Optimization on product manifolds: preconditioned methods and applications

Bin Gao

Academy of Mathematics and Systems Science
Chinese Academy of Sciences

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

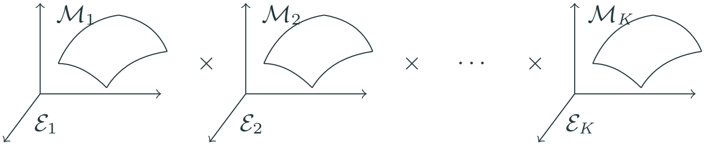
Outline

Optimization on product manifolds

Optimization on a product manifold

$$\min_{x \in \mathcal{M}} f(x)$$

$f : \mathcal{M} \to \mathbb{R}$: a smooth function



product manifold $\quad \mathcal{M} := \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_K$

Applications

- Canonical correlation analysis (CCA) [Yger-Berar-Gasso-Rakotomamonjy'12; Shustin-Aeron'23]

$$\mathcal{M} = \mathrm{St}_{\Sigma_{xx}}(m, d_x) \times \mathrm{St}_{\Sigma_{yy}}(m, d_y)$$

- Singular value decomposition (SVD) [Sato-Iwai'13]

$$\mathcal{M} = \mathrm{St}(p, m) \times \mathrm{St}(p, n)$$

- Joint approximate tensor diagonalization problem *maximize diagonal elements* [Usevich-Li-Comon'20]

$$\mathcal{M} = \times_{k=1}^{\ell} \mathrm{St}(r, n_k, \mathbb{C})$$

- Dimensionality reduction of EEG covariance matrices *EEG classification* [Yamamoto-Yger-Chevallier'21]

$$\mathcal{M} = \times_{k=1}^{\ell} \mathrm{St}(p, m)$$

- Matrix completion (MC) [Mishra-Apuroop-Sepulchre'12]

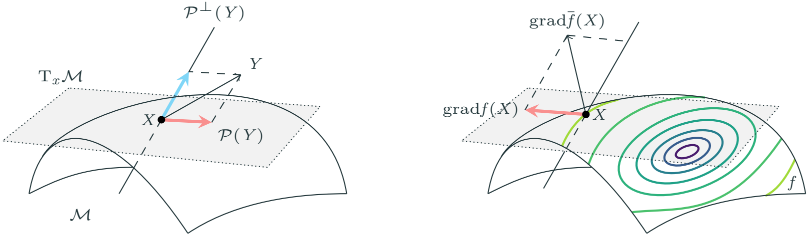
$$\mathcal{M} = \mathbb{R}_*^{n \times r} \times \mathbb{R}_*^{m \times r}$$

- Tensor ring completion (TRTC) [**G.**-Peng-Yuan'24]

$$\mathcal{M} = \mathbb{R}^{n_1 \times r_1 r_2} \times \mathbb{R}^{n_2 \times r_2 r_3} \times \cdots \times \mathbb{R}^{n_d \times r_d r_1}$$

- Tensor completion/decomposition problems [Kolda-Bader'09; Kasai-Mishra'16; Dong-**G.**-Guan-Glineur'22]

$$\mathcal{M} = \times_{k=1}^{3} \mathrm{St}(r_k, n_k) \times \mathbb{R}^{r_1 \times r_2 \times r_3}$$



Riemannian gradient descent method (RGD)

0. Develop Riemannian geometry — **metric** g
1. Search direction: $\xi = -\mathrm{grad}f(x) = -\mathrm{Proj}_{\mathrm{T}_x\mathcal{M}}(\mathrm{grad}\bar{f}(X))$
2. Stepsize: s
3. Retraction: $\mathrm{R}_x(s\xi)$

$g(\mathrm{grad}f(x), \eta) = \langle \nabla f(x), \eta \rangle, \ \eta \in \mathrm{T}_x\mathcal{M}$

$\rightsquigarrow$ Different metric, different gradient!

