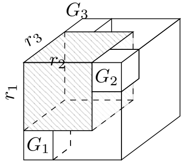
Low-rank optimization on matrix and tensor varieties

Bin Gao

Academy of Mathematics and Systems Science
Chinese Academy of Sciences

Joint work with
Renfeng Peng (AMSS, CAS)
Ya-xiang Yuan (AMSS, CAS)



$$\times_{k=1}^{3} \begin{bmatrix}\mathbf{U}_k & \mathbf{U}_{k,1} & \mathbf{U}_{k,2}\end{bmatrix}$$

Outline

Low-rank optimization on matrix/tensor spaces

Low-rank problems and applications

Low-rank problems

- Low-rank matrix/tensor completion [Wen-Yin-Zhang'12; Xu-Yin-Wen-Zhang'12; Kressner-Steinlechner-Vandereycken'14; Steinlechner'16; Kasai-Mishra'16; Shen-Liu'20; Dong-**G.**-Guan-Glineur'22; Zhao-Bai-Sun-Zheng'22; Yu-Zang-Huang'23; **G.**-Peng-Yuan'24]
- Low-rank approximation of higher-dimensional functions [Grasedyck-Kressner-Tobler'13; Uschmajew-Vandereycken'20]
- Low-rank solution of tensor equations [Kressner-Steinlechner-Vandereycken'16]
- Low-rank SDP [Lemon-So-Ye'16; Wang-Deng-Liu-Wen'23; Tang-Toh'23]
- Low-rank solution of high-dimensional PDEs [Eigel-Schneider-Sommer'22; Bachmayr-Eisenmann-Uschmajew'23; Wang-Lin-Liao-Liu-Xie'23]

Applications

- Recommendation system: movie ratings [Frolov-Oseledets'17]
- Hyperspectral Images [Zhang-He-Zhang-Shen-Yuan'13; Zhuang-Fu-Ng'21]
- Image and video inpainting [Bertalmio-Sapiro-Caselles-Ballester'00; Fu-Ruan-Luo-An-Jin'21; Luo-Zhao-Li-Ng-Meng'23; Bai-Zhang-Ni-Cui'16]
- EEG (brain signals) data [Mørup-Hansen-Herrmann-Parnas-Arnfred'06; Kong-Kong-Fan-Zhao-Cichoki'17]
- Magnetic resonance imaging (MRI) [Banco-Aeron-Hoge'16; Choi-Bao-Zhang'18; Fessler'20]
- Data analysis, e.g., Weather forecast [Loucheur-Absil-Journee'23] and Markov models [Zhu-Li-Wang-Zhang'22]

Matrix rank, singular value decomposition (SVD)

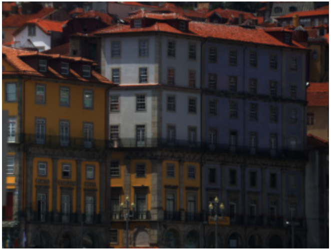
Low-rank matrix factorizations

- Input data (A): **Traffic** matrix (size: $\sim 10^3 \times 10^5$)
- Low-rank approx. by $\hat{X}_k := U_k \Sigma_k V_k^\top$ (truncated SVD) $\quad k = 10$

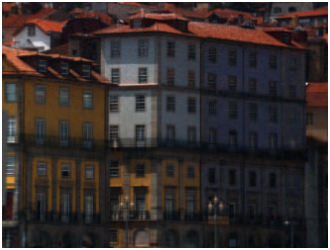
accuracy	$(1 - \frac{\left\| U_k \Sigma_k V_k^\top - A \right\|_F}{\| A \|_F})$:	$\mathbf{67.6\%}$
storage	$\frac{\text{\#parameters } (U_k, \Sigma_k, V_k)}{\text{\#parameters } (A)}$:	$\mathbf{1.1\%}$ only!

Low-rank assumption

- ☹ Store a full tensor: $\mathcal{O}(n^d)$ number of parameters!
- ☺ Low-rank tensor decomposition: save storage



Full image: 20MB



Compressed image: 0.4MB

Tucker rank: $[65, 65, 5]$ *Relative error: 0.0743*

Optimization on the set of fixed-rank matrices

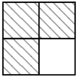
$$\begin{aligned} \min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \quad & f(\mathbf{X}) \\ \text{s. t.} \quad & \mathbf{X} \in \mathbb{R}_{\underline{r}}^{m\times n} := \{\mathbf{X} \in \mathbb{R}^{m\times n} : \operatorname{rank}(\mathbf{X}) = \underline{r}\} \end{aligned}$$

- $\mathbb{R}_{\underline{r}}^{m\times n}$: smooth *manifold* [Helmke-Shayman'95]
- $\underline{r} \leq \min(m, n)$: rank parameter

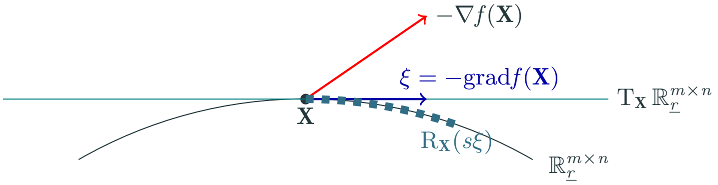
Tangent space [Vandereycken'13]

Given a thin SVD $\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\top}$

$$\mathrm{T}_{\mathbf{X}}\mathbb{R}_{\underline{r}}^{m\times n} = \begin{bmatrix}\mathbf{U} & \mathbf{U}^{\perp}\end{bmatrix} \begin{bmatrix}\mathbb{R}^{\underline{r}\times\underline{r}} & \mathbb{R}^{\underline{r}\times(n-\underline{r})} \\ \mathbb{R}^{(m-\underline{r})\times\underline{r}} & 0\end{bmatrix} \begin{bmatrix}\mathbf{V} & \mathbf{V}^{\perp}\end{bmatrix}^{\top}$$

$$= \begin{bmatrix}\mathbf{U} & \mathbf{U}^{\perp}\end{bmatrix}$$  $$\begin{bmatrix}\mathbf{V} & \mathbf{V}^{\perp}\end{bmatrix}^{\top}$$

Optimization on manifold $\mathbb{R}_{\underline{r}}^{m\times n}$



- Online-learning procedure [Shalit'12]
- Riemannian conjugate gradient descent [Vandereycken'13]
- Quotient geometry [Mishra-Meyer-Bonnabel-Sepulchre'14; Luo-Li-Zhang'23]

$\rightsquigarrow$ $\mathbb{R}_{\underline{r}}^{m\times n}$ is NOT closed!

How to choose a rank parameter?

