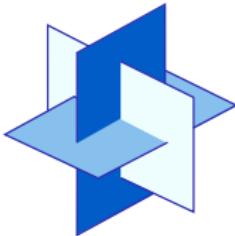


Tensor completion with hierarchical tensors

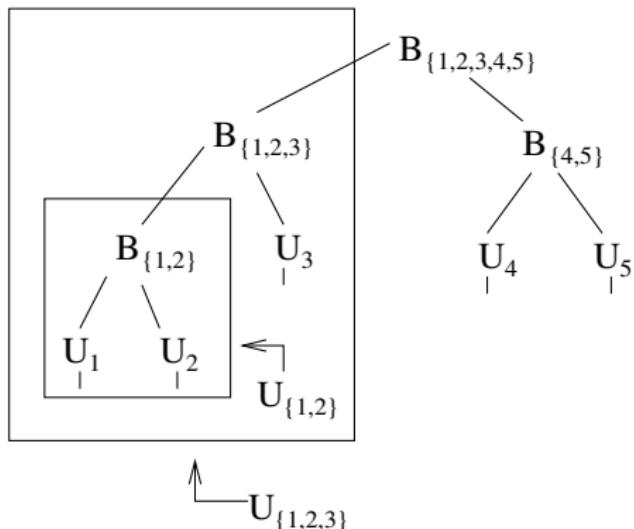
R. Schneider (TUB Matheon),
joint work with H. Rauhut and Z. Stojanac

Berlin December 2013



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)

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

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

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

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

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

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

As for matrices, incomplete SVD: reduces only to

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

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

Setting - Tensors of order d - hyper matrices

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

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

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

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

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

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

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

As for matrices, incomplete SVD: reduces only to

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

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

Setting - Tensors of order d - hyper matrices

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

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

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

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

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

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

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

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

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

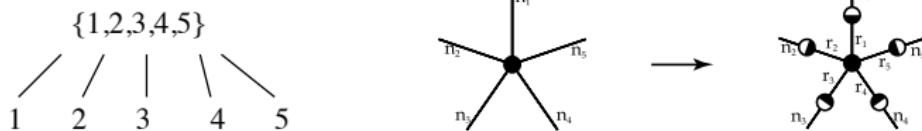
Subspace approximation instead of canonical dec.

Subspace approximation instead of canonical dec.

- ▷ Tucker format (Q : MCTDH(F)) - robust Segre -variety (closed)
But complexity $\mathcal{O}(r^d + ndr)$
Is there a robust tensor format, but polynomial in d ?

Univariate bases $x_i \mapsto (U_i[k_i, x_i])_{k_i=1}^{r_i}$ (\rightarrow Graßmann man.)

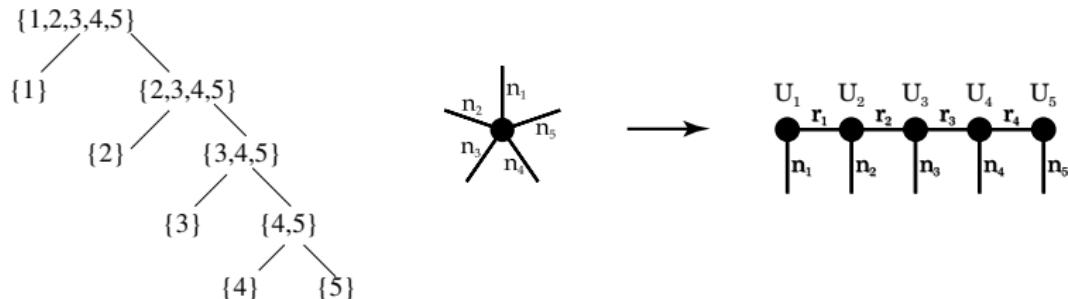
$$U[x_1, \dots, x_d] = \sum_{k_1=1}^{r_1} \dots \sum_{k_d=1}^{r_d} B[k_1, \dots, k_d] \bigotimes_{i=1}^d \mathbf{u}_i[k_i, x_i]$$



Subspace approximation instead of canonical dec.