Motivation 2: alleviate ill-conditioning

Local convergence rate of RGD

RGD with fixed stepsize $x^{(t+1)} = \mathrm{R}_{x^{(t)}}(-\frac{1}{L}\mathrm{grad}f(x^{(t)}))$

- Strict local minima x^*: $\mathrm{grad}f(x^*) = 0$ and $\mathrm{Hess}f(x^*) \succ 0$
- Linear convergence rate: at most $1 - 1/\kappa_g(\mathrm{Hess}f(x^*))$

Metric-related Rayleigh quotient [Boumal'23]

- Rayleigh quotient of $\mathrm{Hess}f(x)$:

$$q_x(\xi) := \frac{g_x(\xi, \mathrm{Hess}f(x)[\xi])}{g_x(\xi, \xi)}$$

- Condition number: $\kappa_g(\mathrm{Hess}f(x)) := \lambda_{\max}/\lambda_{\min}$

  Eigenvalues: $\lambda_{\max} = \sup_{\xi \in \mathrm{T}_x\mathcal{M}} q_x(\xi)$; $\lambda_{\min} = \inf_{\xi \in \mathrm{T}_x\mathcal{M}} q_x(\xi)$

$\rightsquigarrow$ Different metric, different condition number!

Visualization

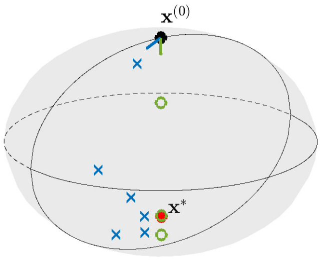
A toy example

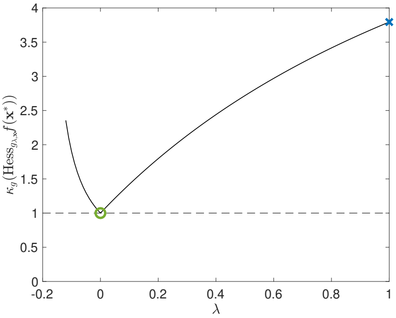
$$\min \ f(\mathbf{x}) := -\mathbf{b}^\top \mathbf{x}$$
$$\text{s.t.} \ \ \mathbf{x} \in \mathcal{M} := \{\mathbf{x} \in \mathbb{R}^n : \ \mathbf{x}^\top \mathbf{B}\mathbf{x} = 1\},$$

- $\mathbf{B} = \mathrm{diag}(2^2, 3^2, 1)$ *and* $\mathbf{b} = (1, 1, 1)$
- $\mathbf{x}^* = \mathbf{B}^{-1}\mathbf{b}/\|\mathbf{B}^{-1}\mathbf{b}\|_{\mathbf{B}}$: *closed-form solution*

Different metric, different performance

$$g_{\lambda,\mathbf{x}} := \langle \xi, (\lambda \mathbf{I}_n + (1-\lambda)\mathbf{B})\eta \rangle \quad \text{for tangent vectors } \xi \text{ and } \eta$$





Developing a preconditioned metric

Approximating the Newton direction

Inspired by matrix case: $\mathbf{A}$ is SPD and $\mathcal{M} = \mathbb{R}^n$

$$g_x(\xi, \eta) := \langle \xi, \mathbf{A}\eta \rangle \quad \longrightarrow \quad \mathrm{grad} f(x) = \mathbf{A}^{-1} \nabla f(x)$$

General manifold

Construct an operator $\bar{\mathcal{H}}$ on $\mathrm{T}\mathcal{E}$ such that

$\mathcal{E}$: the ambient space of $\mathcal{M}$

$$g_x(\xi, \eta) := \langle \xi, \bar{\mathcal{H}}(x)[\eta] \rangle \approx \langle \xi, \mathrm{Hess_e} f(x)[\eta] \rangle$$

$$\downarrow$$

$$\mathrm{grad}_g f(x) = \Pi_{g,x} \left(\bar{\mathcal{H}}(x)^{-1} [\nabla f(x)] \right) \approx (\mathrm{Hess_e} f(x))^{-1} [\mathrm{grad_e} f(x)]$$

- $\Pi_{g,x}$: *projection on tangent*
- $\mathrm{Hess_e}\, f$: *Riemannian Hessian under Euclidean metric*