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \quad f(\mathbf{X}) \qquad \text{s.t.} \quad \mathbf{X}\in\mathbb{R}^{m\times n}_{\leq r} := \{\mathbf{X}\in\mathbb{R}^{m\times n} : \operatorname{rank}(\mathbf{X}) \leq r\}$$

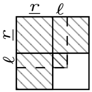
Set of bounded-rank matrices $\mathbb{R}^{m\times n}_{\leq r}$

- closure of $\mathbb{R}^{m\times n}_{r}$
- *real-algebraic variety*
- more flexible choices of rank

Tangent cone [Schneider-Uschmajew'15]

Given a thin SVD $\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$

$$\mathrm{T}_{\mathbf{X}}\mathbb{R}^{m\times n}_{\leq r} = \begin{bmatrix}\mathbf{U} & \mathbf{U}^\perp\end{bmatrix} \begin{bmatrix} \mathbb{R}^{\underline{r}\times\underline{r}} & \mathbb{R}^{\underline{r}\times(n-\underline{r})} \\ \mathbb{R}^{(m-\underline{r})\times\underline{r}} & \mathbb{R}^{(m-\underline{r})\times(n-\underline{r})}_{\leq r-\underline{r}} \end{bmatrix} \begin{bmatrix}\mathbf{V} & \mathbf{V}^\perp\end{bmatrix}^\top$$

$$= \begin{bmatrix}\mathbf{U} & \mathbf{U}_1 & \mathbf{U}_2\end{bmatrix}$$  $$\begin{bmatrix}\mathbf{V} & \mathbf{V}_1 & \mathbf{V}_2\end{bmatrix}^\top$$

with $[\mathbf{U}\ \ \mathbf{U}_1\ \ \mathbf{U}_2] \in \mathcal{O}(m)$, $[\mathbf{V}\ \ \mathbf{V}_1\ \ \mathbf{V}_2] \in \mathcal{O}(n)$, *and* $\ell = 0, 1, \ldots, r - \underline{r}$

Existing work

Line-search methods

- Projected gradient descent method [Schneider-Uschmajew'15]

$$\mathbf{X}^{(t+1)} = \mathrm{P}_{\mathbb{R}^{m\times n}_{\leq r}}\left(\mathbf{X}^{(t)} + \alpha^{(t)} \mathrm{P}_{\mathrm{T}_{\mathbf{X}^{(t)}}\mathbb{R}^{m\times n}_{\leq r}}(-\nabla f(\mathbf{X}^{(t)}))\right)$$

- Gradient sampling method [Hosseini-Uschmajew'19]
- Riemannian rank-adaptive method [**G.**-Absil'22]

Optimization on (product) manifold through a lift $(\mathbf{L}, \mathbf{R}) \mapsto \mathbf{L}\mathbf{R}^\top$

- Riemannian trust-region method [Levin-Kileel-Boumal'23]
- Gauss–Southwell type methods [Olikier-Uschmajew-Vandereycken'23]

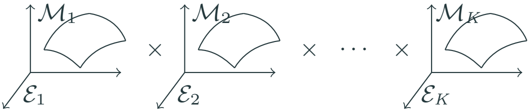
Low-rank optimization in semidefinite programming

- Riemannian method for SDP relaxation [Tang-Toh'23]

Tensor format: a view of product manifold

$\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$

$$\min_{\mathcal{X} \in \mathcal{M} := \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_K} f(\mathcal{X})$$



- CANDECOMP/PARAFAC (CP) decomposition [Hitchcock'27]

$$\mathcal{M} = \mathbb{R}^{n_1 \times r} \times \cdots \times \mathbb{R}^{n_d \times r}$$

- Tucker decomposition [Tucker'63]

$$\mathcal{M} = \mathrm{St}(r_1, n_1) \times \cdots \times \mathrm{St}(r_d, n_d) \times \mathbb{R}^{r_1 \times \cdots \times r_d}$$

- Tensor train decomposition [Oseledet'11]

$$\mathcal{M} = \mathbb{R}^{n_1 \times r_2} \times \mathbb{R}^{r_2 \times n_2 \times r_3} \times \cdots \times \mathbb{R}^{r_d \times n_d}$$

- Tensor ring decomposition [Zhao et al.'16]

$$\mathcal{M} = \mathbb{R}^{r_1 \times n_1 \times r_2} \times \cdots \times \mathbb{R}^{r_d \times n_d \times r_1}$$

$$\begin{array}{ll} \min_{\mathcal{X}} & f(\mathcal{X}) \\ \text{s.t.} & \mathcal{X} \in \mathcal{M}_{\mathbf{r}} \end{array}$$

Tensors with fixed Tucker rank

- A smooth manifold [Uschmajew and Vandereycken'13]
- Riemannian conjugate gradient method [Kressner-Steinlechner-Vandereycken'14]
- Riemannian conjugate gradient method under quotient geometry [Kasai-Mishra'16]

Tensors with fixed tensor train rank

- A smooth manifold [Uschmajew-Vandereycken'13]
- Riemannian conjugate gradient method [Steinlechner'16]
- Quotient geometry [Cai-Huang-Wang-Wei'22]

$\rightsquigarrow$ $\mathcal{M}_{\mathbf{r}}$ is NOT closed!

How to choose a rank parameter?

Optimization on tensor varieties

$$\begin{array}{ll} \min_{\mathcal{X}} & f(\mathcal{X}) \\ \text{s.t.} & \mathcal{X} \in \mathcal{M}_{\leq \mathbf{r}} \end{array}$$

$f : \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \to \mathbb{R}$: *a smooth function*

Tensor train varieties

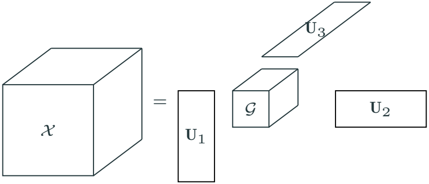
- Tangent cone [Kutschan'18]
- Rank-estimation method [Vermeylen-Olikier-Absil'23]

Tucker tensor varieties

- $\mathcal{M}_{\leq \mathbf{r}}$: real-algebraic varieties, closed
- Optimality condition [Luo-Qi'23]

$\rightsquigarrow$ Geometry of $\mathcal{M}_{\leq \mathbf{r}}$ is intricate!

Geometry of Tucker tensor varieties



$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$$

$\mathcal{G}$: *core tensor* $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$

$\mathbf{U}_k$: *"principle components"* $\mathbf{U}_k \in \mathbb{R}^{n_k \times r_k}$

Tucker decomposition [Tucker'63]

- Matrix case: $\mathbf{X} = \mathbf{S} \times_1 \mathbf{U} \times_2 \mathbf{V} = \mathbf{USV}^\top$
- Search space:

$$\mathrm{St}(r_1, n_1) \times \cdots \times \mathrm{St}(r_d, n_d) \times \mathbb{R}^{r_1 \times \cdots \times r_d}$$

- Storage:

$$n_1 r_1 + \cdots + n_d r_d + r_1 \cdots r_d$$

- Tucker rank: $\mathrm{rank}_{\mathrm{tc}}(\mathcal{X}) = (r_1, \ldots, r_d)$
- *Fixed-rank Tucker manifold*: $\mathcal{M}_{\mathbf{r}} = \{\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} : \mathrm{rank}_{\mathrm{tc}}(\mathcal{X}) = \mathbf{r}\}$

Geometry of fixed-rank Tucker manifold

Given $\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \cdots \times_d \mathbf{U}_d$ with $\mathrm{rank_{tc}}(\mathcal{X}) = \underline{\mathbf{r}} = \mathbf{r}$

