Low-Rank Tensor Completion by Riemannian Optimization

Daniel Kressner

CADMOS Chair of Numerical Algorithms and HPC
MATHICSE / SMA / SB
EPF Lausanne
http://anchp.epfl.ch

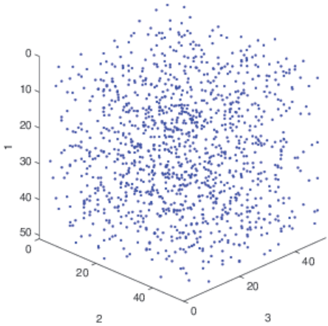
Joint work with
Michael Steinlechner (EPF Lausanne) and
Bart Vandereycken (U Genève)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

center for advanced modeling science

General Setting

Goal: Complete multidimensional data.



Applications:

- Completion of corrupted hyperspectral images, CT Scans, . . .
- Compression of multivariate functions with singularities
- Non-intrusive methods for stochastic/parametric PDEs
- Context-aware recommender systems
- . . .

General Setting

Goal: Complete multidimensional data.

Mathematical setting:

- Consider tensor $\mathcal{X}$ with very few entries known.
- Encode known entries by linear projection P_Ω.

Tensor reconstruction:

$$\min_{\mathcal{X}} \quad \frac{1}{2}\|\mathrm{P}_\Omega \mathcal{X} - \text{known entries}\|^2$$

- Ill-posed problem.
- Regularize with (multilinear) low-rank model for $\mathcal{X}$.

General Setting

Goal: Complete multidimensional data.

Mathematical setting:

- Consider tensor $\mathcal{X}$ with very few entries known.
- Encode known entries by linear projection P_Ω.

Low-rank tensor reconstruction:

$$\begin{array}{ll} \min\limits_{\mathcal{X}} & \frac{1}{2}\|\mathrm{P}_\Omega \mathcal{X} - \text{known entries}\|^2 \\ \text{subject to} & \mathcal{X} \in \mathcal{M}_{\mathbf{k}} := \{\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} : \mathrm{rank}(\mathcal{X}) = \mathbf{k}) \end{array}$$

- In this talk: Assume that $\mathcal{M}_{\mathbf{k}}$ is a smooth embedded submanifold.
- Multilinear ranks (Tucker, TT) OK.
  Tensor rank (CANDECOMP/PARAFAC) not OK.

Contents

1. Low-rank matrix completion
2. Low-rank tensor completion: Low order
3. Low-rank tensor completion: High order

Talk based on:

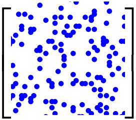
1. B. Vandereycken. Low-rank matrix completion by Riemannian optimization. SIAM Journal on Optimization, 23(2):1214–1236, 2013.
2. DK, M. Steinlechner, and B. Vandereycken. Low-rank tensor completion by Riemannian optimization. *BIT*, 54(2):447–468, 2014.
3. M. Steinlechner. Riemannian optimization for high-dimensional tensor completion. Technical report, 2015.

Papers and software available from `http://anchp.epfl.ch`.

Low-rank matrix completion by Riemannian optimization

Matrix Completion

$$\mathrm{P}_\Omega\, A =$$



$$\overset{\text{recover?}}{\rightsquigarrow} \; A$$

$$\mathrm{P}_\Omega : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}, \qquad \mathrm{P}_\Omega\, X = \begin{cases} X_{ij} & \text{if } \; (i,j) \in \Omega, \\ 0 & \text{else.} \end{cases}$$

Applications: image reconstruction, image inpainting, Netflix problem

Low-rank matrix completion:

$$\begin{array}{rll} \min\limits_{X} & \operatorname{rank}(X)\,, & X \in \mathbb{R}^{m\times n} \\ \text{subject to} & \mathrm{P}_\Omega\, X = \mathrm{P}_\Omega\, A & \end{array}$$

Low-rank matrix completion: *($\rightsquigarrow$ NP-Hard)*

$$\begin{aligned} \min_{X} \quad & \operatorname{rank}(X)\,, \qquad X \in \mathbb{R}^{m\times n} \\ \text{subject to} \quad & \mathrm{P}_\Omega\, X = \mathrm{P}_\Omega\, A \end{aligned}$$

Nuclear norm relaxation: *($\rightsquigarrow$ convex, but expensive)*

$$\begin{aligned} \min_{X} \quad & \|X\|_* = \sum_i \sigma_i\,, \qquad X \in \mathbb{R}^{m\times n} \\ \text{subject to} \quad & \mathrm{P}_\Omega\, X = \mathrm{P}_\Omega\, A \end{aligned}$$

Robust low-rank completion: *(Assume rank is known)*

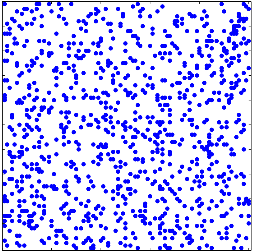
$$\begin{aligned} \min_{X} \quad & \frac{1}{2}\|\,\mathrm{P}_\Omega\, X - \mathrm{P}_\Omega\, A\|_F^2\,, \qquad X \in \mathbb{R}^{m\times n} \\ \text{subject to} \quad & \operatorname{rank}(X) = k \end{aligned}$$

