[k_1, x_2, k_2] V_2[k_2, x_3, \ldots, x_d] \text{ etc.}$$

$$\rightsquigarrow U[x_1, \ldots, x_d] = \sum_{k_1, \ldots, k_{d-1}} U_1[x_1, k_1] U_2[k_1, x_2, k_2] \cdots U_i[k_{i-1}, x_i, k_i] \cdots U_d[k_{d-1}, x_d]$$

Hierarchical subspace approximation, e.g. TT

Let $U \in \mathcal{H}$. For all $j = 1, \ldots, d-1$ we reshape U into matrices

$$U[x_1, \ldots, x_j, x_{j+1}, \ldots, x_d] =: \mathcal{M}_j(U)[\mathbf{x}, \mathbf{y}] \in V_{\mathbf{x}}^j \otimes (V_{\mathbf{y}}^j)'$$

where $V_{\mathbf{x}}^j := V_1 \otimes \cdots \otimes V_j$, $V_{\mathbf{y}}^j := V_{j+1} \otimes \cdots \otimes V_d$

1. Low dim. subspace assumption : $\forall j = 1, \ldots, d-1$, dim $V_{\mathbf{x}}^j =: r_j$ is moderate (sub-space approximation)

$$\mathbb{V}^j = \text{span}\{\phi_{k_j}[\mathbf{x}] = \phi_{k_j}[x_1, \ldots, x_j] : k_j = 1, \ldots, r_j\}$$

and

$$\mathcal{V}^j := \mathbb{V}^j \otimes V_{j+1} \otimes \cdots \otimes V_d$$

$$\Rightarrow \quad V_{\mathbf{x}}^{j+1} \subset \mathbb{V}^j \otimes V^{j+1} \quad \Rightarrow \text{ nestedness } \mathcal{V}^{j+1} \subset \mathcal{V}^j$$

we have a tensorial multi-resolution analysis,
$\rightsquigarrow$ a tensor MRA or T-MRA.

However we have modify the concept slightly. The unbalanced tree for TT is only an example for general dimension trees $\mathbb{T}$

Hierarchical subspace approximation (e.g. TT) and tensor MRA

Nestedness:

$$\mathcal{V}^{j+1} \subset \mathcal{V}^{j} \ , \ \ \mathcal{V}^{j} = \mathcal{V}^{j+1} + \mathcal{W}^{j+1} \ \Rightarrow \ \mathbb{V}^{j+1} \subset \mathbb{V}^{j} \otimes V_{j+1}$$

so far $\mathcal{W}^{j+1}$ has been ignored!!!

recursive SVD (HSVD) $\rightsquigarrow$ 2-scale refinement rel.: $1 \leq k_j \leq r_j$

$$\phi_{k_j}[x_1, \ldots, x_{j-1}, x_j] := \sum_{k_{j-1}=1}^{r_{j-1}} U_j[k_{j-1}, \alpha_j, k_j] \phi_{k_{j-1}}[x_1, \ldots, x_{j-1}] \otimes \mathbf{e}_{\alpha_j}[x_j]$$

for simplicity let us take $\mathbf{e}_{\alpha_j}[x_j] = \delta_{\alpha_j, x_j}$. We need only

$$U_j[k_{j-1}, x_j, k_j] \ , \ j = 1, \ldots, d$$

to define full tensor $U \Rightarrow$ complexity $\mathcal{O}(nr^2d)$

$$U[x_1, \ldots, x_d] = \sum_{k_1, \ldots, k_{d-1}} U_1[x_1, k_1] U_2[k_1, x_2, k_2] \cdots U_i[k_{i-1}, x_i, k_i] \cdots U_d[k_{d-1}, x_d]$$

This is an adaptive MRA, or non stationary sub-division like algorithm where

$$\mathcal{V}^d = \operatorname{span}\{\phi^d\}, \phi^d[x_1, \ldots, x_d] = U[x_1, \ldots, x_d] \ , \ \dim \mathcal{V}^d = 1!$$

General Hierarchical Tensor (HT) format

- ▷ General hierarchical tensor setting
- ▷ Subspace approach (Hackbusch/Kühn, 2009)

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

General Hierarchical Tensor (HT) format

- ▷ Given dimension tree

$\rightsquigarrow$ a manifold!

- ▷ Subspace approach (Hackbusch/Kühn, 2009)

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

General Hierarchical Tensor (HT) format

- Given dimension tree

$\rightsquigarrow$ a manifold!