$\rightsquigarrow$ **What is $\mathrm{Hess_e}\, f$ and how to approximate?**

Block-diagonal approximation

$$\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_K$$

Block structure of Riemannian Hessian on product manifolds

$$\begin{aligned}
\mathrm{Hess_e}\, f(x)[\eta] = (&H_{11}(x)[\eta_1] + H_{12}(x)[\eta_2] + \cdots + H_{1K}(x)[\eta_K], \\
&H_{21}(x)[\eta_1] + H_{22}(x)[\eta_2] + \cdots + H_{2K}(x)[\eta_K], \\
&\vdots \\
&H_{K1}(x)[\eta_1] + H_{K2}(x)[\eta_2] + \cdots + H_{KK}(x)[\eta_K])
\end{aligned}$$

Block-diagonal approximation

$$\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_K$$

Block structure of Riemannian Hessian on product manifolds

$$\begin{aligned}\operatorname{Hess_e} f(x)[\eta] = (&H_{11}(x)[\eta_1] + H_{12}(x)[\eta_2] + \cdots + H_{1K}(x)[\eta_K],\\ &H_{21}(x)[\eta_1] + H_{22}(x)[\eta_2] + \cdots + H_{2K}(x)[\eta_K],\\ &\vdots\\ &H_{K1}(x)[\eta_1] + H_{K2}(x)[\eta_2] + \cdots + H_{KK}(x)[\eta_K])\end{aligned}$$

Approximating "block-diagonal" terms

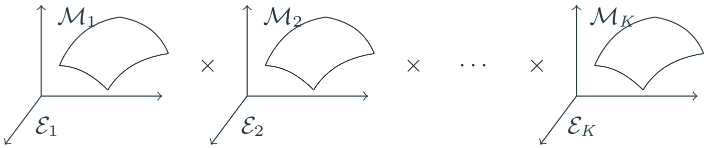
$\bar{\mathcal{H}}_k(x) \approx H_{kk}(x)$: *easy to construct; easy to compute inverse*

$$g_{x_k}(\xi_k, \eta_k) := \operatorname{tr}(\xi_k^\top \bar{\mathcal{H}}_k(x)[\eta_k]) \approx \langle \xi_k, H_{kk}(x)[\eta_k] \rangle$$

"Block-Jacobi" preconditioning in matrix case [Demmel'23]

- Block-diagonal matrix $\mathbf{D} := \operatorname{diag}(\mathbf{D}_{11}, \mathbf{D}_{22}, \ldots, \mathbf{D}_{KK})$

$$\mathbf{M} \to \mathbf{D}\mathbf{M}\mathbf{D}^\top$$



product manifold $\mathcal{M} := \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_K$

Constructing metric on each component $\mathcal{M}_i$

$$
\begin{aligned}
g_x(\xi, \eta) &:= g_{x_1}(\xi_1, \eta_1) + g_{x_2}(\xi_2, \eta_2) + \cdots + g_{x_K}(\xi_K, \eta_K) \\
&= \mathrm{tr}(\xi_1^\top \bar{\mathcal{H}}_1(x)[\eta_1]) + \cdots + \mathrm{tr}(\xi_k^\top \bar{\mathcal{H}}_k(x)[\eta_k])
\end{aligned}
$$

for $\xi, \eta \in \mathrm{T}_x\mathcal{M}$

Riemannian gradient under the preconditioned metric

$$\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_K$$

Non-Euclidean metric

$\xi, \eta \in \mathrm{T}_x\mathcal{M}$

$$g_x(\xi, \eta) := \sum_{k=1}^{K} \mathrm{tr}(\xi_k^\top \bar{\mathcal{H}}_k(x)[\eta_k]) \approx \sum_{k=1}^{K} \langle \xi_k, H_{kk}(x)[\eta_k] \rangle$$

Riemannian gradient

$$\begin{aligned} \mathrm{grad}_g f(x) = \Pi_{g,x}\Big(& \bar{\mathcal{H}}_1(x)^{-1}[\partial_1 f(x)], \\ & \bar{\mathcal{H}}_2(x)^{-1}[\partial_2 f(x)], \\ & \vdots \\ & \bar{\mathcal{H}}_K(x)^{-1}[\partial_K f(x)]\Big) \end{aligned}$$

$\Pi_{g,x} : \mathrm{T}_x\,\mathcal{E} \simeq \mathcal{E} \to \mathrm{T}_x\,\mathcal{M}$ *the orthogonal projection operator w.r.t. the metric* g *onto* $\mathrm{T}_x\,\mathcal{M}$