Huge body of work! Overview: http://perception.csl.illinois.edu/matrix-rank/

Basic setting

$$\begin{array}{ll} \underset{X}{\text{minimize}} & f(X) := \frac{1}{2}\|P_\Omega(X - A)\|_F^2 \\ \text{subject to} & X \in \mathcal{M}_k := \left\{X \in \mathbb{R}^{m\times n} : \operatorname{rank}(X) = k\right\} \end{array}$$

$$\begin{aligned} P_\Omega : \quad & \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} \\ & X_{ij} \mapsto \begin{cases} X_{ij} & \text{if } (i,j) \in \Omega, \\ 0 & \text{if } (i,j) \notin \Omega. \end{cases} \end{aligned}$$



Manifold of Low-Rank Matrices

$$\mathcal{M}_k := \{X \in \mathbb{R}^{m \times n} : \text{rank}(X) = k\}$$

- $\mathcal{M}_k$ is a smooth manifold, e.g., [Bruns/Vetter'1988].
- Riemannian metric induced by inner product $\langle A, B\rangle = \text{tr}(A^T B)$.

$\rightsquigarrow$ Minimization on $\mathcal{M}_k$ by Riemannian optimization:

Constraint Optimization

$$\min_X \quad \frac{1}{2}\| \mathrm{P}_\Omega X - \mathrm{P}_\Omega A\|_F^2$$
$$X \in \mathbb{R}^{m \times n}$$
$$\text{subject to} \quad \text{rank}(X) = k$$

Riemannian Optimization

$$\min_X \quad \frac{1}{2}\| \mathrm{P}_\Omega X - \mathrm{P}_\Omega A\|_F^2$$
$$X \in \mathcal{M}_k$$

$\Rightarrow$ unconstrained!

- Newton-type [Simonsson/Eldén'2010], [Vandereycken/Vandewalle'2010].
- Trust-region methods for low-rank matrix completion [Boumal/Absil'2011].
- Nonlinear CG [Vandereycken'2012, Ngo/Saad'2012]
- Gradient descent [Journée et al.'2010, Mishra et al.'2012, Shalit/Weinshall/Chechik'2010].
- ...

Tangent space $T_X\mathcal{M}_k$

Consider SVD of rank-k matrix

$$X = \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V & V_\perp \end{bmatrix}^T, \qquad \Sigma \in \mathbb{R}^{k\times k}.$$

Tangent space of $\mathcal{M}_k$ at X:

$$T_X\mathcal{M}_k = \left\{ \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} \mathbb{R}^{k\times k} & \mathbb{R}^{k\times(n-k)} \\ \mathbb{R}^{(m-k)\times k} & 0 \end{bmatrix} \begin{bmatrix} V & V_\perp \end{bmatrix}^T \right\}$$

Riemannian gradient $\operatorname{grad} f(X) \in T_X\mathcal{M}_k$ defined by

$$\langle \operatorname{grad} f(X), \xi \rangle = \mathrm{D}\, f(X)[\xi] \quad \forall \xi \in T_X\mathcal{M}_k.$$

For $f(X) := \frac{1}{2}\|P_\Omega(X-A)\|_F^2$:

$$\operatorname{grad} f(X) = P_{T_X\mathcal{M}_k}\big(P_\Omega(X-A)\big)$$

with orthogonal projection $P_{T_X\mathcal{M}_k} : \mathbb{R}^{m\times n} \to T_X\mathcal{M}_k$.

Retraction

Gradient descent: $X \leftarrow X - \gamma \operatorname{grad} f(X) \notin \mathcal{M}_k$ 

Retraction

Gradient descent: $X \leftarrow X - \gamma \,\mathrm{grad}\, f(X) \notin \mathcal{M}_k$ 

Retraction $=$ Mapping $R_X : T_X\mathcal{M}_k \to \mathcal{M}_k$ such that

1. R is locally smooth on the tangent bundle
2. $R_X(0) = X$;
3. $\mathrm{D}R_X(0)[\xi] = \xi$ holds locally.

The metric projection

$$R_X(\xi) = P_X(X+\xi) = \arg\min_{Z\in\mathcal{M}_k} \|X+\xi-Z\|_F$$

is a retraction.

- Computed by truncated SVD [Absil/Malick'2010].
- Alternatives: orthographic projection; matching first terms of Taylor series expansion of exponential map.

Vector transport

Conjugate gradient method requires combination of gradients for subsequent iterates:

$$
\begin{aligned}
& \operatorname{grad} f(X) \in T_X \mathcal{M}_k, \quad \operatorname{grad} f(Y) \in T_Y \mathcal{M}_k \\
\Rightarrow \quad & \operatorname{grad} f(X) + \operatorname{grad} f(Y) \ ???
\end{aligned}
$$

Vector transport

Conjugate gradient method requires combination of gradients for subsequent iterates:

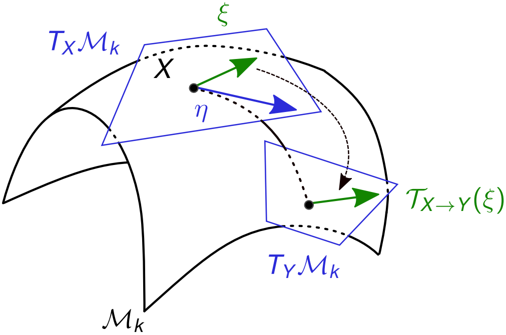
$$\operatorname{grad} f(X) \in T_X\mathcal{M}_k, \quad \operatorname{grad} f(Y) \in T_Y\mathcal{M}_k$$

$$\Rightarrow \quad \operatorname{grad} f(X) + \operatorname{grad} f(Y) \;???$$

Can be addressed by vector transport:

$$\mathcal{T}_{X\to Y} : T_X\mathcal{M}_k \to T_Y\mathcal{M}_k$$

$$\mathcal{T}_{X\to Y}(\xi) = P_{T_Y\mathcal{M}_k}(\xi).$$



Can be implemented in $O((m+n)k^2)$ ops.