- ▷ Tucker format (Q: MCTDH(F)) - robust Segre -variety (closed)
But complexity $\mathcal{O}(r^d + ndr)$
Is there a robust tensor format, but polynomial in d ?
- ▷ Hierarchical Tucker format
(HT; Hackbusch/Kühn, Grasedyck, : Tree-tensor networks)
- ▷ Tensor Train (TT)-format \simeq Matrix product states (MPS)
(Oseledets/Tyrtyshnikov (09))

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



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

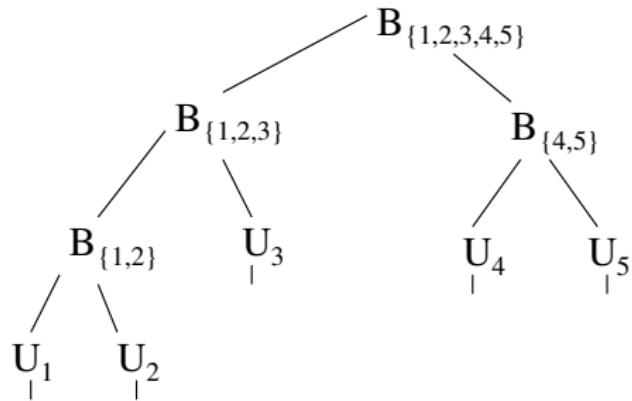
Hierarchical tensor (HT) format

- ▷ 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

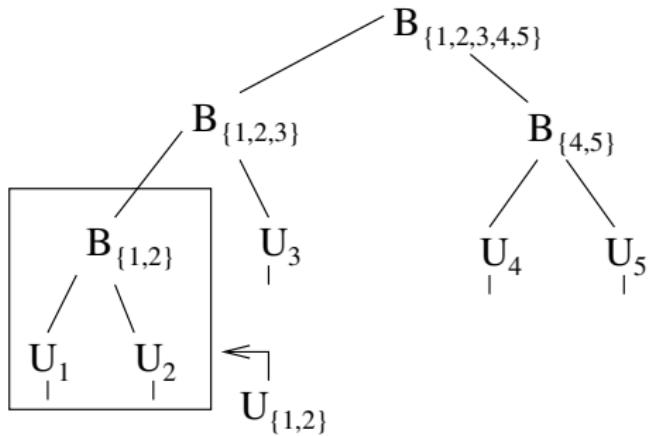
- ▷ Canonical decomposition **not closed, no embedded manifold!**
- ▷ Subspace approach (Hackbusch/Kühn, 2009)



(Example: $d = 5$, $\mathbf{U}_i \in \mathbb{R}^{n \times k_i}$, $\mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t1} \times k_{t2}}$)

Hierarchical tensor (HT) format

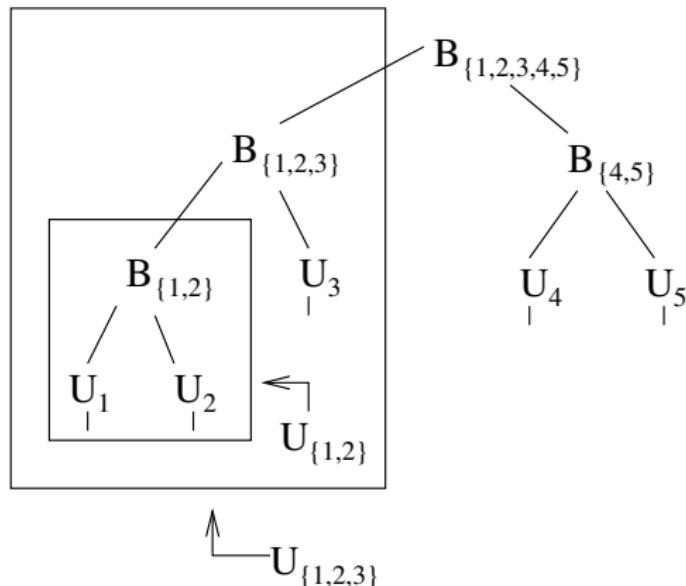
- ▷ Canonical decomposition **not closed, no embedded manifold!**
- ▷ Subspace approach (Hackbusch/Kühn, 2009)



(Example: $d = 5$, $\mathbf{U}_i \in \mathbb{R}^{n \times k_i}$, $\mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t1} \times k_{t2}}$)

Hierarchical tensor (HT) format

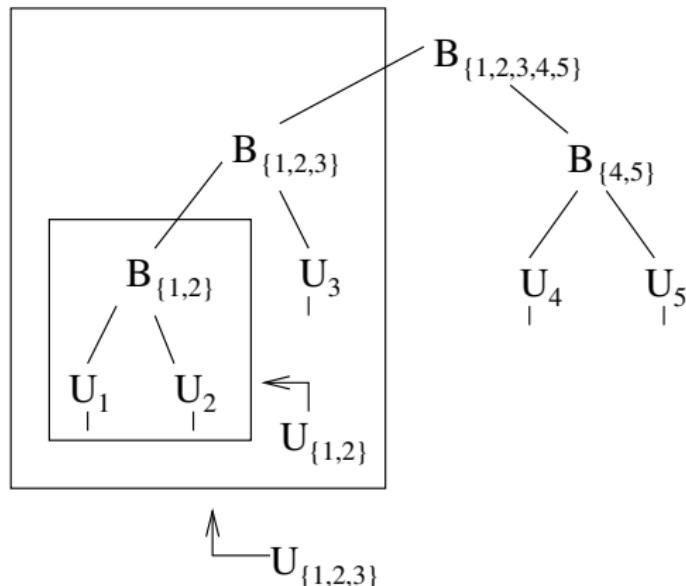
- ▷ Canonical decomposition **not closed, no embedded manifold!**
- ▷ Subspace approach (Hackbusch/Kühn, 2009)