Tangent space [Koch-Lubich'10]

$$\mathrm{T}_{\mathcal{X}}\mathcal{M}_{\mathbf{r}} = \left\{ \begin{array}{l} \dot{\mathcal{G}} \times_1 \mathbf{U}_1 \cdots \times_d \mathbf{U}_d + \sum_{k=1}^{d} \mathcal{G} \times_k \dot{\mathbf{U}}_k \times_{j \neq k} \mathbf{U}_j : \\ \dot{\mathcal{G}} \in \mathbb{R}^{r_1 \times \cdots \times r_d}, \dot{\mathbf{U}}_k \in \mathbb{R}^{n_k \times r_k}, \dot{\mathbf{U}}_k^\top \mathbf{U}_k = 0 \end{array} \right\}$$

A new reformulation

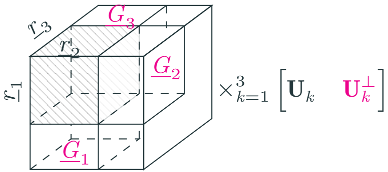
Given $\mathcal{V} \in \mathrm{T}_{\mathcal{X}}\mathcal{M}_{\mathbf{r}}$

$$\begin{aligned} \mathcal{V} &= \dot{\mathcal{G}} \times_1 \mathbf{U}_1 \cdots \times_d \mathbf{U}_d + \sum_{k=1}^{d} \mathcal{G} \times_k \dot{\mathbf{U}}_k \times_{j \neq k} \mathbf{U}_j \\ &= \dot{\mathcal{G}} \times_1 \mathbf{U}_1 \cdots \times_d \mathbf{U}_d + \sum_{k=1}^{d} \mathcal{G} \times_k (\mathbf{U}_k^\perp \mathbf{R}_k) \times_{j \neq k} \mathbf{U}_j \\ &= \dot{\mathcal{G}} \times_1 \mathbf{U}_1 \cdots \times_d \mathbf{U}_d + \sum_{k=1}^{d} (\mathcal{G} \times_k \mathbf{R}_k) \times_k \mathbf{U}_k^\perp \times_{j \neq k} \mathbf{U}_j \end{aligned}$$

Given $\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \cdots \times_d \mathbf{U}_d$ with $\mathrm{rank}_{\mathrm{tc}}(\mathcal{X}) = \underline{\mathbf{r}} = \mathbf{r}$

An illustration for third-order tensor ($d = 3$)

$\underline{G}_k := \mathcal{G} \times_k \mathbf{R}_k$



Orthogonal projection onto $\mathrm{T}_{\mathcal{X}}\mathcal{M}_{\mathbf{r}}$

$$\mathrm{P}_{\mathrm{T}_{\mathcal{X}}\mathcal{M}_{\mathbf{r}}} \mathcal{A} = \mathcal{A} \times_{k=1}^{d} \mathrm{P}_{\mathbf{U}_k} + \sum_{k=1}^{d} \mathcal{G} \times_k \left(\mathrm{P}_{\mathbf{U}_k}^{\perp} \left(\mathcal{A} \times_{j \neq k} \mathbf{U}_j^{\top} \right)_{(k)} \mathbf{G}_{(k)}^{\dagger} \right) \times_{j \neq k} \mathbf{U}_j$$

- Projection onto each “block”

Tangent cone of Tucker tensor varieties

Given $\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \cdots \times_d \mathbf{U}_d$ with $\mathrm{rank_{tc}}(\mathcal{X}) = \underline{\mathbf{r}} \leq \mathbf{r}$

Tucker tensor varieties: $\mathcal{M}_{\leq \mathbf{r}} = \{\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} : \mathrm{rank_{tc}}(\mathcal{X}) \leq \mathbf{r}\}$

(Bouligand) tangent cone

$$\mathrm{T}_{\mathcal{X}} \mathcal{M}_{\leq \mathbf{r}} := \{\mathcal{V} : \exists t^{(i)} \to 0,\ \mathcal{X}^{(i)} \to \mathcal{X} \text{ in } \mathcal{M}_{\leq \mathbf{r}},\ \text{s.t.}\ \frac{\mathcal{X}^{(i)} - \mathcal{X}}{t^{(i)}} \to \mathcal{V}\}$$
$$\subseteq \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$$

Connection to matrix varieties

$$\mathcal{M}_{\leq \mathbf{r}} = \bigcap_{k=1}^{d} \mathrm{ten}_{(k)} \left(\mathbb{R}^{n_k \times n_{-k}}_{\leq r_k}\right)$$

- A subset of the intersection of the tangent cone of matrix varieties

$$\mathrm{T}_{\mathcal{X}} \mathcal{M}_{\leq \mathbf{r}} \subseteq \bigcap_{k=1}^{d} \mathrm{ten}_{(k)} \left(\mathrm{T}_{\mathbf{X}_{(k)}} \mathbb{R}^{n_k \times n_{-k}}_{\leq r_k}\right)$$

An explicit parametrization

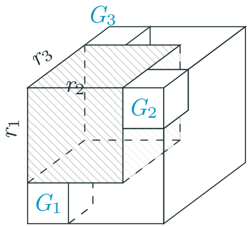
$\mathcal{C} \in \mathbb{R}^{r_1 \times \cdots \times r_d}$, $\mathbf{R}_{k,2} \in \mathbb{R}^{(n_k - r_k) \times \underline{r}_k}$, $\mathbf{U}_{k,1} \in \mathrm{St}(r_k - \underline{r}_k, n_k)$ *and* $\mathbf{U}_{k,2} \in \mathrm{St}(n_k - r_k, n_k)$ *are arbitrary that satisfy* $[\mathbf{U}_k \ \mathbf{U}_{k,1} \ \mathbf{U}_{k,2}] \in \mathcal{O}(n_k)$ *for* $k \in [d]$

$$\mathcal{V} = \mathcal{C} \times_{k=1}^{d} \begin{bmatrix} \mathbf{U}_k & \mathbf{U}_{k,1} \end{bmatrix} + \sum_{k=1}^{d} \mathcal{G} \times_k (\mathbf{U}_{k,2}\mathbf{R}_{k,2}) \times_{j \neq k} \mathbf{U}_j,$$

Characterization of tangent cone

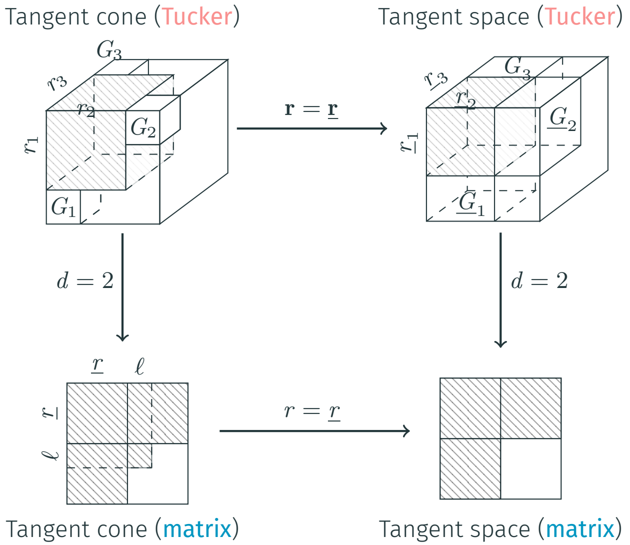
$$\mathrm{T}_{\mathcal{X}} \mathcal{M}_{\leq \mathbf{r}} = \bigcap_{k=1}^{d} \mathrm{ten}_{(k)} \left(\mathrm{T}_{\mathbf{X}_{(k)}} \mathbb{R}^{n_k \times n_{-k}}_{\leq r_k} \right)$$