Geometric nonlinear CG for matrix completion

Input: Initial guess $X_0 \in \mathcal{M}_k$.

$\eta_0 \leftarrow -\mathrm{grad}\, f(X_0)$

$\alpha_0 \leftarrow \mathrm{argmin}_\alpha\, f(X_0 + \alpha\eta_0)$

$X_1 \leftarrow R_{X_0}(\alpha_0\eta_0)$

for $i = 1, 2, \ldots$ **do**

Compute gradient:

$\xi_i \leftarrow \mathrm{grad}\, f(X_i)$

Conjugate direction by PR+ updating rule:

$\eta_i \leftarrow -\xi_i + \beta_i \mathcal{T}_{X_{i-1} \to X_i} f(\eta_{i-1})$

Initial step size from linearized line search:

$\alpha_i \leftarrow \mathrm{argmin}_\alpha\, f(X_i + \alpha\eta_i)$

Armijo backtracking for sufficient decrease:

Find smallest integer $m \geq 0$ such that

$f(X_i) - f(R_{X_i}(2^{-m}\alpha_i\eta_i)) \geq -1 \cdot 10^{-4} \langle \xi_i, 2^{-m}\alpha_i\eta_i \rangle$

Obtain next iterate:

$X_{i+1} \leftarrow R_{X_i}(2^{-m}\alpha_i\eta_i)$

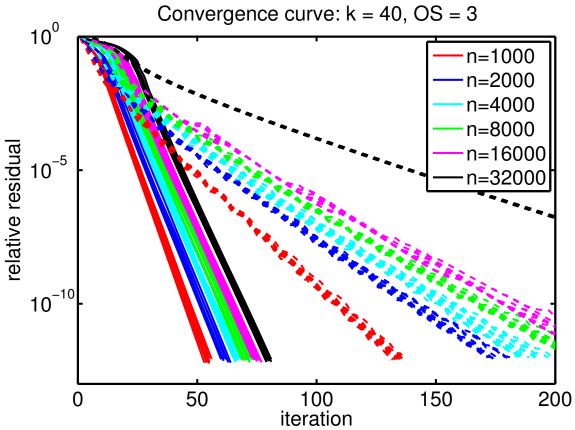
end for

Cost/iteration: $O((m+n)k^2 + |\Omega|k)$ ops.

Numerical experiments

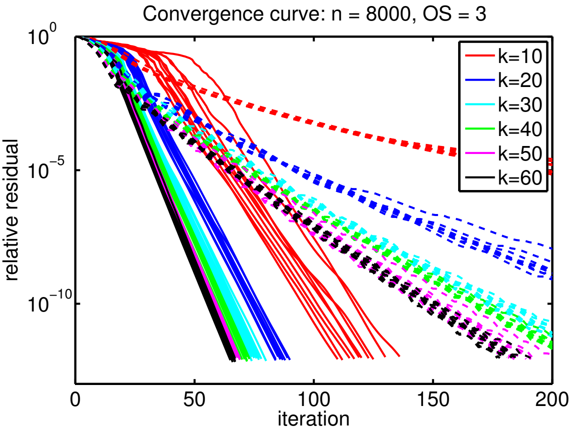
- Comparison to LMAFit [Wen/Yin/Zhang'2010].
  `http://lmafit.blogs.rice.edu/` .
- Oversampling factor $\mathrm{OS} = |\Omega|/(k(2n-k))$.
- Purely academic example $A = A_L A_R^T$ with $A_L, A_R = \texttt{randn}$.

Influence of n



- Dashed lines: LMAFit. Solid lines: Nonlinear CG.
- time(1 iteration of Nonlinear CG)
  $\approx$ 2$\times$ time(1 iteration of LMAFit)

Influence of rank



- Dashed lines: LMAFit. Solid lines: Nonlinear CG.
- time(1 iteration of Nonlinear CG)
  $\approx$ $2\times$ time(1 iteration of LMAFit)

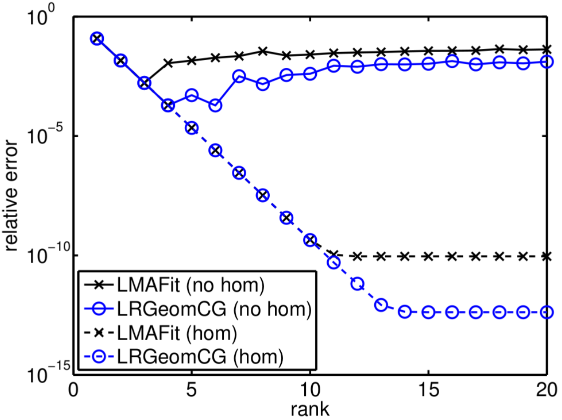
Numerical experiments

- Comparison to LMAFit [Wen/Yin/Zhang'2010]. `http://lmafit.blogs.rice.edu/` .
- Oversampling factor $\mathrm{OS} = |\Omega|/(k(2n-k)) = 8$.
- $8\,000 \times 8\,000$ matrix A is obtained from evaluating

$$f(x,y) = \frac{1}{1+|x-y|^2}$$

on $[0,1] \times [0,1]$.