First-order methods

Riemannian gradient descent (RGD) method

- Search direction: $\eta^{(t)} = -\mathrm{grad}_g f(x^{(t)})$
- Stepsize: $s^{(t)}$
- Update: $x^{(t+1)} = \mathrm{R}_{x^{(t)}}(s^{(t)}\eta^{(t)})$

Riemannian conjugate gradient (RCG) method

- Search direction: $\eta^{(t)} = -\mathrm{grad}_g f(x^{(t)}) + \beta^{(t)}\mathcal{T}_{t\leftarrow t-1}\eta^{(t-1)}$

  $\mathcal{T}_{t\leftarrow t-1}$: vector transport; $\beta^{(t)}$: CG parameter
- Stepsize: $s^{(t)}$
- Update: $x^{(t+1)} = \mathrm{R}_{x^{(t)}}(s^{(t)}\eta^{(t)})$

Works interpreted by preconditioned metrics

Problem and methods	Search space $\mathcal{M}$ and variable	Metric $g_x(\xi, \eta)$, $\xi, \eta \in \mathrm{T}_x \mathcal{M}$
MC [Mishra-Apuroop-Sepulchre'12] **RGD, RCG, RTR**	$\mathbb{R}_*^{m\times r} \times \mathbb{R}_*^{n\times r}$ $(\mathbf{L}, \mathbf{R})$	$\langle \xi_1, \eta_1(\mathbf{R}^\top \mathbf{R})\rangle + \langle \xi_2, \eta_2(\mathbf{L}^\top \mathbf{L})\rangle$
Matrix sensing [Tong-Ma-Chi'21] **ScaledGD**	$\mathbb{R}_*^{m\times r} \times \mathbb{R}_*^{n\times r}$ $(\mathbf{L}, \mathbf{R})$	$\langle \xi_1, \eta_1(\mathbf{R}^\top \mathbf{R})\rangle + \langle \xi_2, \eta_2(\mathbf{L}^\top \mathbf{L})\rangle$
Tucker TC [Kasai-Mishra'16] **RCG**	$\times_{k=1}^{3} \mathrm{St}(r_k, n_k) \times \mathbb{R}^{r_1\times r_2\times r_3}$ $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathcal{G})$	$\sum_{k=1}^{3} \langle \xi_k, \eta_k(\mathbf{G}_{(k)}\mathbf{G}_{(k)}^\top)\rangle + \langle \xi_{\mathcal{G}}, \eta_{\mathcal{G}}\rangle$
CP TC [Dong-**G.**-Guan-Glineur'22] **RGD, RCG**	$\times_{k=1}^{d} \mathbb{R}^{n_k\times r}$ $(\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_d)$	$\sum_{k=1}^{d} \langle \xi_k, \eta_k((\mathbf{U}^{\odot j\neq k})^\top \mathbf{U}^{\odot j\neq k} + \delta \mathbf{I}_r)\rangle$
TT TC [Cai-Huang-Wang-Wei'22] **RGD, RCG, RGN**	$\times_{k=1}^{d} \mathbb{R}_*^{r_{k-1}\times n_k\times r_k}$ $(\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_d)$	$\sum_{k=1}^{d} \langle \xi_k, \eta_k(\mathbf{H}_k^\top \mathbf{H}_k)\rangle$
TR TC [**G.**-Peng-Yuan'24] **RGD, RCG**	$\times_{k=1}^{d} \mathbb{R}^{n_k\times r_{k-1}r_k}$ $(\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_d)$	$\sum_{k=1}^{d} \langle \xi_k, \eta_k(\mathbf{W}_{\neq k}^\top \mathbf{W}_{\neq k} + \delta \mathbf{I}_{r_{k-1}r_k})\rangle$
CCA [Yger et al.'12; Shustin-Aeron'23] **RCG**	$\mathrm{St}_{\Sigma_{xx}}(m, d_x) \times \mathrm{St}_{\Sigma_{yy}}(m, d_y)$ $(\mathbf{U}, \mathbf{V})$	$\langle \xi_1, \Sigma_{xx}\eta_1\rangle + \langle \xi_2, \Sigma_{yy}\eta_2\rangle$