An illustration for third-order tensor ($d = 3$)



$$\times_{k=1}^{3} \begin{bmatrix} \mathbf{U}_k & \mathbf{U}_{k,1} & \mathbf{U}_{k,2} \end{bmatrix}$$

Connection to matrix varieties



Geometric methods

Projected gradient descent (P2GD) [Matrix: Schneider-Uschmajew'15]

$$\mathcal{X}^{(t+1)} = \mathrm{P}_{\leq \mathbf{r}} \left(\mathcal{X}^{(t)} + s^{(t)} \mathrm{P}_{\mathrm{T}_{\mathcal{X}^{(t)}} \mathcal{M}_{\leq \mathbf{r}}} (-\nabla f(\mathcal{X}^{(t)})) \right)$$

$\rightsquigarrow$ Two projections are computationally intractable!

Approximate projections

- HOSVD instead of $\mathrm{P}_{\leq \mathbf{r}}$:

$$\mathrm{P}^{\mathrm{HO}}_{\leq \mathbf{r}}(\mathcal{A}) := \mathrm{P}^{d}_{\leq r_d}(\mathrm{P}^{d-1}_{\leq r_{d-1}} \cdots (\mathrm{P}^{1}_{\leq r_1}(\mathcal{A})))$$

- An approximate projection of $\mathrm{P}_{\mathrm{T}_{\mathcal{X}^{(t)}} \mathcal{M}_{\leq \mathbf{r}}}$:

$$\tilde{\mathrm{P}}_{\mathrm{T}_{\mathcal{X}} \mathcal{M}_{\leq \mathbf{r}}}(\mathcal{A}) = \mathcal{A} \times_{k=1}^{d} \mathrm{P}_{\tilde{\mathbf{S}}_k} + \sum_{k=1}^{d} \mathcal{G} \times_k \left(\mathrm{P}^{\perp}_{\tilde{\mathbf{S}}_k} \left(\mathcal{A} \times_{j \neq k} \mathbf{U}_j^{\top} \right)_{(k)} \mathbf{G}^{\dagger}_{(k)} \right) \times_{j \neq k} \mathbf{U}_j,$$

where $\tilde{\mathbf{S}}_k := [\mathbf{U}_k \; \tilde{\mathbf{U}}_{k,1}]$ is orthogonal

$$\mathcal{X}^{(t+1)} = \mathrm{P}^{\mathrm{HO}}_{\leq \mathbf{r}} \left(\mathcal{X}^{(t)} + s^{(t)} \tilde{\mathrm{P}}_{\mathrm{T}_{\mathcal{X}^{(t)}} \mathcal{M}_{\leq \mathbf{r}}} (-\nabla f(\mathcal{X}^{(t)})) \right)$$

Gradient-Related Approximate Projection method (GRAP)

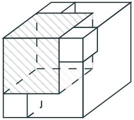
- Search direction: $g^{(t)} = \tilde{\mathrm{P}}_{\mathrm{T}_{\mathcal{X}^{(t)}} \mathcal{M}_{\leq \mathbf{r}}} (-\nabla f(\mathcal{X}^{(t)}))$
- Stepsize: $s^{(t)}$ Armijo backtracking line search
- Update: $\mathcal{X}^{(t+1)} = \mathrm{P}^{\mathrm{HO}}_{\leq \mathbf{r}} (\mathcal{X}^{(t)} + s^{(t)} g^{(t)})$

Retraction-free search directions

Revisiting the parametrization of tangent cone

$$\mathcal{V} = \mathcal{C} \times_{k=1}^{d} \begin{bmatrix} \mathbf{U}_k & \mathbf{U}_{k,1} \end{bmatrix} + \sum_{k=1}^{d} \mathcal{G} \times_k (\mathbf{U}_{k,2}\mathbf{R}_{k,2}) \times_{j\neq k} \mathbf{U}_j$$

$$= \mathcal{V}_0 + \sum_{k=1}^{d} \mathcal{V}_k$$



Surprising observations

$$\mathcal{X} + \mathcal{V}_0 = \mathcal{G} \times_{k=1}^{d} \mathbf{U}_k + \mathcal{C} \times_{k=1}^{d} \begin{bmatrix} \mathbf{U}_k & \mathbf{U}_{k,1} \end{bmatrix}$$

$$\in \bigotimes_{k=1}^{d} \mathrm{span}([\mathbf{U}_k \ \mathbf{U}_{k,1}]) \subseteq \mathcal{M}_{\leq \mathbf{r}}$$

$$\mathcal{X} + \mathcal{V}_k = \mathcal{G} \times_{i=1}^{d} \mathbf{U}_i + \mathcal{G} \times_k (\mathbf{U}_{k,2}\mathbf{R}_{k,2}) \times_{j\neq k} \mathbf{U}_k$$

$$= \mathcal{G} \times_k (\mathbf{U}_k + \mathbf{U}_{k,2}\mathbf{R}_{k,2}) \times_{j\neq k} \mathbf{U}_k \in \mathcal{M}_{\leq \mathbf{r}}$$

♠ None of any two combination is feasible!

$$\mathcal{X} + \mathcal{V}_0 + \mathcal{V}_k \notin \mathcal{M}_{\leq \mathbf{r}}$$

Retraction-free search directions

$$\mathrm{P}_0(\mathcal{A}) := \underset{\mathcal{V}_0}{\arg\min} \left\{ \|\mathcal{V}_0 - \mathcal{A}\| : \mathcal{V}_0 = \mathcal{C} \times_{k=1}^{d} \begin{bmatrix} \mathbf{U}_k & \mathbf{U}_{k,1} \end{bmatrix} \in \mathrm{T}_{\mathcal{X}} \mathcal{M}_{\leq \mathbf{r}} \right\}$$

$$\mathrm{P}_k(\mathcal{A}) := \underset{\mathcal{V}_k}{\arg\min} \left\{ \|\mathcal{V}_k - \mathcal{A}\| : \mathcal{V}_k = \mathcal{G} \times_k (\mathbf{U}_{k,2} \mathbf{R}_{k,2}) \times_{j \neq k} \mathbf{U}_j \in \mathrm{T}_{\mathcal{X}} \mathcal{M}_{\leq \mathbf{r}} \right\}$$