Influence of rank



- Hom: Start with $k = 1$ and subsequently increase k, using previous result as initial guess.

Further remarks

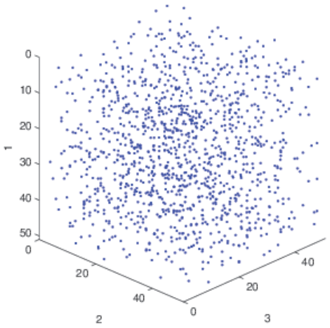
- Convergence analysis complicated by the fact that $\mathcal{M}_k$ is not closed.
- Second-order methods (Newton-like) require Hessian: painful and not of much help for low-rank matrix completion.
- Matrices generated by functions that are smooth only almost everywhere $\rightsquigarrow$ most low-rank matrix completion methods have difficulties in achieving high accuracy in such a setting.
- Potential way out: Adaptive choice of metric [Ngo/Saad'2012].

Low-rank tensor completion by Riemannian optimization low order

Tensor Completion

Low-rank tensor completion:

$$\begin{aligned} \min_{\mathcal{X}} \quad & \operatorname{rank}(\mathcal{X}), \qquad \mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_d} \\ \text{subject to} \quad & \mathrm{P}_\Omega \, \mathcal{X} = \mathrm{P}_\Omega \, \mathcal{A} \end{aligned}$$

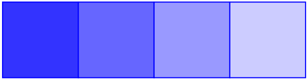


Applications:

- Completion of multidimensional data, e.g. hyperspectral images, CT Scans
- Compression of multivariate functions with singularities
- …

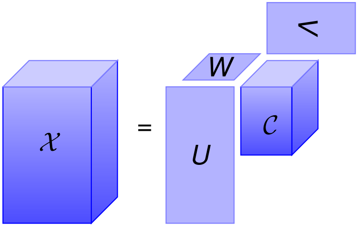
Multilinear Rank & Tucker Format

Reshape tensor into matrix by slicing, e.g. for first dimension:

$\mathcal{X} =$  $\rightsquigarrow \; X_{(1)} =$  $\in \mathbb{R}^{n_1 \times (n_2 \cdot n_3)}$

Multilinear rank of tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ defined by tuple

$$\mathbf{k} = (k_1, k_2, k_3), \quad \text{with} \quad k_i = \operatorname{rank}\big(X_{(i)}\big).$$



Representation of rank-**k**-tensor:

Tucker decomposition:

$$\mathcal{X} = \mathcal{C} \times_1 U \times_2 V \times_3 W$$

$U \in \mathbb{R}^{n_1 \times k_1}$, $V \in \mathbb{R}^{n_2 \times k_2}$, $W \in \mathbb{R}^{n_3 \times k_3}$,
and core tensor $\mathcal{C} \in \mathbb{R}^{k_1 \times k_2 \times k_3}$

Higher-Order SVD (HOSVD)

Goal: Approximate given tensor $\mathcal{X}$ by Tucker decomposition with prescribed multilinear rank $\mathbf{k} = (k_1, k_2, k_3)$.

1. Calculate SVD of matricizations:

$$X_{(\mu)} = \widetilde{U}_\mu \widetilde{\Sigma}_\mu \widetilde{V}_\mu^T \quad \text{for } \mu = 1, 2, 3.$$

2. Truncate basis matrices:

$$U_\mu := \widetilde{U}_\mu(:, 1 : k_\mu) \quad \text{for } \mu = 1, 2, 3.$$

3. Form core tensor:

$$\mathcal{C} := U_1^T \times_1 U_2^T \times_2 U_3^T \times_3 \mathcal{X}.$$

Truncated tensor produced by HOSVD [Lathauwer/De Moor/Vandewalle'2000]:

$$\widetilde{\mathcal{X}} = \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 U_3.$$

Quasi-optimality: $\|\mathcal{X} - \widetilde{\mathcal{X}}\| \leq \sqrt{d}\|\mathcal{X} - \mathcal{X}_{\text{best}}\|$.

Manifold of Low-Rank Tensors

$$\mathcal{M}_{\mathbf{k}} := \left\{\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d} : \mathrm{rank}(\mathcal{X}) = \mathbf{k}\right\},$$

$$\dim(\mathcal{M}_{\mathbf{k}}) = \prod_{j=1}^{d} k_j + \sum_{i=1}^{d} \left(k_i n_i - \frac{k_i(k_i-1)}{2}\right).$$

- $\mathcal{M}_{\mathbf{k}}$ is a smooth manifold. Discussed for more general formats in [Holtz/Rohwedder/Schneider'2012], [Uschmajew/Vandereycken'2012]
- Riemannian with metric induced by standard inner product $\langle \mathcal{X}, \mathcal{Y} \rangle = \langle \mathcal{X}_{(1)}, \mathcal{Y}_{(1)} \rangle$ *(sum of element-wise product)*