MC=matrix completion; TC=tensor completion

Works interpreted by preconditioned metrics

Problem and methods	Search space $\mathcal{M}$ and variable	Metric $g_x(\xi, \eta)$, $\xi, \eta \in \mathrm{T}_x \mathcal{M}$
MC [Mishra-Apuroop-Sepulchre'12] **RGD, RCG, RTR**	$\mathbb{R}_*^{m\times r} \times \mathbb{R}_*^{n\times r}$ $(\mathbf{L}, \mathbf{R})$	$\langle \xi_1, \eta_1 (\mathbf{R}^\top \mathbf{R}) \rangle + \langle \xi_2, \eta_2 (\mathbf{L}^\top \mathbf{L}) \rangle$
Matrix sensing [Tong-Ma-Chi'21] **ScaledGD**	$\mathbb{R}_*^{m\times r} \times \mathbb{R}_*^{n\times r}$ $(\mathbf{L}, \mathbf{R})$	$\langle \xi_1, \eta_1 (\mathbf{R}^\top \mathbf{R}) \rangle + \langle \xi_2, \eta_2 (\mathbf{L}^\top \mathbf{L}) \rangle$
Tucker TC [Kasai-Mishra'16] **RCG**	$\times_{k=1}^3 \mathrm{St}(r_k, n_k) \times \mathbb{R}^{r_1\times r_2\times r_3}$ $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathcal{G})$	$\sum_{k=1}^{3} \langle \xi_k, \eta_k (\mathbf{G}_{(k)} \mathbf{G}_{(k)}^\top) \rangle + \langle \xi_{\mathcal{G}}, \eta_{\mathcal{G}} \rangle$
CP TC [Dong-**G.**-Guan-Glineur'22] **RGD, RCG**	$\times_{k=1}^d \mathbb{R}^{n_k\times r}$ $(\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_d)$	$\sum_{k=1}^{d} \langle \xi_k, \eta_k ((\mathbf{U}^{\odot j\neq k})^\top \mathbf{U}^{\odot j\neq k} + \delta \mathbf{I}_r) \rangle$
TT TC [Cai-Huang-Wang-Wei'22] **RGD, RCG, RGN**	$\times_{k=1}^d \mathbb{R}_*^{r_{k-1}\times n_k\times r_k}$ $(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_d)$	$\sum_{k=1}^{d} \langle \xi_k, \eta_k (\mathbf{H}_k^\top \mathbf{H}_k) \rangle$
TR TC [**G.**-Peng-Yuan'24] **RGD, RCG**	$\times_{k=1}^d \mathbb{R}^{n_k\times r_{k-1} r_k}$ $(\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_d)$	$\sum_{k=1}^{d} \langle \xi_k, \eta_k (\mathbf{W}_{\neq k}^\top \mathbf{W}_{\neq k} + \delta \mathbf{I}_{r_{k-1} r_k}) \rangle$
CCA [Yger et al.'12; Shustin-Aeron'23] **RCG**	$\mathrm{St}_{\Sigma_{xx}}(m, d_x) \times \mathrm{St}_{\Sigma_{yy}}(m, d_y)$ $(\mathbf{U}, \mathbf{V})$	$\langle \xi_1, \Sigma_{xx} \eta_1 \rangle + \langle \xi_2, \Sigma_{yy} \eta_2 \rangle$