Approximations

Select basis $\tilde{\mathbf{U}}_{k,1}$, and $\tilde{\mathbf{S}}_k := [\mathbf{U}_k \ \tilde{\mathbf{U}}_{k,1}]$ is orthogonal

$$\tilde{\mathrm{P}}_0(\mathcal{A}) := \mathcal{A} \times_{k=1}^{d} \mathrm{P}_{\tilde{\mathbf{S}}_k},$$

$$\tilde{\mathrm{P}}_k(\mathcal{A}) := \mathcal{G} \times_k \left(\mathrm{P}_{\mathbf{U}_k}^{\perp} \left(\mathcal{A} \times_{j \neq k} \mathbf{U}_j^{\top} \right)_{(k)} \mathbf{G}_{(k)}^{\dagger} \right) \times_{j \neq k} \mathbf{U}_j.$$

Search direction

$$\hat{\mathrm{P}}_{\mathrm{T}_{\mathcal{X}} \mathcal{M}_{\leq \mathbf{r}}}(\mathcal{A}) := \underset{\mathcal{V} \in \{\tilde{\mathrm{P}}_0(\mathcal{A}), \ldots, \tilde{\mathrm{P}}_d(\mathcal{A})\}}{\arg\max} \|\mathcal{V}\|_{\mathrm{F}}.$$

$$\mathcal{X}^{(t+1)} = \mathcal{X}^{(t)} + s^{(t)} \hat{\mathrm{P}}_{\mathrm{T}_{\mathcal{X}^{(t)}} \mathcal{M}_{\leq \mathbf{r}}} (-\nabla f(\mathcal{X}^{(t)}))$$

Retraction-free GRAP method (rfGRAP)

- Search direction: $g^{(t)} = \hat{\mathrm{P}}_{\mathrm{T}_{\mathcal{X}^{(t)}} \mathcal{M}_{\leq \mathbf{r}}} (-\nabla f(\mathcal{X}^{(t)}))$
- Stepsize: $s^{(t)}$ Armijo backtracking line search
- Update: $\mathcal{X}^{(t+1)} = \mathcal{X}^{(t)} + s^{(t)} g^{(t)}$

Convergence

Global convergence

- Stationary measurement:

$$\lim_{t\to\infty} \| \mathrm{P}_{\mathrm{T}_{\mathcal{X}^{(t)}}\mathcal{M}_{\leq \mathbf{r}}}(-\nabla f(\mathcal{X}^{(t)}))\|_{\mathrm{F}} = 0$$

- Complexity: $\mathcal{O}(\epsilon^{-2})$ iterations to achieve

$$\|\mathrm{P}_{\mathrm{T}_{\mathcal{X}^{(t)}}\mathcal{M}_{\leq \mathbf{r}}}(-\nabla f(\mathcal{X}^{(t)}))\|_{\mathrm{F}} < \epsilon$$

Local convergence

- Assumption: *Łojasiewicz gradient inequality*

$$|f(\mathcal{X}) - f(\mathcal{Y})|^{1-\theta} \leq L\| \mathrm{P}_{\mathrm{T}_{\mathcal{X}}\mathcal{M}_{\leq \mathbf{r}}}(-\nabla f(\mathcal{Y}))\|_{\mathrm{F}}$$

- An accumulation point is a limit point
- If $\mathrm{rank}_{\mathrm{tc}}(\mathcal{X}^*) = \mathbf{r}$, then the stationary measure $\|\mathrm{grad} f(\mathcal{X}^*)\|_{\mathrm{F}} = 0$ and

$$\|\mathcal{X}^{(t)} - \mathcal{X}^*\|_{\mathrm{F}} \leq C \begin{cases} \mathrm{e}^{-ct}, & \text{if } \theta = \frac{1}{2}, \\ t^{-\frac{\theta}{1-2\theta}}, & \text{if } 0 < \theta < \frac{1}{2} \end{cases}$$

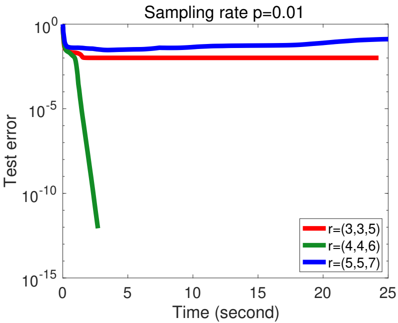
Tucker rank-adaptive method

Toy example on tensor completion

$$\min_{\mathcal{X} \in \mathcal{M}_{\mathbf{r}}} \|\mathrm{P}_\Omega(\mathcal{X}) - \mathrm{P}_\Omega(\mathcal{A})\|_{\mathrm{F}}^2$$

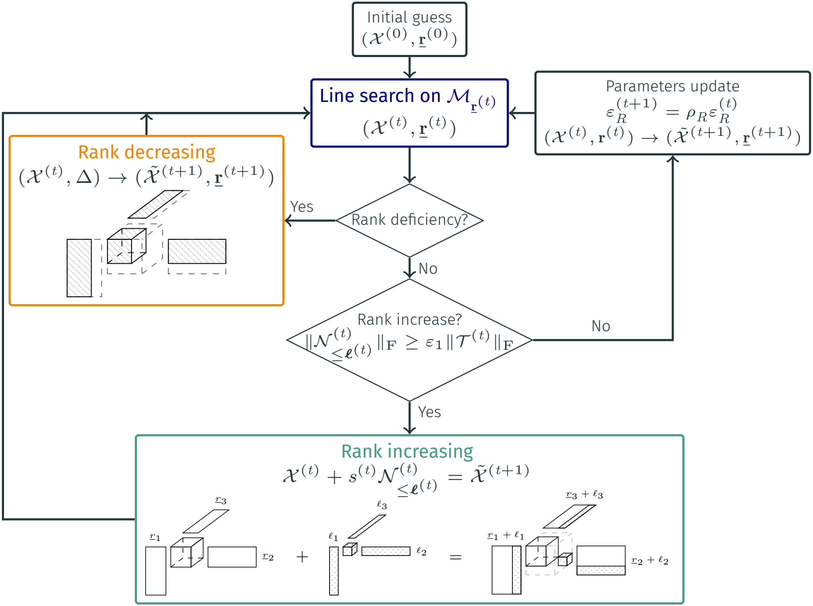
Performance of GeomCG [Kressner et al.'14]

- $\mathcal{A}$: synthetic Tucker tensor
- Size: $100 \times 100 \times 200$
- True rank: $\mathbf{r}^* = (4, 4, 6)$
- Rank parameters: $\mathbf{r} = (3, 3, 5), (4, 4, 6), (5, 5, 7)$

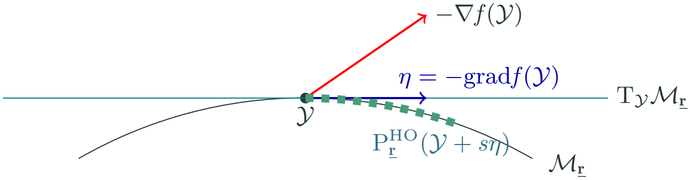


$\rightsquigarrow$ Performance is sensitive to rank selection!