Manifold structure used in

- dynamical low-rank approximation [Koch/Lubich'2010], [Arnold/Jahnke'2012], [Lubich/Rohwedder/Schneider/Vandereycken'2012], [Khoromskij/Oseledets/Schneider'2012], ...
- best multilinear approximation [Eldén/Savas'2009], [Ishteva/Absil/Van Huffel/De Lathauwer'2011], [Curtef/Dirr/Helmke'2012]

Gradients and Tangent Space $T_{\mathcal{X}}\mathcal{M}_{\mathbf{k}}$

Every ξ in the tangent space $T_{\mathcal{X}}\mathcal{M}_{\mathbf{k}}$ at $\mathcal{X} = \mathcal{C} \times_1 U \times_2 V \times_3 W$ can be written as:

$$\begin{aligned}\xi = {} & \mathcal{S} \times_1 U \times_2 V \times_3 W \\ & + \mathcal{C} \times_1 U_\perp \times_2 V \times_3 W \\ & + \mathcal{C} \times_1 U \times_2 V_\perp \times_3 W \\ & + \mathcal{C} \times_1 U \times_2 V \times_3 W_\perp\end{aligned}$$

for some $\mathcal{S} \in \mathbb{R}^{k_1 \times k_2 \times k_3}$, $U_\perp \in \mathbb{R}^{n_1 \times k_1}$ with $U_\perp^T U = 0, \ldots$
Again, we obtain the Riemannian gradient of the objective function

$$f(\mathcal{X}) := \frac{1}{2} \| \mathrm{P}_\Omega \, \mathcal{X} - \mathrm{P}_\Omega \, \mathcal{A} \|_F^2$$

by projecting the Euclidean gradient into the tangent space:

$$\operatorname{grad} f(\mathcal{X}) = \mathrm{P}_{T_{\mathcal{X}}\mathcal{M}_{\mathbf{k}}}(\mathrm{P}_\Omega \, \mathcal{X} - \mathrm{P}_\Omega \, \mathcal{A})$$

Retraction

Retraction = Mapping $R_{\mathcal{X}} : T_{\mathcal{X}}\mathcal{M}_{\mathbf{k}} \to \mathcal{M}_{\mathbf{k}}$ such that

1. $R_{\mathcal{X}}$ is locally smooth wrt $\mathcal{X}$;
2. $R_{\mathcal{X}}(0) = \mathcal{X}$;
3. $\mathrm{D}R_{\mathcal{X}}(0)[\xi] = \xi$ holds locally.

Metric projection

$$R_{\mathcal{X}}(\xi) = P_{\mathcal{X}}(\mathcal{X} + \xi) = \operatorname*{arg\,min}_{\mathcal{Z} \in \mathcal{M}_{\mathbf{k}}} \|\mathcal{X} + \xi - \mathcal{Z}\|.$$

No closed-form solution available

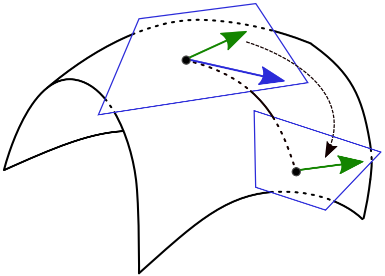
- Replaced by HOSVD truncation.
- Seems to work fine.
- HOSVD truncation is a retraction [K./Steinlechner/Vandereycken'14].

Vector transport

Conjugate gradient method requires combination of gradients for subsequent iterates:

$$\operatorname{grad} f(\mathcal{X}) \in T_{\mathcal{X}}\mathcal{M}_{\mathbf{k}}, \quad \operatorname{grad} f(\mathcal{Y}) \in T_{\mathcal{Y}}\mathcal{M}_{\mathbf{k}}$$

$$\Rightarrow \quad \operatorname{grad} f(X) + \operatorname{grad} f(Y) \text{ ???}$$



Can be addressed by vector transport:

$$\mathcal{T}_{\mathcal{X}\to\mathcal{Y}} : T_{\mathcal{X}}\mathcal{M}_{\mathbf{k}} \to T_{\mathcal{Y}}\mathcal{M}_{\mathbf{k}}$$

$$\mathcal{T}_{\mathcal{X}\to\mathcal{Y}}(\xi) = \mathrm{P}_{T_{\mathcal{Y}}\mathcal{M}_{\mathbf{k}}}(\xi).$$

Can be implemented in $O(nk^d)$ ops.

Geometric Nonlinear CG for Tensor Completion

Input: Initial guess $\mathcal{X}_0 \in \mathcal{M}_{\mathbf{k}}$.