(Example: $d = 5$, $\mathbf{U}_i \in \mathbb{R}^{n \times k_i}$, $\mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t1} \times k_{t2}}$)

Hierarchical tensor (HT) format

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



(Example: $d = 5$, $\mathbf{U}_i \in \mathbb{R}^{n \times k_i}$, $\mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t1} \times k_{t2}}$)

TT - Tensors - Matrix product representation

Noteable special case of HT:

TT format (Oseledets & Tyrtiyshnikov, 2009)

\simeq matrix product states (MPS) in quantum physics (Heisenberg model - spin systems), Affleck, Kennedy, Lieb & Tagasaki (87.), Fannes, Nachtergale & Werner (91), Römmer & Ostlund (94), Vidal (03), Schöllwock, Cirac, Verstraete, Wolf, Eisert ...)

TT tensor U can be written as matrix product form

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

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

with matrices or component tensors

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

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

Fundamental properties of HT (particularly TT)

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

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

into row or column index of $\mathbf{A}_t = \mathbf{A}_t(U) = (\mathbf{A}_{\mathcal{I}_t, \mathcal{I}_D \setminus \mathcal{I}_t}) \Rightarrow$
matricisation or unfolding of

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

e.g. TT format $r_i = \text{rank } A_{x_1, \dots, x_i}^{x_{i+1}, \dots, x_d}$. There exist a well defined
rank tuple $\mathbf{r} := (r_t)_{t \in \mathbb{T}}$, e.g. $\mathbf{r} = (r_1, \dots, r_{d-1})$ for TT

$\mathcal{M}_{\mathbf{r}} = \{U \in \mathcal{H} : r_t = \text{rank } \mathbf{A}_t, t \in \mathbb{T}\}$ is analytic manifold

$$\mathcal{M}_{\underline{r}} \simeq (\times_{i=1}^d X_i) / \mathcal{G}_{\underline{r}}$$

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

Hackbusch & Falco

$\mathcal{M}_{\leq \mathbf{r}}$ is a an algebraic variety

Fundamental properties of HT (particularly TT)

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

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

into row or column index of $\mathbf{A}_t = \mathbf{A}_t(U) = (\mathbf{A}_{\mathcal{I}_t, \mathcal{I}_D \setminus \mathcal{I}_t}) \Rightarrow$
matricisation or unfolding of

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

e.g. TT format $r_i = \text{rank } A_{x_1, \dots, x_i}^{x_{i+1}, \dots, x_d}$. There exist a well defined rank tuple $\mathbf{r} := (r_t)_{t \in \mathbb{T}}$, e.g. $\mathbf{r} = (r_1, \dots, r_{d-1})$ for TT
 $\mathcal{M}_{\underline{r}} = \{U \in \mathcal{H} : r_t = \text{rank } \mathbf{A}_t, t \in \mathbb{T}\}$ is analytic manifold

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

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

Hackbusch & Falco

$\mathcal{M}_{\leq \underline{r}}$ is a an algebraic variety

Fundamental properties of HT (particularly TT)

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

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

into row or column index of $\mathbf{A}_t = \mathbf{A}_t(U) = (\mathbf{A}_{\mathcal{I}_t, \mathcal{I}_D \setminus \mathcal{I}_t}) \Rightarrow$
matricisation or unfolding of

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

e.g. TT format $r_i = \text{rank } A_{x_1, \dots, x_i}^{x_{i+1}, \dots, x_d}$. There exist a well defined rank tuple $\mathbf{r} := (r_t)_{t \in \mathbb{T}}$, e.g. $\mathbf{r} = (r_1, \dots, r_{d-1})$ for TT
 $\mathcal{M}_{\underline{r}} = \{U \in \mathcal{H} : r_t = \text{rank } \mathbf{A}_t, t \in \mathbb{T}\}$ is analytic manifold