MC=matrix completion; TC=tensor completion

Works interpreted by preconditioned metrics

Problem and methods	Search space $\mathcal{M}$ and variable	Metric $g_x(\xi, \eta)$, $\xi, \eta \in \mathrm{T}_x \mathcal{M}$
MC [Mishra-Apuroop-Sepulchre'12] **RGD, RCG, RTR**	$\mathbb{R}_*^{m\times r} \times \mathbb{R}_*^{n\times r}$ $(\mathbf{L}, \mathbf{R})$	$\langle \xi_1, \eta_1(\mathbf{R}^\top\mathbf{R})\rangle + \langle \xi_2, \eta_2(\mathbf{L}^\top\mathbf{L})\rangle$
Matrix sensing [Tong-Ma-Chi'21] **ScaledGD**	$\mathbb{R}_*^{m\times r} \times \mathbb{R}_*^{n\times r}$ $(\mathbf{L}, \mathbf{R})$	$\langle \xi_1, \eta_1(\mathbf{R}^\top\mathbf{R})\rangle + \langle \xi_2, \eta_2(\mathbf{L}^\top\mathbf{L})\rangle$
Tucker TC [Kasai-Mishra'16] **RCG**	$\times_{k=1}^{3} \mathrm{St}(r_k, n_k) \times \mathbb{R}^{r_1\times r_2\times r_3}$ $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathcal{G})$	$\sum_{k=1}^{3} \langle \xi_k, \eta_k(\mathbf{G}_{(k)}\mathbf{G}_{(k)}^\top)\rangle + \langle \xi_\mathcal{G}, \eta_\mathcal{G}\rangle$
CP TC [Dong-**G.**-Guan-Glineur'22] **RGD, RCG**	$\times_{k=1}^{d} \mathbb{R}^{n_k\times r}$ $(\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_d)$	$\sum_{k=1}^{d} \langle \xi_k, \eta_k((\mathbf{U}^{\odot j\neq k})^\top \mathbf{U}^{\odot j\neq k} + \delta \mathbf{I}_r)\rangle$
TT TC [Cai-Huang-Wang-Wei'22] **RGD, RCG, RGN**	$\times_{k=1}^{d} \mathbb{R}_*^{r_{k-1}\times n_k\times r_k}$ $(\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_d)$	$\sum_{k=1}^{d} \langle \xi_k, \eta_k(\mathbf{H}_k^\top\mathbf{H}_k)\rangle$
TR TC [**G.**-Peng-Yuan'24] **RGD, RCG**	$\times_{k=1}^{d} \mathbb{R}^{n_k\times r_{k-1}r_k}$ $(\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_d)$	$\sum_{k=1}^{d} \langle \xi_k, \eta_k(\mathbf{W}_{\neq k}^\top\mathbf{W}_{\neq k} + \delta \mathbf{I}_{r_{k-1}r_k})\rangle$
CCA [Yger et al.'12; Shustin-Aeron'23] **RCG**	$\mathrm{St}_{\Sigma_{xx}}(m, d_x) \times \mathrm{St}_{\Sigma_{yy}}(m, d_y)$ $(\mathbf{U}, \mathbf{V})$	$\langle \xi_1, \Sigma_{xx}\eta_1\rangle + \langle \xi_2, \Sigma_{yy}\eta_2\rangle$
CCA (this work) **RGD, RCG**	$\mathrm{St}_{\Sigma_{xx}}(m, d_x) \times \mathrm{St}_{\Sigma_{yy}}(m, d_y)$ $(\mathbf{U}, \mathbf{V})$	$\langle \xi_1, \Sigma_{xx}\eta_1\mathbf{M}_{1,2}\rangle + \langle \xi_2, \Sigma_{yy}\eta_2\mathbf{M}_{2,2}\rangle$
SVD (this work) **RGD, RCG**	$\mathrm{St}(p, m) \times \mathrm{St}(p, n)$ $(\mathbf{U}, \mathbf{V})$	$\langle \xi_1, \eta_1(\mathrm{sym}(\mathbf{U}^\top\mathbf{A}\mathbf{V}\mathbf{N})^2 + \delta\mathbf{I}_p)^{1/2}\rangle$ $+\langle \xi_2, \eta_2(\mathrm{sym}(\mathbf{V}^\top\mathbf{A}^\top\mathbf{U}\mathbf{N})^2 + \delta\mathbf{I}_p)^{1/2}\rangle$

MC=matrix completion; TC=tensor completion

Example: canonical correlation analysis

Problem formulation

Given two data matrices $\mathbf{X} \in \mathbb{R}^{n \times d_x}$ and $\mathbf{Y} \in \mathbb{R}^{n \times d_y}$

Optimization on the product of two generalized Stiefel manifolds

$$\min_{\mathbf{U} \in \mathbb{R}^{d_x \times m}, \mathbf{V} \in \mathbb{R}^{d_y \times m}} \quad f(\mathbf{U}, \mathbf{V}) := -\mathrm{tr}(\mathbf{U}^\top \Sigma_{xy} \mathbf{V} \mathbf{N})$$
$$\text{s. t.} \quad (\mathbf{U}, \mathbf{V}) \in \mathrm{St}_{\Sigma_{xx}}(m, d_x) \times \mathrm{St}_{\Sigma_{yy}}(m, d_y)$$