$\eta_0 \leftarrow -\mathrm{grad}\, f(\mathcal{X}_0)$

$\alpha_0 \leftarrow \mathrm{argmin}_\alpha\, f(\mathcal{X}_0 + \alpha\eta_0)$

$\mathcal{X}_1 \leftarrow R_{\mathcal{X}_0}(\alpha_0\eta_0)$

for $i = 1, 2, \dots$ **do**

Compute gradient:

$\xi_i \leftarrow \mathrm{grad}\, f(\mathcal{X}_i)$

Conjugate direction by PR+ updating rule:

$\eta_i \leftarrow -\xi_i + \beta_i \mathcal{T}_{\mathcal{X}_{i-1}\to\mathcal{X}_i} f(\eta_{i-1})$

Initial step size from linearized line search:

$\alpha_i \leftarrow \mathrm{argmin}_\alpha\, f(\mathcal{X}_i + \alpha\eta_i)$

Armijo backtracking for sufficient decrease:

Find smallest integer $m \geq 0$ such that

$f(\mathcal{X}_i) - f(R_{\mathcal{X}_i}(2^{-m}\alpha_i\eta_i)) \geq -1 \cdot 10^{-4}\langle \xi_i, 2^{-m}\alpha_i\eta_i\rangle$

Obtain next iterate:

$\mathcal{X}_{i+1} \leftarrow R_{\mathcal{X}_i}(2^{-m}\alpha_i\eta_i)$

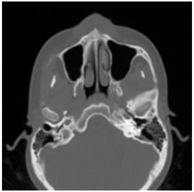
end for

Cost/iteration: $O(nk^d + |\Omega|k^{d-1})$ ops.

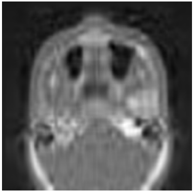
Reconstruction of CT Scan

$199 \times 199 \times 150$ tensor from scaled CT data set “INCISIX”,
(taken from OSIRIX MRI/CT data base
[www.osirix-viewer.com/datasets/])

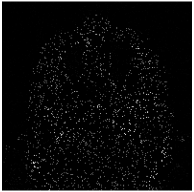
Slice of original tensor



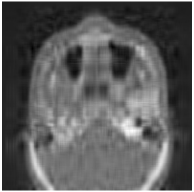
HOSVD approx. of rank 21



Sampled tensor (6.7%)



Low-rank completion of rank 21



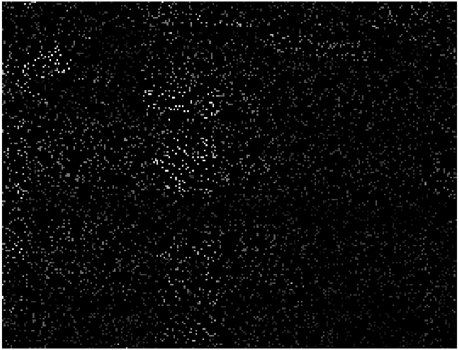
Compares very well with existing results w.r.t. low-rank recovery and speed, e.g., [Gandy/Recht/Yamada/’2011].

Hyperspectral Image

Set of photographs, (204 $\times$ 268 px) taken across a large range of wavelengths. 33 samples from ultraviolet to infrared [Image data: Foster et al.'2004]

Stacked into a tensor of size 204 $\times$ 268 $\times$ 33

10% of the Original Hyperspectral Imega Tensor, 16th Slice
Size of Tensor is [204, 268, 33]



Completed Tensor, 16th Slice
Final Rank is k = [50 50 6]

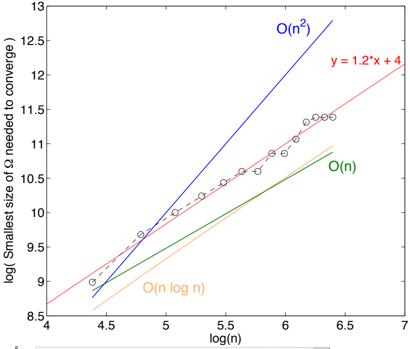
Here: Only 10% of entries known; [Signoretti et al.'2011] use 50%.

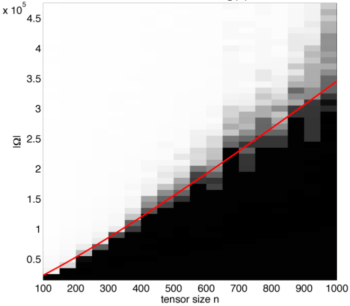
How many samples do we need?

Matrix case:
$O(n \cdot \log^{\beta} n)$ samples suffice!
[Candès/Tao'2009]
$\Rightarrow$ *Completion of tensor by applying matrix completion to matricization:* $O(n^2 \log(n))$. Gives upper bound!

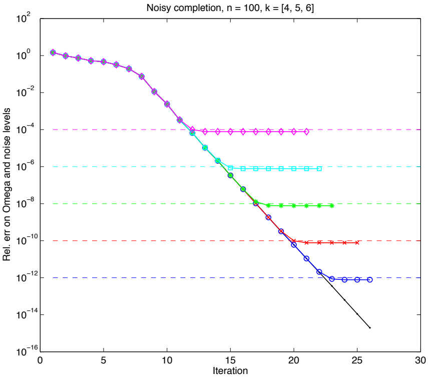
Tensor case:
Certainly: $|\Omega| \ll O(n^2)$
In all cases of convergence
$\rightsquigarrow$ exact reconstruction.

Conjecture: $|\Omega| = O(n \cdot \log^{\beta} n)$





Robustness of Convergence



- Random $100 \times 100 \times 100$ tensor of multilinear rank $(4, 5, 6)$ perturbed by white noise.
- Upon convergence $\rightsquigarrow$ reconstruction up to noise level.

Low-rank tensor completion by Riemannian optimization HIGH order

Going to high order

Applications leading to high-order tensors d:

- Stochastic and parameter-dependent PDEs [DK/Tobler'2011]
- Machine learning [Ishteva et al.]
- Learning of multivariate functions [Cevher et al.]
- ...