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

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

Hackbusch & Falco

$\mathcal{M}_{\leq \underline{r}}$ is a an algebraic variety

HOSVD and Summary

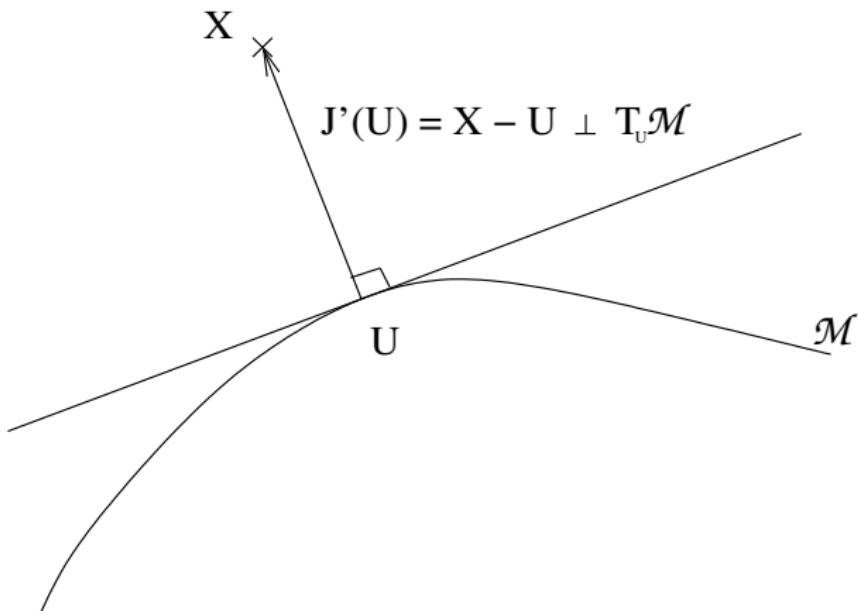
HOSVD- thresholding of singular values in of A_t for all $t \in \mathbb{T}$.

$$\|U - U_{best}\| \leq \|U - U_{HOSVD}\| \leq 2\sqrt{d}\|U - U_{best}\|$$

- ▶ includes the case of low rank matrices $d = 2$
 - ▶ rank tuple $(r_t)_{t \in \mathbb{T}}$
 - ▶ complexity HT $\mathcal{O}(r^3 + nrd)$ (or TT $\mathcal{O}(ndr^2)$),
 $r := \max\{r_t : t \in \mathbb{T}\}$
 - ▶ HOSVD provides exact reconstruction and a quasi-optimal approximation
 - ▶ \mathcal{M}_r is a manifold, and $\mathcal{M}_{\leq r}$ is a variety, (but not convex!)
 - ▶ But, for $d > 2$, best approximation $\operatorname{argmin}_{V \in \mathcal{M}_{\leq}} \|U - V\|$ is NP hard, in general (Hillar & Lim 2012)
 - ▶ $\operatorname{conv hull} \{U \in \mathcal{H} : \|U\|_0 \leq 1\} \neq \{U : \sum_{t \in \mathbb{T}} \|\mathbf{A}_t\|_{*,1}\}$, we do not know a good convex yet!
- ⇒ HT is a good candidate to extend matrix completion to higher order tensors, circumventing principle difficulties of the canonical format

II.

Tensor Completions - Low Rank Tensor Recovery



Low Rank Tensor Recovery - Tensor Completion

Given p measurements

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

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

Tensor completion: given values

$$U(\mathbf{k}_i) \quad , \quad i = 1, \dots, p \quad << N = n^d .$$

at randomly chosen points \mathbf{k}_i ,

Can one reconstruct $\mathbf{U} \in \mathcal{M}_{\underline{\mathbf{r}}}$?

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

$\mathbf{U} \in \mathcal{M}_{\leq \mathbf{r}}$

E.g. as a prototype example TT-format in matrix product representation, oracle dimension

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

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

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

$$\begin{aligned} Y_{n+1} &:= \textcolor{blue}{U}_n - \alpha_n (\mathcal{A}^T(\mathcal{A}U_n - \mathbf{y})) \text{ gradient step} \\ \textcolor{blue}{U}_{n+1} &:= \mathcal{R}_n(Y_{n+1}). \end{aligned}$$

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

$$\mathcal{R}_n : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathcal{M}_r$$

e.g HOSVD $\sigma_s := \sigma_{s_t}$ singular values of $\mathbf{A}_t = \mathbf{A}_t(Y_{n+1})$, $t \in \mathbb{T}$,

1. Hard thresholding, $\sigma_s := 0$, $s > r$, $\sigma_s \leftarrow \sigma_s$, $s \leq r$ CS:
Blumensath et al. , matrix case: Tanner et al.
2. Riemannian techniques including ALS: e.g. Kressner et al. (2013), da Silva & Herrmann (2013)
3. Soft thresholding, $\sigma_s \leftarrow \max\{\sigma_s - \varepsilon, 0\}$