- Subspace approach (Hackbusch/Kühn, 2009)

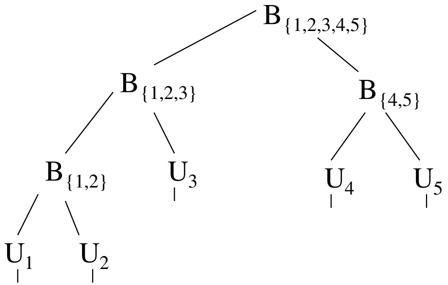
(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n \times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t_1} \times k_{t_2}}$)

General Hierarchical Tensor (HT) format

- ▷ Given dimension tree

$\rightsquigarrow$ a manifold!

- ▷ Subspace approach (Hackbusch/Kühn, 2009)



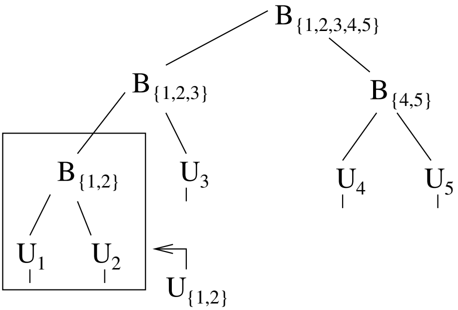
(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n \times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t_1} \times k_{t_2}}$)

General Hierarchical Tensor (HT) format

- ▷ Given dimension tree

$\rightsquigarrow$ a manifold!

- ▷ Subspace approach (Hackbusch/Kühn, 2009)



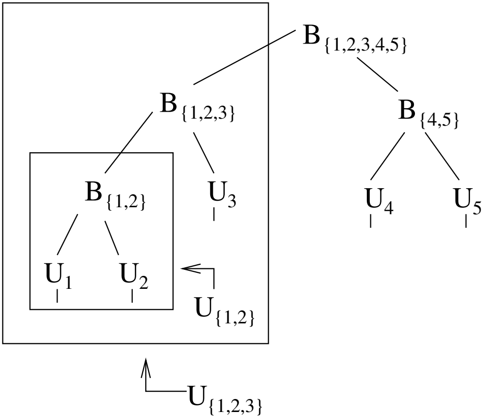
(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n \times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t_1} \times k_{t_2}}$)

General Hierarchical Tensor (HT) format

- ▷ Given dimension tree

$\rightsquigarrow$ a manifold!

- ▷ Subspace approach (Hackbusch/Kühn, 2009)



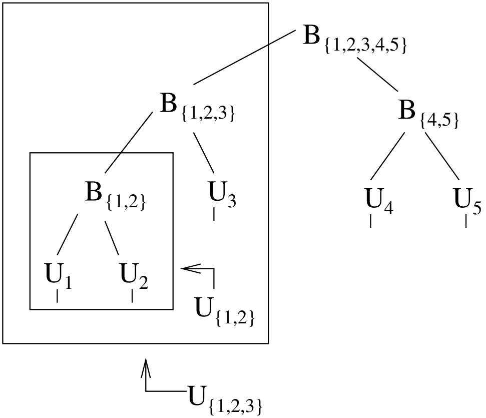
(Example: $d = 5, \mathbf{U}_i \in \mathbb{R}^{n \times k_i}, \mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t_1} \times k_{t_2}}$)

General Hierarchical Tensor (HT) format

- ▷ Given dimension tree

  $\rightsquigarrow$ a manifold!

- ▷ Subspace approach (Hackbusch/Kühn, 2009)



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

Application of HT concepts

- Hidden Markov models ...
- Quantum physics - 1 D spin systems - density matrix renormalization group DMRG S. White (1992) MPS with open boundary conditions best know tool - standard
- 2D or 3 D spin systems or Hubbard model - tensor networks (Vidal, Verstraete, Cirac, Schollwöck, Jens Eisert, Kitaev ...) standard tool $N = 2^d$, $d \approx 100 - 200$, $r \geq 10000$.
- Quantum Chemistry - Q-DMRG (G. Chan (Princeton), Legeza, Reiher (ETHZ), ..., our group) only for strong correlation effects, $N = 2^d$, $d \approx 100$, $r \sim 1000 - 10000$.
- Molecular dynamics -Langevin dynamics (new) (Noe & Nske & & Vitali our group . 2014) $N = n^d$, e.g. $n = 2$, $d = 254$, $r \leq 8!$.
- Uncertainty quantification (UQ): Oseledets & Khoromskij, Grasedyck, Espig & Matthies & Hackbusch, our group) $N \sim n^d$, $n \leq 10$, $d \leq 150$.
- Signal analysis: daSilva & Herrmann (great paper!), Kressner et al.
- machine learning: Cickochi, Oseledets,
- combination with variable transformation (see Vybiral& Fournasier): Oseledets