$\Sigma_{xy} := \mathbf{X}^\top \mathbf{Y}$, $\mathbf{N} := \mathrm{diag}(\mu_1, \mu_2, \ldots, \mu_m)$: $\mu_1 > \mu_2 > \cdots > \mu_m$

- Generalized Stiefel manifold:

$$\mathrm{St}_{\Sigma_{xx}}(m, d_x) := \{\mathbf{U} \in \mathbb{R}^{d_x \times m} : \mathbf{U}^\top \Sigma_{xx} \mathbf{U} = \mathbf{I}_m\}$$
$$\mathrm{St}_{\Sigma_{yy}}(m, d_y) := \{\mathbf{V} \in \mathbb{R}^{d_y \times m} : \mathbf{V}^\top \Sigma_{yy} \mathbf{V} = \mathbf{I}_m\}$$

$\Sigma_{xx} := \mathbf{X}^\top \mathbf{X} + \lambda_x \mathbf{I}_{d_x}$, $\Sigma_{yy} := \mathbf{Y}^\top \mathbf{Y} + \lambda_y \mathbf{I}_{d_y}$

Left preconditioning [Yger-Berar-Gasso-Rakotomamonjy'12; Shustin-Aeron'23]

- Metric: $g_{(\mathbf{U},\mathbf{V})}(\xi,\eta) := \langle \xi_1, \bar{\mathcal{H}}_1(\mathbf{U},\mathbf{V})[\eta_1]\rangle + \langle \xi_2, \bar{\mathcal{H}}_2(\mathbf{U},\mathbf{V})[\eta_2]\rangle$

$$\bar{\mathcal{H}}_1(\mathbf{U},\mathbf{V})[\eta_1] := \Sigma_{xx}\eta_1$$
$$\bar{\mathcal{H}}_2(\mathbf{U},\mathbf{V})[\eta_2] := \Sigma_{yy}\eta_2$$

$\rightsquigarrow$ Can we develop a better metric by our framework?

- Riemannian Hessian: (diagonal blocks)

$$\begin{aligned}\operatorname{Hess}_g f(\mathbf{U},\mathbf{V})[\eta] = \Pi_{(\mathbf{U},\mathbf{V})}\Big(&\eta_1 \operatorname{sym}(\mathbf{U}^\top\Sigma_{xy}\mathbf{V}\mathbf{N}) + \mathbf{U}\operatorname{sym}(\eta_1^\top\Sigma_{xy}\mathbf{V}\mathbf{N}) \\ &+ \mathbf{U}\operatorname{sym}(\mathbf{U}^\top\Sigma_{xy}\eta_2\mathbf{N}) - \Sigma_{xx}^{-1}\Sigma_{xy}\eta_2\mathbf{N}, \\ &\eta_2 \operatorname{sym}(\mathbf{V}^\top\Sigma_{xy}^\top\mathbf{U}\mathbf{N}) + \mathbf{V}\operatorname{sym}(\eta_2^\top\Sigma_{xy}^\top\mathbf{U}\mathbf{N}) \\ &+ \mathbf{V}\operatorname{sym}(\mathbf{U}^\top\Sigma_{xy}^\top\eta_1\mathbf{N}) - \Sigma_{yy}^{-1}\Sigma_{xy}^\top\eta_1\mathbf{N}\Big)\end{aligned}$$

Left and right preconditioning

- Metric: $g_{(\mathbf{U},\mathbf{V})}(\xi,\eta) := \langle \xi_1, \bar{\mathcal{H}}_1(\mathbf{U},\mathbf{V})[\eta_1]\rangle + \langle \xi_2, \bar{\mathcal{H}}_2(\mathbf{U},\mathbf{V})[\eta_2]\rangle$

$$\bar{\mathcal{H}}_1(\mathbf{U},\mathbf{V})[\eta_1] := \Sigma_{xx}\eta_1 (\mathrm{sym}(\mathbf{U}^\top \Sigma_{xy}\mathbf{V}\mathbf{N})^2 + \delta \mathbf{I}_m)^{\frac{1}{2}}$$

$$\bar{\mathcal{H}}_2(\mathbf{U},\mathbf{V})[\eta_2] := \Sigma_{yy}\eta_2 (\mathrm{sym}(\mathbf{V}^\top \Sigma_{xy}^\top\mathbf{U}\mathbf{N})^2 + \delta \mathbf{I}_m)^{\frac{1}{2}}$$