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

$$\begin{aligned} Y_{n+1} &:= \textcolor{blue}{U}_n - \alpha_n (\mathcal{A}^T(\mathcal{A}U_n - \mathbf{y})) \text{ gradient step} \\ \textcolor{blue}{U}_{n+1} &:= \mathcal{R}_n(Y_{n+1}). \end{aligned}$$

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

$$\mathcal{R}_n : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathcal{M}_r$$

e.g HOSVD $\sigma_s := \sigma_{s_t}$ singular values of $\mathbf{A}_t = \mathbf{A}_t(Y_{n+1})$, $t \in \mathbb{T}$,

1. Hard thresholding, $\sigma_s := 0$, $s > r$, $\sigma_s \leftarrow \sigma_s$, $s \leq r$ CS:
Blumensath et al. , matrix case: Tanner et al.
2. Riemannian techniques including ALS: e.g. Kressner et al. (2013), da Silva & Herrmann (2013)
3. Soft thresholding, $\sigma_s \leftarrow \max\{\sigma_s - \varepsilon, 0\}$

Hard Thresholding - Riemannian optimization

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

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

$$\begin{aligned} Y_{n+1} &:= \color{blue}{U_n} - \color{red}{P_{\mathcal{T}_U}\alpha_n(\mathcal{A}^T(\mathcal{A}U_n - \mathbf{y}))} \text{ projected gradient step} \\ &= \color{blue}{U_n + \xi_n} , \color{blue}{\mathcal{M}_{\mathbf{r}}} + \color{red}{\mathcal{T}_U} \\ U_{n+1} &:= \color{blue}{\mathcal{R}_n(Y_n)} := \color{blue}{R(U_n, \xi_n)} . \end{aligned}$$

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

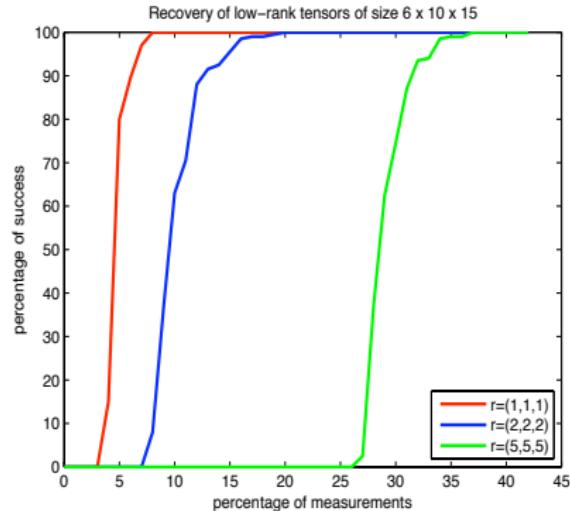
$$R(U, \xi) = U + \xi + \mathcal{O}(\|\xi\|^2)$$

e.g. R is an approximate exponential map

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

III.

Iterative Hard Thresholding (local) convergence



HOSVD - high order SVD

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

Matricisation or unfolding

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

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

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

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

with matrices or component functions

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

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

HOSVD - high order SVD

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

Matricisation or unfolding

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

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

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

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

with matrices or component functions

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

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

HOSVD - high order SVD

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

Matricisation or unfolding

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

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

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

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

with matrices or component functions

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

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

Iterative Hard Thresholding

$$U_{n+1} := H_r(U_n + \alpha_n \mathcal{A}^*(\mathbf{y} - \mathcal{A}U_n)) \in \mathcal{M}_r .$$

where $H_r : \mathcal{H} \rightarrow \mathcal{M}_{\leq r}$ an operator resulting a quasi-best approximation, e.g. H_r is provided by the **HOSVD** for hard thresholding,

$$\|H_r V - V\| \leq \inf_{W \in \mathcal{M}_{\leq r}} \|V - W\| .$$

Here we focus on locally almost best

$$\|H_r V - V\| \leq \inf_{V_r \in \mathcal{M}_{\leq r}} \|V - V_r\| + \mathcal{O}(\|H_r V - V\|^2)$$

is valid for $H_r(V) := R(U, P_{\mathcal{T}_U}(V - U)$ for Riemannian optimization and HOSVD.