Flowchart of Tucker rank-adaptive method



Step 1: Line search on fixed-rank manifold



Riemannian gradient descent method (RGD)

- Search direction: $\eta^{(t)} = -\mathrm{grad} f(\mathcal{Y}^{(t)})$
- Stepsize: $s^{(t)}$ Armijo backtracking line search
- Update: $\mathcal{Y}^{(t+1)} = \mathrm{P}_{\underline{\mathbf{r}}^{(t)}}^{\mathrm{HO}}(\mathcal{Y}^{(t)} + s^{(t)}\eta^{(t)})$

Step 2: Detection of rank deficiency

Given an iterate $\mathcal{X}$ generated by RGD

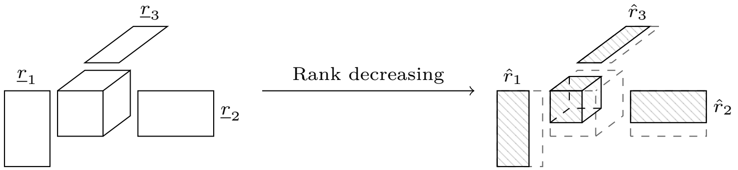
Ratio of largest and smallest singular values

$$\frac{\sigma_{1,k}}{\sigma_{\underline{r}_k,k}} > \Delta$$

- $\sigma_{1,k} \geq \cdots \geq \sigma_{\underline{r}_k,k}$: singular values of $\mathbf{X}_{(k)}^{(t)}$
- Δ: threshold

Rank-decreasing procedure

- Rank-$\hat{\mathbf{r}}$ truncation of $\mathcal{X}$: $\hat{r}_k := \min\left\{i : \frac{\sigma_{1,k}}{\sigma_{\underline{r}_k,k}} > \Delta\right\}$



Searching in "normals" *matrix case:* [G.-Absil'22]

- Decomposition: $\mathrm{N}_{\leq \boldsymbol{\ell}}(\mathcal{X}) := \mathcal{M}_{\leq \boldsymbol{\ell}} \cap \left(\bigotimes_{k=1}^{d} \mathrm{span}(\mathbf{U}_k)^{\perp}\right) \subseteq \mathrm{N}_{\mathcal{X}} \mathcal{M}_{\underline{\mathbf{r}}}$

$$\mathrm{T}_{\mathcal{X}} \mathcal{M}_{\underline{\mathbf{r}}} + \mathrm{N}_{\leq \boldsymbol{\ell}}(\mathcal{X}) \subseteq \mathrm{T}_{\mathcal{X}} \mathcal{M}_{\leq \mathbf{r}},$$

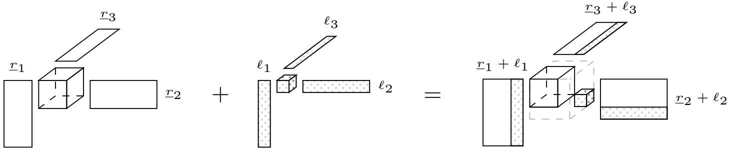
- Line search along $\mathcal{N}_{\leq \boldsymbol{\ell}} \in \mathrm{N}_{\leq \boldsymbol{\ell}}(\mathcal{X})$:

$$\underline{\mathbf{r}} < \mathrm{rank}_{\mathrm{tc}}(\mathcal{X} + s\mathcal{N}_{\leq \boldsymbol{\ell}}) \leq \mathbf{r}$$

Given bases $\mathbf{U}_{k,1} \in \mathrm{St}(\ell_k, n_k)$ and $0 < \ell < \mathbf{r} - \underline{\mathbf{r}}$

Rank-increasing procedure

- Compute $\mathcal{N}_{\leq \boldsymbol{\ell}} := -\nabla f(\mathcal{X}) \times_{k=1}^{d} \mathrm{P}_{\mathbf{U}_{k,1}} \in \mathrm{N}_{\leq \boldsymbol{\ell}}(\mathcal{X})$
- Stepsize: s Armijo backtracking line search
- Rank increasing: $\mathcal{X} + s\mathcal{N}_{\leq \boldsymbol{\ell}}$



A big picture

P2GD	$\mathcal{X}^{(t+1)} = \mathrm{P}_{\le \mathbf{r}} \left(\mathcal{X}^{(t)} + s^{(t)} \mathrm{P}_{\mathrm{T}_{\mathcal{X}^{(t)}} \mathcal{M}_{\le \mathbf{r}}} (-\nabla f(\mathcal{X}^{(t)})) \right)$

A big picture

P2GD	$\mathcal{X}^{(t+1)} = \mathrm{P}_{\leq \mathbf{r}} \left(\mathcal{X}^{(t)} + s^{(t)} \mathrm{P}_{\mathrm{T}_{\mathcal{X}^{(t)}} \mathcal{M}_{\leq \mathbf{r}}} (-\nabla f(\mathcal{X}^{(t)})) \right)$
GRAP	$\mathcal{X}^{(t+1)} = \mathrm{P}_{\leq \mathbf{r}}^{\mathrm{HO}} \left(\mathcal{X}^{(t)} + s^{(t)} \tilde{\mathrm{P}}_{\mathrm{T}_{\mathcal{X}^{(t)}} \mathcal{M}_{\leq \mathbf{r}}} (-\nabla f(\mathcal{X}^{(t)})) \right)$

A big picture

P2GD	$\mathcal{X}^{(t+1)} = \mathrm{P}_{\leq \mathbf{r}}\left(\mathcal{X}^{(t)} + s^{(t)}\mathrm{P}_{\mathrm{T}_{\mathcal{X}^{(t)}}\mathcal{M}_{\leq \mathbf{r}}}(-\nabla f(\mathcal{X}^{(t)}))\right)$
GRAP	$\mathcal{X}^{(t+1)} = \mathrm{P}_{\leq \mathbf{r}}^{\mathrm{HO}}\left(\mathcal{X}^{(t)} + s^{(t)}\tilde{\mathrm{P}}_{\mathrm{T}_{\mathcal{X}^{(t)}}\mathcal{M}_{\leq \mathbf{r}}}(-\nabla f(\mathcal{X}^{(t)}))\right)$
rfGRAP	$\mathcal{X}^{(t+1)} = \mathcal{X}^{(t)} + s^{(t)}\hat{\mathrm{P}}_{\mathrm{T}_{\mathcal{X}^{(t)}}\mathcal{M}_{\leq \mathbf{r}}}(-\nabla f(\mathcal{X}^{(t)}))$