Hierarchical tensor or tensor networks is tool which has been successfully applied to high dimensional ($d >> 1$) problems in linear spaces of dimensions $N \sim n^d \sim 10^{80}$ number large than the number of all atoms in the earth $\leq 10^{62}$ or the sun $\leq 10^{68}$).

$$n^d \quad \rightsquigarrow \quad ndr^2 \;\text{ or }\; ndr + r^3 = \mathcal{O}(d) \;\text{ (so far)}$$

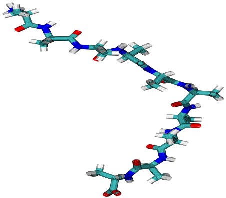
Transfer operator for MD simulation

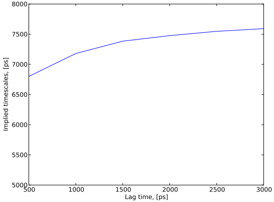
ongoing joint work with F. Nüsken & F. Noe (FU Berlin, ZIB), F. Vitali

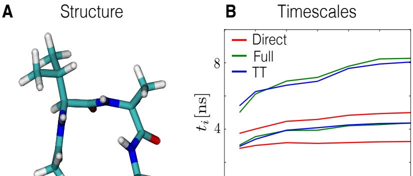
We look for the first $N = 3(2)$ eigenfunctions of the transfer operator

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

Dimension $d = 18$ largest example 58-residue protein BPTI produced on the Anton supercomputer provided by D.E. Shaw research 4d=258







Conclusions

Most matrix techniques can be extended to hierarchical tensors

1. SVD $\rightsquigarrow$ HSVD (but only quasi-optimal approximation)
2. hard and soft thresholding iteration
3. Riemanian optimization Riemanian gardient iteration,

   Tangent space has almost the same structure and can be straightforwardly deduced from the matrix case
4. matrix completion $\rightsquigarrow$ tensor completion ?

Conrtributions to HT

- HT - Hackbusch & Kühn (2009), TT - Oseledets & Tyrtyshnikov (2009)
- MPS- Affleck et al. AKLT (87), Fannes et al. (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, Rohwedder, S. Uschmajew, Vandereycken, daSilva, Herrman Kressner, Steinlechner, ...
- Oseledets, Khoromskij, Savostyanov, Dolgov, Kazeev, ...
- Grasedyck, Ballani, Bachmayr, Dahmen, ...
- Physics: Cirac, Verstraete, Schollwöck, G. Chan, Eisert, ...

Low Rank Tensor Recovery - Tensor Completion

Given p measurements

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

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

Tensor completion: given values at randomly chosen points $\mathbf{k}_i$,

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

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

E.g. TT-format oracle dimension

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

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

Remark: (HT -) TT representation of

$$\mathcal{A}^T \mathbf{y} \ \ = \ \ \sum_{i=1}^{p} y[i] \mathbf{e}_{x_{1,i}} \otimes \cdots \otimes \mathbf{e}_{x_{d,i}}$$

$U_j[k_{j-1}, x_j, k_j] = \tilde{y}[i,j]\delta_{k_{j-1},i}\delta_{k_j,i}\delta_{x_{j,i},x_j}\ \ ,\ \ U_j \in \mathbb{R}^{p \times n_j \times p}$ but sparse

Hard Thresholding

Projected Gradient Algorithms: Minimize residual

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

w.r.t. low rank constraints

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

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

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

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

1. Hard thresholding, $\sigma_s := 0$, $s > r$, $\sigma_s \leftarrow \sigma_s$, $s \leq r$ compressive sensing: Blumensath et al. , matrix recovery : Tanner et al., Jain et al.

Hard Thresholding - Riemannian gradient iteration

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

Projected gradient is the Riemannian gradient w.r.t. to the embedded metric

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

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

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

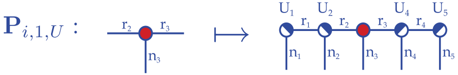
e.g. R is an approximate exponential map

in matrix completion: e.g. MLAFIT and several others, e.g Kershavan, Montanari, & O, Vandereycken, Saad et al., Sepulchre et al., Kressner et al.., W. Yin et al. etc.