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

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

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

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

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:

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

- ▶ 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

Let $V_n := U_n + \mathcal{A}^*(\mathbf{y} - \mathcal{A}U_n)$, assume that \mathcal{A} satisfies the RIP of order $3r$, $2r$ with RIC's and

$$\|U - U_m\| \leq \delta , \quad \delta \sim \text{dist}(U, \partial \mathcal{M}_r)$$

both sufficiently small,

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

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

Remark: Suppose $\|U\| = 1$ then

$$\text{dist}(U, \partial \mathcal{M}_r) \leq \min_{t \in \mathbb{T}, 0 < k \leq r_t} \sigma_{t,k}$$

is smallest (non-zero) singular value of $\mathbf{A}_t(U)$!

Iterative Hard Thresholding - Convergence Proof

For $X \in \mathcal{M}_{\leq r}$, define the orthogonal projections $P_X : \mathcal{H} \rightarrow \mathcal{M}_{\leq r}$ s.t $P_X X = X$, (e.g $P_X = E_d^T E_d$),
 $\Omega^n := P_{U^n}(\mathcal{H}) \oplus P_{U^{n+1}}(\mathcal{H}) \subset \mathcal{M}_{\leq 3r}$ and $P^n : \mathcal{H} \rightarrow \Omega^n$.

We estimate

$$\begin{aligned}\|U^{n+1} - U\| &= \|P^n(U^{n+1} - U)\| \\ &\leq \|P^n(U^{n+1} - V^n)\| + \|P^n(U - V^n)\| \\ &\leq (1+1)\|P^n(U - V^n)\| + \mathcal{O}(\|U - U^n\|^2) \\ &= 2\|P^n((U - U^n) - \mathcal{A}^*(\mathbf{y} - \mathcal{A}U^n))\| + \mathcal{O}(\|U - U^n\|^2) \\ &= 2(\|P^n((U - U^n) - P^n \mathcal{A}^* \mathcal{A}(U - U^n))\|) + \mathcal{O}(\|U - U^n\|^2) \\ &= 2\left(\|P^n((U - U^n - (\mathcal{A}P^n)^* (\mathcal{A}P^n(U - U^n))))\|\right) + \mathcal{O}(\|U - U^n\|^2) \\ &\leq \rho \|P^n(U - U^n)\| \\ &= \rho \|(U - U^n)\|.\end{aligned}$$

For δ , $\|U - U_0\|$ sufficiently small

Since $\Omega^n = P^n(\mathcal{H}) \subset \mathcal{M}_{\leq 3r}$, for suff. small RICs $0 < \delta_{\leq 3r}$ there ex. $0 \leq \rho < 1$, and the series converges.

Iterative Hard Thresholding - Convergence Proof

For $X \in \mathcal{M}_{\leq r}$, define the orthogonal projections $P_X : \mathcal{H} \rightarrow \mathcal{M}_{\leq r}$ s.t $P_X X = X$, (e.g $P_X = E_d^T E_d$),
 $\Omega^n := P_{U^n}(\mathcal{H}) \oplus P_{U^{n+1}}(\mathcal{H}) \subset \mathcal{M}_{\leq 3r}$ and $\textcolor{red}{P^n} : \mathcal{H} \rightarrow \Omega^n$.
We estimate

$$\begin{aligned}\|U^{n+1} - U\| &= \|\textcolor{red}{P^n}(U^{n+1} - U)\| \\ &\leq \|\textcolor{red}{P^n}(U^{n+1} - V^n)\| + \|\textcolor{red}{P^n}(U - V^n)\| \\ &\leq (1+1)\|\textcolor{blue}{P^n}(U - V^n)\| + \mathcal{O}(\|U - U^n\|^2) \\ &= 2\|\textcolor{red}{P^n}((U - U^n) - \mathcal{A}^*(\mathbf{y} - \mathcal{A}U^n))\| + \mathcal{O}(\|U - U^n\|^2) \\ &= 2(\|\textcolor{red}{P^n}((U - U^n) - \textcolor{red}{P^n}\mathcal{A}^*\mathcal{A}(U - U^n))\|) + \mathcal{O}(\|U - U^n\|^2) \\ &= 2\left(\|\textcolor{red}{P^n}((U - U^n - (\mathcal{A}\textcolor{red}{P^n})^*(\mathcal{A}\textcolor{red}{P^n}(U - U^n))))\|\right) + \mathcal{O}(\|U - U^n\|^2) \\ &\leq \rho\|\textcolor{red}{P^n}(U - U^n)\| \\ &= \rho\|(U - U^n)\|.\end{aligned}$$

For δ , $\|U - U_0\|$ sufficiently small

Since $\Omega^n = \textcolor{red}{P^n}(\mathcal{H}) \subset \mathcal{M}_{\leq 3r}$, for suff. small RICs $0 < \delta_{\leq 3r}$ there ex. $0 \leq \rho < 1$, and the series converges.