A big picture

P2GD	$\mathcal{X}^{(t+1)} = \mathrm{P}_{\leq \mathbf{r}}\left(\mathcal{X}^{(t)} + s^{(t)} \mathrm{P}_{\mathrm{T}_{\mathcal{X}^{(t)}}\mathcal{M}_{\leq \mathbf{r}}}(-\nabla f(\mathcal{X}^{(t)}))\right)$
GRAP	$\mathcal{X}^{(t+1)} = \mathrm{P}^{\mathrm{HO}}_{\leq \mathbf{r}}\left(\mathcal{X}^{(t)} + s^{(t)} \tilde{\mathrm{P}}_{\mathrm{T}_{\mathcal{X}^{(t)}}\mathcal{M}_{\leq \mathbf{r}}}(-\nabla f(\mathcal{X}^{(t)}))\right)$
rfGRAP	$\mathcal{X}^{(t+1)} = \mathcal{X}^{(t)} + s^{(t)} \hat{\mathrm{P}}_{\mathrm{T}_{\mathcal{X}^{(t)}}\mathcal{M}_{\leq \mathbf{r}}}(-\nabla f(\mathcal{X}^{(t)}))$
TRAM	$\tilde{\mathcal{X}}^{(t)} \xrightarrow{\text{Line search on } \mathcal{M}_{\underline{\mathbf{r}}^{(t)}}} \mathcal{X}^{(t)} \xrightarrow{\text{Rank adjustment}} \tilde{\mathcal{X}}^{(t+1)}$

Numerical experiments: tensor completion

Tensor completion

- $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$: partially observed on $\Omega \subseteq [n_1] \times \cdots \times [n_d]$

  $[n_k] = \{1, 2, \cdots, n_k\},\ k = 1, \ldots, d$

Problem formulation

$$\begin{aligned} \min \quad & \frac{1}{2} \| \mathrm{P}_\Omega(\mathcal{X}) - \mathrm{P}_\Omega(\mathcal{A}) \|_\mathrm{F}^2 \\ \mathrm{s.\,t.} \quad & \mathcal{X} \in \mathcal{M}_{\leq \mathbf{r}}, \end{aligned}$$

- P_Ω: the projection operator onto Ω
- $\mathbf{r} = (r_1, r_2, \ldots, r_d)$: an array of d positive integers

Default settings

Running platform

- Workstation with two Intel(R) Xeon(R) Processors Gold 6330 @ 2.00GHz×28 and 512GB of RAM
- Matlab R2019b under Ubuntu 22.04.3

Compared methods

★ **GRAP:** gradient-related approximate-projection method *Tucker*

★ **rfGRAP:** retraction-free GRAP *Tucker*

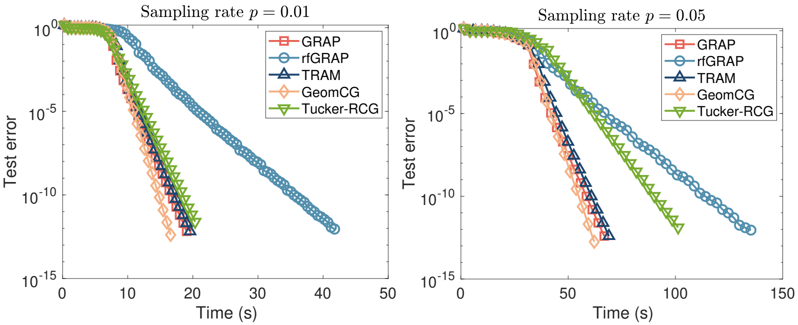
★ **TRAM:** Tucker rank-adaptive method *Tucker*

- **GeomCG:** Riemannian conjugate gradient *Tucker* [Kressner-Steinlechner-Vandereycken'14]
- **Tucker-RCG:** preconditioned Riemannian conjugate gradient *Tucker* [Kasai-Mishra'16]
- **CP-AltMin:** graph-based alternating minimization *CP* [Guan-Dong-**G.**-Absil-Glineur'20]
- **TT-RCG:** Riemannian conjugate gradient *TT* [Steinlechner'15]
- **TR-RGD:** preconditioned Riemannian gradient descent *TR* [**G.**-Peng-Yuan'24]

Synthetic data: test with true rank

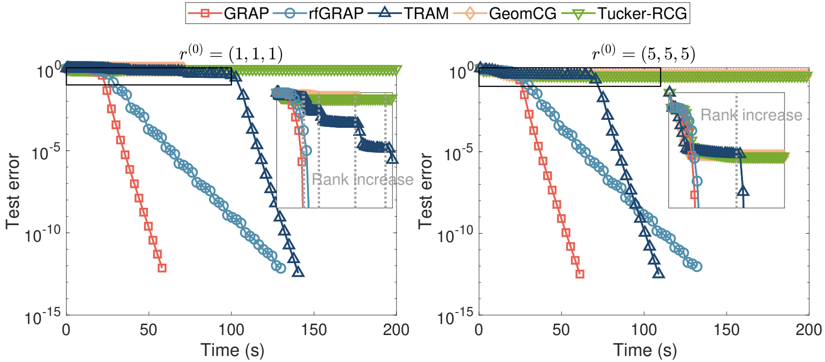
True rank $\mathbf{r} = \mathbf{r}^*$

- $\mathcal{A}$: Low-rank tensor with $n_1 = n_2 = n_3 = 3$, and Tucker rank $\mathbf{r}^* = (6, 6, 6)$
- $p = 0.01, 0.05$



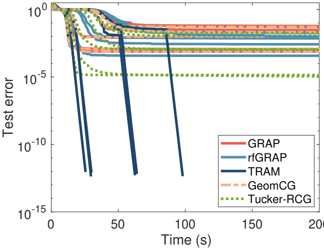
Under-estimated initial rank $\mathbf{r}^{(0)} < \mathbf{r}^*$

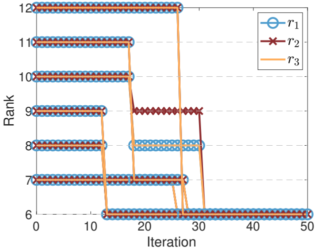
- $\mathcal{A}$: Low-rank tensor with $n_1 = n_2 = n_3 = 3$, and Tucker rank $\mathbf{r}^* = (6, 6, 6)$
- $p = 0.05$, $\mathbf{r}^{(0)} = (1, 1, 1), (5, 5, 5)$

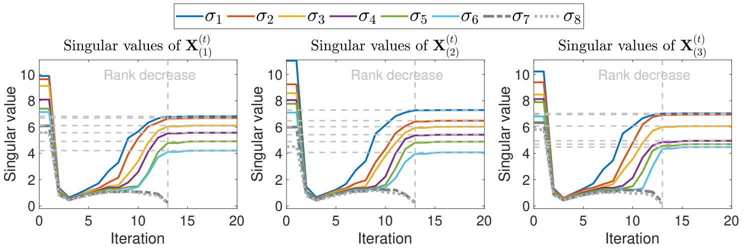


Over-estimated initial rank $\mathbf{r} > \mathbf{r}^*$

- $\mathcal{A}$: Low-rank tensor with $n_1 = n_2 = n_3 = 3$, and Tucker rank $\mathbf{r}^* = (6, 6, 6)$
- $p = 0.01$, $\mathbf{r} = (7, 7, 7), (8, 8, 8), \ldots, (12, 12, 12)$





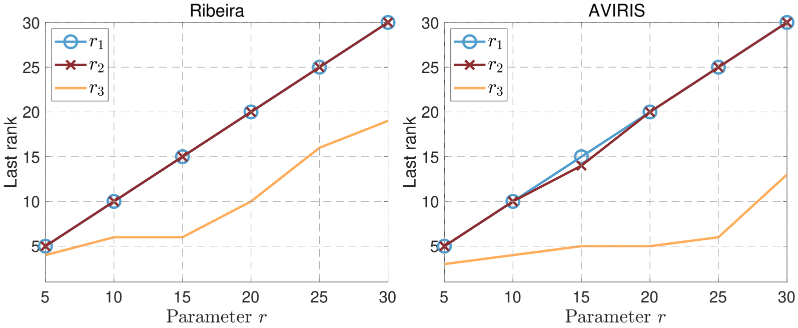


Two hyperspectral images

- Sampling rate $p = 0.1$.
- Tucker rank $\mathbf{r} = (5,5,5), (10,10,10), \ldots, (30,30,30)$

Left: "Ribeira" Right: "AVIRIS"

Last rank obtained in TRAM



$(15, 15, 6)$ coincides with fine-tuning (Kressner et al.'14)!