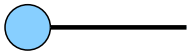
Tensor completion in Tucker format requires $O(nk^d + |\Omega|k^{d-1})$ operations.
$\rightsquigarrow$ Need other formats.

Formats described by tensor network diagrams:

- Introduced by Roger Penrose.
- Heavily used in quantum mechanics (spin networks).

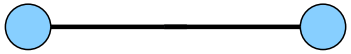
This is a scalar $\gamma \in \mathbb{R}$

This is a vector $x \in \mathbb{R}^n$



These are two vectors $x, y \in \mathbb{R}^n$

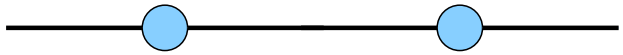
This is the inner product between $x, y \in \mathbb{R}^n$



$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$$

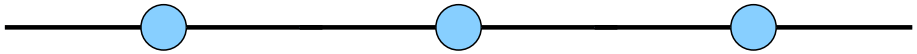
These are two matrices A, B

This is the matrix product $C = AB$



$$C_{ij} = \sum_{k=1}^{r} A_{ik} B_{kj}$$

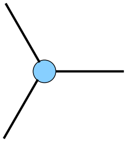
This is the matrix product $C = U\Sigma V^T$



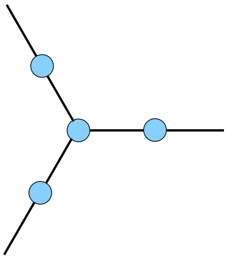
$$C_{ij} = \sum_{k=1}^{r} \sum_{\ell=1}^{r} U_{ik} \Sigma_{k\ell} V_{j\ell}$$

If $r \ll n$: Implicit representation of C via smaller matrices U, V, Σ.

This is a tensor $\mathcal{X}$ of order 3



This is a tensor $\mathcal{X}$ of order 3 in Tucker decomposition

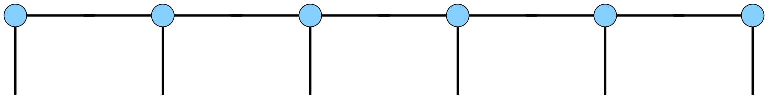


$$\mathcal{X}_{ijk} = \sum_{\ell_1=1}^{r_1} \sum_{\ell_2=1}^{r_2} \sum_{\ell_3=1}^{r_3} \mathcal{C}_{\ell_1\ell_2\ell_3} U_{i\ell_1} V_{j\ell_2} W_{k\ell_3}$$

Implicit representation of $\mathcal{X}$ via

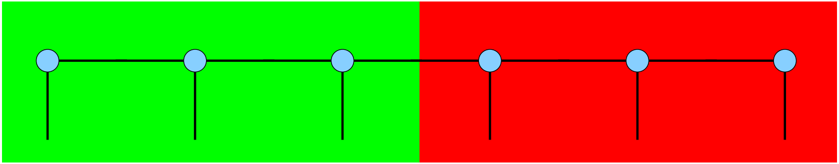
- $r_1 \times r_2 \times r_3$ core tensor $\mathcal{C}$
- $n_1 \times r_1$ matrix U spans first mode
- $n_2 \times r_2$ matrix V spans second mode
- $n_3 \times r_3$ matrix W spans third mode.

Six-dimensional tensor $\mathcal{X}$ in TT format



- $\mathcal{X}$ implicitly represented by four $r \times n \times r$ tensors and two $n \times r$ matrices
- Quantum mechanics: MPS (matrix product states)
- Matrix-based tensor formats introduced in numerical analysis by Grasedyck, Hackbusch, Kühn, Oseledets, Tyrtishnikov.

Six-dimensional tensor $\mathcal{X}$ in TT format



This partition corresponds to low-rank factorization

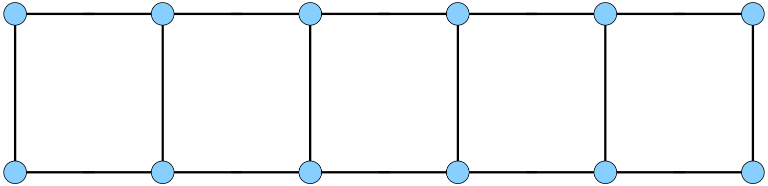
$$X^{(1,2,3)} = UV^T, \qquad X^{(1,2,3)} \in \mathbb{R}^{n_1 n_2 n_3 \times n_4 n_5 n_6},\ U \in \mathbb{R}^{n_1 n_2 n_3 \times r_3},\ V \in \mathbb{R}^{n_4 n_5 n_6 \times r_3}$$

$X^{(1,2,3)}$ is matricization of $\mathcal{X}$:

Merge multi-indices $(1,2,3)$ into row indices and
multi-indices $(4,5,6)$ into column indices

The ranks of $X^{(1,\ldots,\mu)}$ for $\mu = 1,\ldots,d-1$ are the TT ranks of $\mathcal{X}$.

Inner product of two tensors in TT decomposition



- Carrying out contractions requires $O(dnk^4)$ instead of $O(n^d)$ operations for tensors of order d.