Block coordinate search for TT (HT) tensors - ALS

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, $\nu \in \{1, \ldots, d\} \setminus \{j\}$, 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 $\simeq \mathbb{R}^{r_{i-1}} \otimes V_i \otimes \mathbb{R}^{r_i} \subset \mathcal{H}$
3) and orthogonalize left to define a basis for the next step
4) Repeat with $\mathbf{U}^{j+1}$

S. Holtz & Rohwedder & S. (2010), Oseledets et al. (2013),Cickochi et al. (2014)

Single site DMRG /density matrix renormalization alg.

Variant: ADS performs only a gradient step in [4]

(alternating directional search - Grasedyck & Krämer 2016, Espig et al. 2014)

This reduces the computational complexity of ALS

$$\mathcal{O}(pndr^4) \rightsquigarrow \mathcal{O}(pndr^2) \ , \ (p >> n, r, d)$$

Analysis: S. (2016) - (preconditioned) Riemannian gradient it.

Linear measurements and TRIP - tensor RIP

Here $\|U\|_H$ is the norm in $\mathcal{H}$

Definition

Restricted isometry property (RIP) of order $\underline{s}$: there exists a restricted isometry constant (RIC) $0 < \delta_{\underline{s}} < 1$ s.t. for all $U \in \mathcal{M}_{\leq \underline{s}}$ there holds

$$(1 - \delta_{\underline{s}})\|U\|_H^2 \leq \|\mathcal{A}U\|_2^2 \leq (1 + \delta_{\underline{s}})\|U\|_H^2 \,. \tag{1}$$

Bi- Lipschitz estimate : with $0 < \alpha = \alpha_{\leq \underline{s}} \leq \beta = \beta_{\leq \underline{s}}$

$$\alpha\|U\|_H \leq \|\mathcal{A}U\| \leq \beta\|U\|_H \ \ \forall U \in \mathcal{M}_{\leq \underline{s}} \tag{2}$$

TRIP - Tensor RIP

Theorem (Stojanac & Rauhut)

Given $0 < \delta < 1$. *For (sub-)Gaussian measurements* $\mathcal{A}$ *the RIP holds with isometry constant* $0 < \delta_r \leq \delta < 1$ *with probability exceeding* $(1 - e^{-cp})$ *provided that*

- *Tucker format:*

$$p > C\delta^{-2}(dnr + r^d)\log d \sim D(\delta)m \,,$$

- *TT format*

$$p > C\delta^{-2}ndr^2\log(dr) \sim D(\delta)m$$

- *conjecture: HT (work in progress)*

$$p > C\delta^{-2}(ndr + dr^3)\log(dr) \sim D(\delta)m$$

for constants $D(\delta), c > 0$

Iterative Hard Thresholding - Local Convergence

Theorem (Conditional global convergence of IHT)

Let $V^{n+1} := U^n + \mathcal{A}^*(\mathbf{y} - \mathcal{A}U^n)$*, and* $U^{n+1} = \mathbf{H_r} V^{n+1}$ *assume that* $\mathcal{A}$ *satisfies the RIP of order* $3\mathbf{r}$*, If*

$$\|\mathbf{H_r} V^{n+1} - V^{n+1}\|^2 \leq \|U - V^{n+1}\|^2 \quad \textit{assumption A}$$

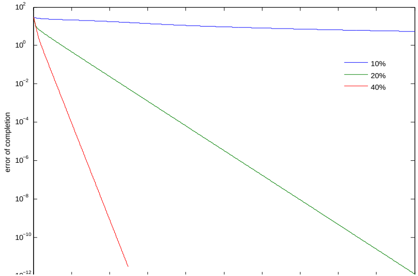
then, there exist $0 < \rho < 1$ *s.t the series* $U^n \in \mathcal{M}_{\leq \mathbf{r}}$ *converges linearly to a unique solution* $U \in \mathcal{M}_{\leq \mathbf{r}}$ *with rate* ρ

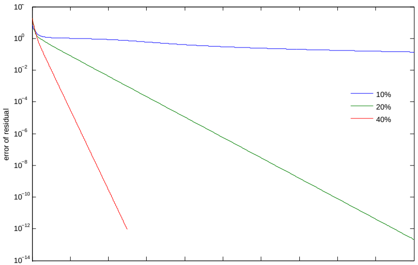
$$\|U^{n+1} - U\| \leq \rho \|U^n - U\|$$

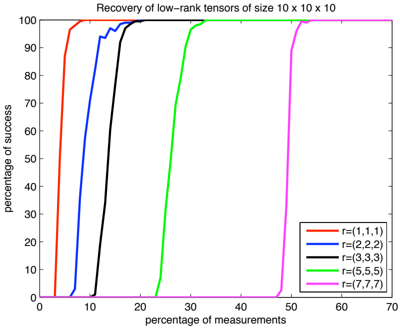
Can we benefit from recent progress in the analysis of matrix completion by ALS: Hardt (2014), Jain, Netrapalli, Sanghavi & Dhillon ...