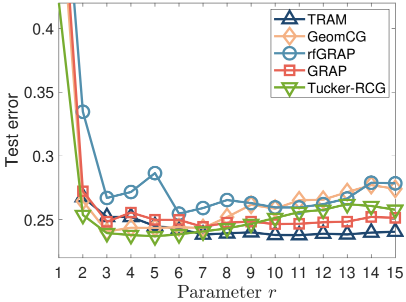
Hyperspectral images: relative errors and PSNR

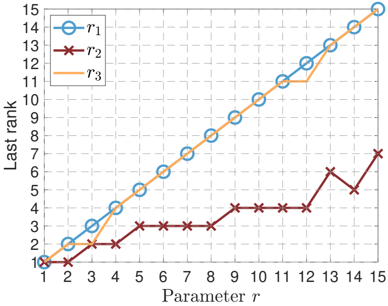
Tucker rank $\mathbf{r}$ (r_1, r_2, r_3)	Results	GRAP	rfGRAP	TRAM	GeomCG	Tucker-RCG
		"Ribeira"				
$(5, 5, 5)$	PSNR	24.9351	24.9325	24.9351	24.9351	24.9350
	relerr	0.2984	0.2985	0.2984	0.2984	0.2984
$(10, 10, 10)$	PSNR	26.8481	26.8482	26.8648	26.8483	26.8482
	relerr	0.2394	0.2394	0.2389	0.2394	0.2394
$(15, 15, 15)$	PSNR	28.3451	28.3450	28.4127	28.3451	28.3451
	relerr	0.2015	0.2015	0.1999	0.2015	0.2015
$(20, 20, 20)$	PSNR	29.3908	29.3934	29.5197	29.3917	29.3924
	relerr	0.1786	0.1786	0.1760	0.1786	0.1786
$(25, 25, 25)$	PSNR	30.2324	30.1852	30.3897	30.2315	30.2332
	relerr	0.1621	0.1630	0.1592	0.1622	0.1621
$(30, 30, 30)$	PSNR	30.7088	30.7182	30.9921	30.7579	30.7566
	relerr	0.1535	0.1533	0.1486	0.1526	0.1527
		"AVIRIS"				
$(5, 5, 5)$	PSNR	31.7181	31.7181	31.6955	31.7181	31.7181
	relerr	0.0835	0.0835	0.0837	0.0835	0.0835
$(10, 10, 10)$	PSNR	33.7393	33.7393	33.7517	33.7393	33.7394
	relerr	0.0661	0.0661	0.0660	0.0661	0.0661
$(15, 15, 15)$	PSNR	35.1308	35.1157	35.1427	35.1144	35.1251
	relerr	0.0564	0.0564	0.0563	0.0565	0.0564
$(20, 20, 20)$	PSNR	36.1776	36.1777	36.5438	36.1781	36.1780
	relerr	0.0500	0.0500	0.0479	0.0500	0.0500
$(25, 25, 25)$	PSNR	36.6010	36.6430	37.5433	36.6142	36.6002
	relerr	0.0476	0.0473	0.0427	0.0475	0.0476
$(30, 30, 30)$	PSNR	36.3106	36.4263	37.4879	36.1278	36.1505
	relerr	0.0492	0.0485	0.0430	0.0502	0.0501

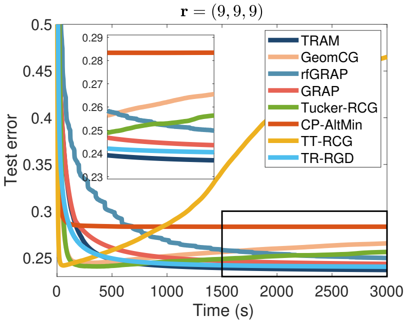
MovieLens 1M dataset

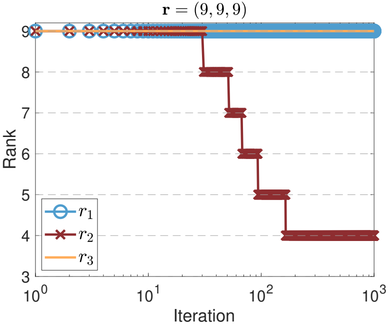
[https://grouplens.org/datasets/movielens/1m/]

- 6040 users, 3952 movies, 150 periods, 1M ratings
- $|\Omega| = 8 \times 10^5$, $|\Gamma| = 2 \times 10^5$, and $p = 2.23 \times 10^{-4}$
- $\mathbf{r} = (1,1,1), (2,2,2), \ldots, (15,15,15)$









Low-rank structure on mode two (movies)!

Conclusion and perspectives

Take-home notes

- Geometry of Tucker tensor varieties
- Geometric methods: GRAP and rfGRAP
- Rank-adaptive methods: TRAM
- Apocalypse-free methods converging to stationary points

References

1. **Bin Gao**, Renfeng Peng, Ya-xiang Yuan. *Low-rank optimization on Tucker tensor varieties.* arXiv:2311.18324, (2023)
2. **Bin Gao**, Renfeng Peng, Ya-xiang Yuan. *Optimization on product manifolds under a preconditioned metric.* arXiv:2306.08873, (2023)
3. **Bin Gao**, Renfeng Peng, Ya-xiang Yuan. *Riemannian preconditioned algorithms for tensor completion via tensor ring decomposition.* Computational Optimization and Applications, (2024)
4. Yu Guan, Shuyu Dong, **Bin Gao**, P.-A. Absil, François Glineur. *Alternating minimization algorithms for graph regularized tensor completion.* arXiv:2008.12876, (2023)
5. Shuyu Dong, **Bin Gao**, Yu Guan, François Glineur. *New Riemannian preconditioned algorithms for tensor completion via polyadic decomposition.* SIAM Journal on Matrix Analysis and Applications, 43-2 (2022), 840-866
6. **Bin Gao**, P.-A. Absil. *A Riemannian rank-adaptive method for low-rank matrix completion.* *Computational Optimization and Applications, 81 (2022), 67-90*

Thanks for your attention!

Email: gaobin@lsec.cc.ac.cn
Homepage: https://www.gaobin.cc