Tensor completion in TT format

Low-rank tensor completion:

$$\begin{aligned} \min_{\mathcal{X}} \quad & \mathrm{rank}(\mathcal{X}), \qquad \mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d} \\ \text{subject to} \quad & \mathrm{P}_\Omega\, \mathcal{X} = \mathrm{P}_\Omega\, \mathcal{A} \end{aligned}$$

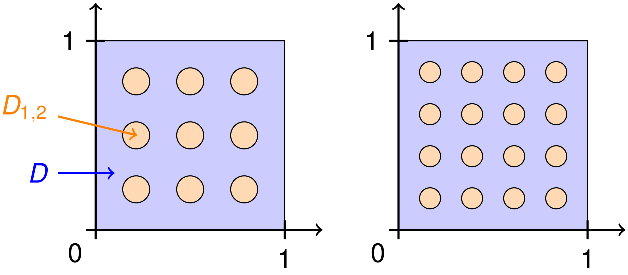
- Tensors of fixed TT rank **k** form a smooth manifold [Holtz/Rohwedder/Schneider'2012]:

$$\mathcal{M}_{\mathbf{k}} := \{\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_d} : \mathrm{rank}(\mathcal{X}) = \mathbf{k}\}.$$

$\rightsquigarrow$ Riemannian optimization can be applied and requires $O(d(n + |\Omega|)r^3)$ operations.

- See [Steinlechner'2015] for details.

Example: Cookie Problem



9 cookies

16 cookies

Consider heat equation:

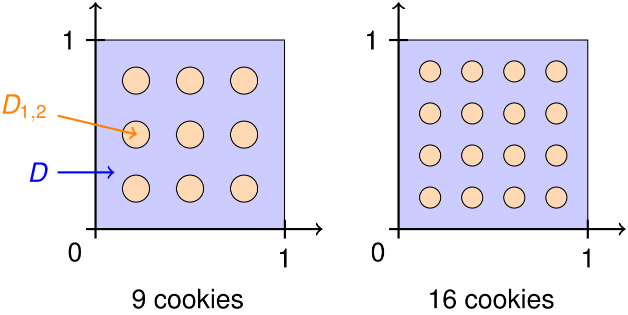
$$\begin{aligned} -\operatorname{div}(a(x,p)\nabla u(x,p)) &= 1, \qquad x \in D, \\ u(x,p) &= 0, \qquad x \in \partial D. \end{aligned}$$

with parametrized piecewise constant heat coefficient

$$a(x,p) := \begin{cases} p_\mu, & \text{if } x \in D_{s,t},\ \mu = m(t-1)+s, \\ 1, & \text{otherwise.} \end{cases}$$

d cookies $\rightsquigarrow$ d parameters $p_1, p_2, \ldots, p_d$.

Example: Cookie Problem



Quantity of interest: Average temperature

$$\overline{u}(p) := \int_{[0,1]^2} u(x,p)\,\mathrm{d}x.$$

- Discretize parameter space with tensorized Chebyshev polynomials with n nodes.
- Discrete values $\overline{u}(p)$ arranged in $n \times n \times \cdot \times n$ tensor $\mathcal{X}$.
- Each entry of $\mathcal{X}$ requires solution of PDE.
- Idea: Sample randomly and do rest by tensor completion!

Example: Cookie Problem

9 cookies

	TT-Cross		Riemannian optimization	
Tol.	Error	Eval.	Error	$\lvert\Omega\rvert$
10^{-3}	$8.35 \cdot 10^{-4}$	11681	$9.93 \cdot 10^{-5}$	1548
10^{-4}	$2.21 \cdot 10^{-5}$	14631	$8.30 \cdot 10^{-6}$	2784
10^{-5}	$1.05 \cdot 10^{-5}$	36291	$6.26 \cdot 10^{-6}$	3224
10^{-6}	$1.00 \cdot 10^{-6}$	42561	$6.50 \cdot 10^{-7}$	5338
10^{-7}	$1.31 \cdot 10^{-7}$	77731	$1.64 \cdot 10^{-7}$	9475

- TT cross approximation by Savostyanov/Oseledets'2011.
- Adaptive choice of ranks, but random choice of sample points in training set.
- Error measured on test set not included in training set.

Example: Cookie Problem

16 cookies

	TT-Cross		Riemannian optimization	
Tol.	Error	Eval.	Error	$\|\Omega\|$
10^{-3}	$8.17 \cdot 10^{-4}$	22951	$2.84 \cdot 10^{-4}$	2959
10^{-4}	$3.93 \cdot 10^{-5}$	68121	$2.10 \cdot 10^{-5}$	5261
10^{-5}	$9.97 \cdot 10^{-6}$	79961	$1.07 \cdot 10^{-5}$	8320
10^{-6}	$1.89 \cdot 10^{-6}$	216391	$1.89 \cdot 10^{-6}$	12736
10^{-7}	—	—	$7.12 \cdot 10^{-7}$	26010

- TT cross approximation by Savostyanov/Oseledets'2011.
- Adaptive choice of ranks, but random choice of sample points in training set.
- Error measured on test set not included in training set.

Conclusions

- Riemannian optimization competitive with state-of-the-art for low-rank matrix completion.
- Riemannian optimization much better than competitors for low-rank tensor completion.
- Nonlinear CG works well in practice. So far, only local convergence analysis.
- Most implementation details not discussed.
- Extension of Riemannian optimization to linear systems and eigenvalue problems using low-rank tensor formats in [DK/Steinlechner/Vandereycken'2015].