Iterative Hard Thresholding - Convergence Proof

$$U^{n+1} = H_U V^n, \quad V^n = U^n + \mathcal{A}^* \mathcal{A}(U - U^n)$$

$$\begin{aligned}\|U^{n+1} - V^n\| &= \|U_{best} - V^n\| + \mathcal{O}(\|U - U^n\|^2) \\ &\leq \|U - V^n\| + \mathcal{O}(\|U - U^n\|^2)\end{aligned}$$

Let $P^n U^{n+1} = V^n$ then $(I - P^n)U^{n+1} = 0$, $(I - P^n)U = 0$ and

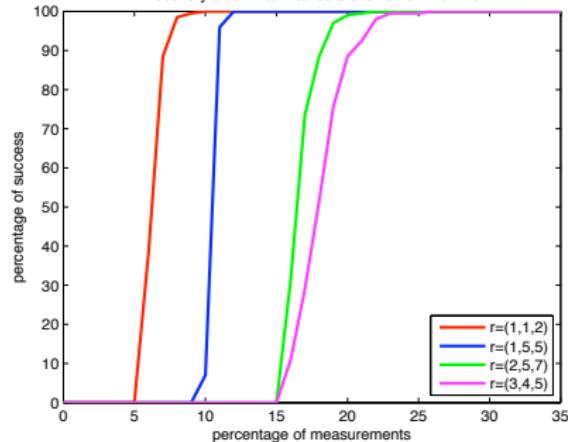
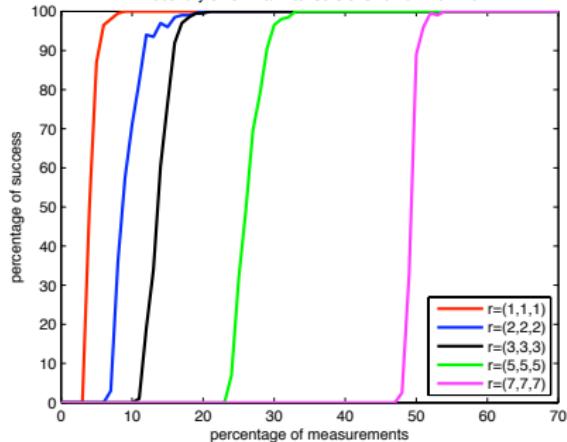
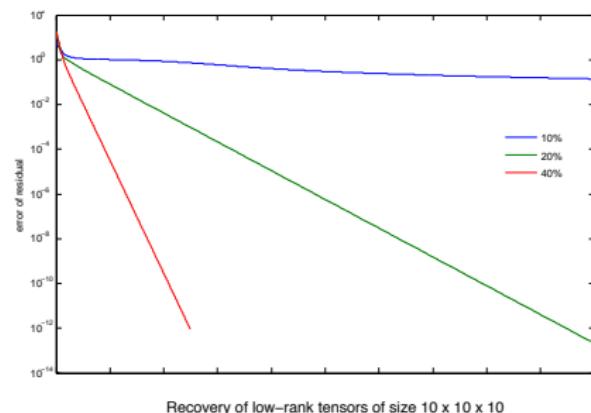
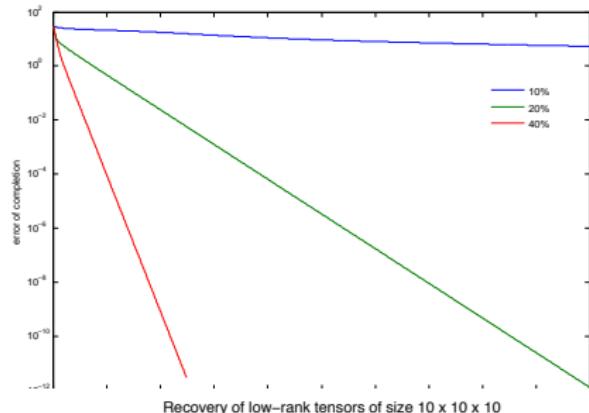
$$\begin{aligned}\|U^{n+1} - V^n\|^2 &= \|P^n(U^{n+1} - V^n)\|^2 + \|(I - P^n)(U^{n+1} - V^n)\|^2 \\ &\leq \|U - V^n\|^2 + \mathcal{O}(\|U - U^n\|^2) \\ &= \|P^n(U - V^n)\|^2 + \|(I - P^n)V^n\|^2 + \mathcal{O}(\|U - U^n\|^2)\end{aligned}$$

hence, by subtracting the last terms on both sides, we obtain

$$\|P^n(U^{n+1} - V^n)\|^2 \leq \|P^n(U - V^n)\|^2 + \mathcal{O}(\|U - U^n\|^2).$$