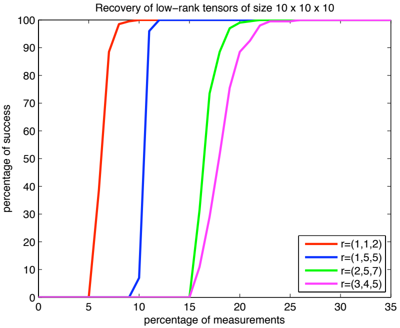
First numerical examples

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



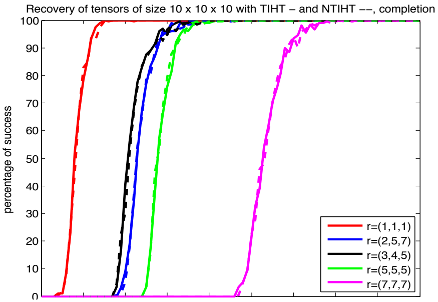


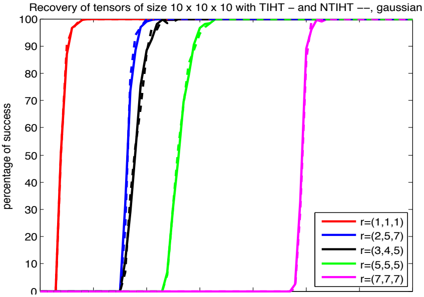


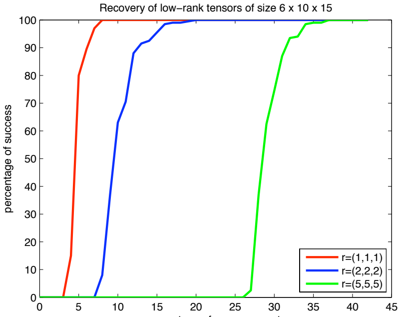


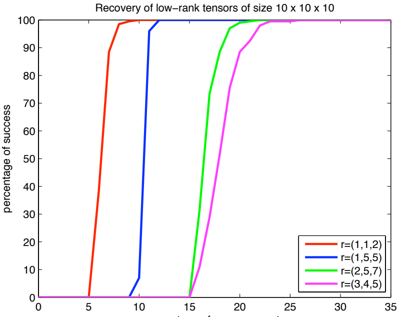
Numerical examples

Stojanac Gaussian measurements



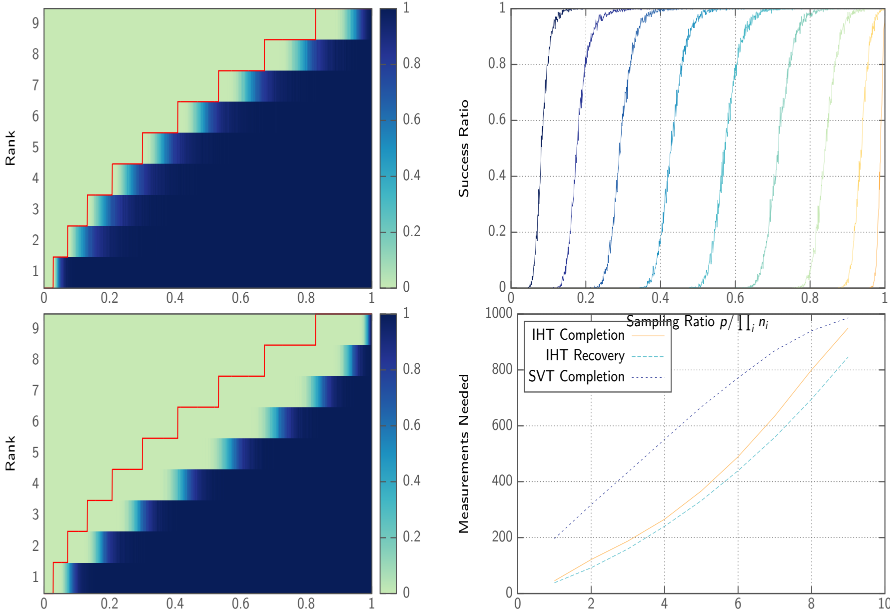






Numerical examples

Sebastian Wolf Master thesis - tensor completion (without and with noise)



Thank you
for your attention.