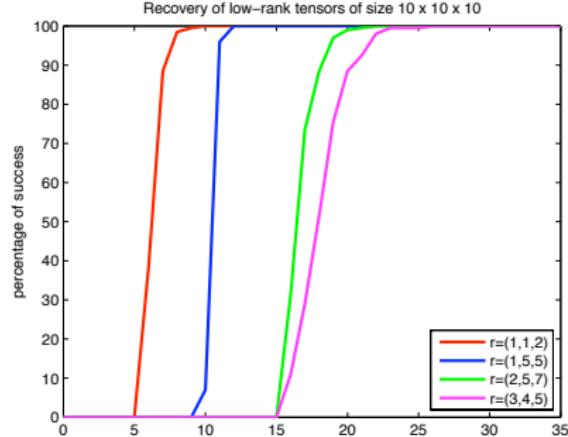
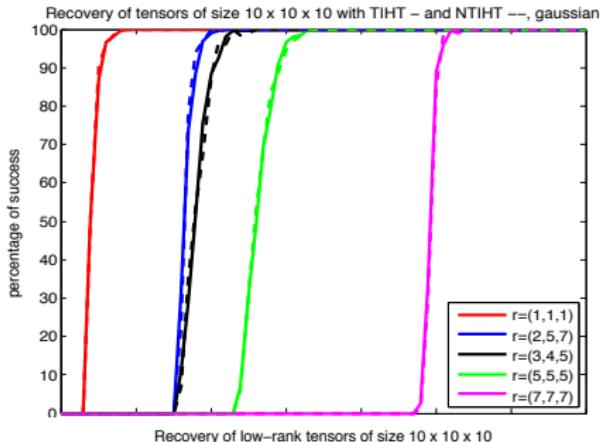
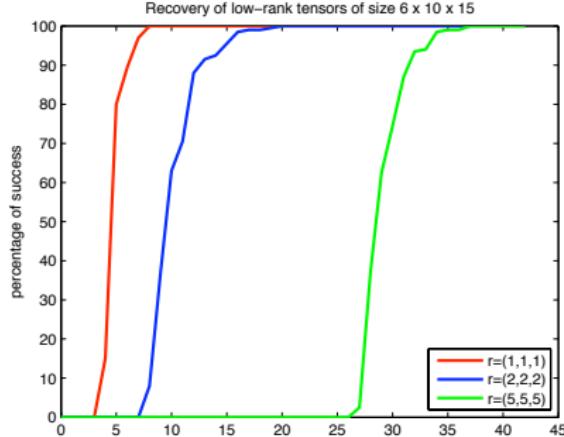
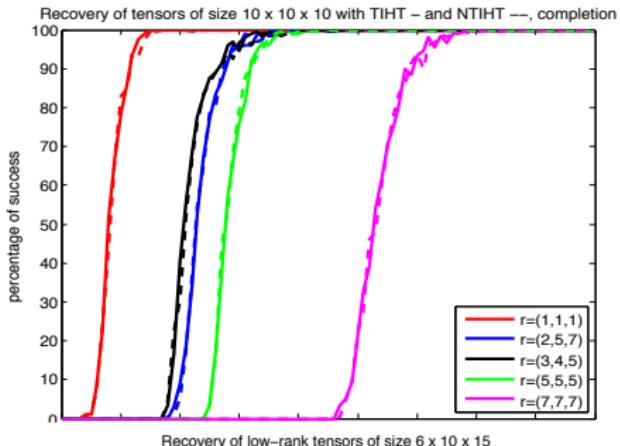
First numerical examples

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



Numerical examples

Stojanac Gaussian measurements



Hierarchical Tensor - Iterative Hard Thresholding - Remarks

- ▶ local convergence - but to global minimum
- ▶ global convergence - only conditional for (HT)-IHT
- ▶ Robustness ???
- ▶ TRIP (tensor RIP) not valid in general for completion ?
- ▶ Convex optimization formulation?
- ▶ the step size μ_n should be chosen adaptively (NIHT), see *Tanner & Wei* for matrix completion.
- ▶ Usually, the correct or appropriate (multi-)rank is not known in advance, and to be adapted through computation
- ▶ Notice that U^n is low rank and in tensor completion $\mathcal{A}^*(\mathbf{y} - \mathcal{A}U^n) = \mathcal{P}_M U - \mathcal{P}_M U^n$ is p -sparse,
 $M = \{\mathbf{k}_i : i = 1, \dots, p\}$.
- ▶ acceleration techniques enhancing convergence
- ▶

Thank you
for your attention.