Riemannian gradient

$\mathbf{M}_{1,2} := (\mathrm{sym}(\mathbf{U}^\top \Sigma_{xy}\mathbf{V}\mathbf{N})^2 + \delta \mathbf{I}_m)^{\frac{1}{2}}$; $\mathbf{M}_{2,2} := (\mathrm{sym}(\mathbf{V}^\top \Sigma_{xy}^\top\mathbf{U}\mathbf{N})^2 + \delta \mathbf{I}_m)^{\frac{1}{2}}$

$$\begin{aligned}\mathrm{grad}_{\mathrm{new}} f(\mathbf{U},\mathbf{V}) = \big(&\Sigma_{xx}^{-1}\partial_{\mathbf{U}} f(\mathbf{U},\mathbf{V})\mathbf{M}_{1,2}^{-1} - \Sigma_{xx}^{-1}\Sigma_{xx}\mathbf{U}\mathbf{S}_1\mathbf{M}_{1,2}^{-1},\\ &\Sigma_{yy}^{-1}\partial_{\mathbf{V}} f(\mathbf{U},\mathbf{V})\mathbf{M}_{2,2}^{-1} - \Sigma_{yy}^{-1}\Sigma_{yy}\mathbf{V}\mathbf{S}_2\mathbf{M}_{2,2}^{-1}\big)\end{aligned}$$

- $\mathbf{S}_1$ and $\mathbf{S}_2$: solutions of Lyapunov equations

$$\mathrm{sym}\left(\mathbf{M}_{1,2}\mathbf{S}_1\right) = \mathrm{sym}\left(\mathbf{M}_{1,2}\mathbf{U}^\top \partial_{\mathbf{U}} f(\mathbf{U},\mathbf{V})\right)$$

$$\mathrm{sym}\left(\mathbf{M}_{2,2}\mathbf{S}_2\right) = \mathrm{sym}\left(\mathbf{M}_{2,2}\mathbf{V}^\top \partial_{\mathbf{V}} f(\mathbf{U},\mathbf{V})\right)$$

Theoretical properties: improved condition number

Condition number of $\mathrm{Hess}_g f(\mathbf{U}^*, \mathbf{V}^*)$ under different metrics

$(\mathbf{U}^*)^\top \Sigma_{xy} \mathbf{V}^* = \Sigma^*$, $\Sigma^* = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$: *leading singular values of* $\Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2}$

- Existing metric: $g_{(\mathbf{U},\mathbf{V})}(\xi, \eta) := \langle \xi_1, \Sigma_{xx} \eta_1 \rangle + \langle \xi_2, \Sigma_{yy} \eta_2 \rangle$
  - $m = 1$ [Shustin-Aeron'23]: $\kappa_g(\mathrm{Hess}_g f(\mathbf{U}^*, \mathbf{V}^*)) = (\sigma_1 + \sigma_2)/(\sigma_1 - \sigma_2)$
  - $m > 1$:

$$\kappa_g(\mathrm{Hess}_g f(\mathbf{U}^*, \mathbf{V}^*)) = \frac{\max\{(\mu_1 + \mu_2)(\sigma_1 + \sigma_2)/2, \mu_1(\sigma_1 + \sigma_{m+1})\}}{\min\{\min_{i,j \in [m], i \neq j} (\mu_i - \mu_j)(\sigma_i - \sigma_j)/2, \mu_m(\sigma_m - \sigma_{m+1})\}}$$

- Proposed metric: $g_{\mathrm{new},(\mathbf{U},\mathbf{V})}(\xi, \eta) := \langle \xi_1, \Sigma_{xx} \eta_1 \mathbf{M}_{1,2} \rangle + \langle \xi_2, \Sigma_{yy} \eta_2 \mathbf{M}_{2,2} \rangle$

$$\kappa_{\mathrm{new}}(\mathrm{Hess}_{\mathrm{new}} f(\mathbf{U}^*, \mathbf{V}^*)) = \frac{\max\left\{ \max\limits_{i,j \in [m], i \neq j} \frac{(\mu_i + \mu_j)(\sigma_i + \sigma_j)}{\sqrt{\sigma_i^2 \mu_i^2 + \delta} + \sqrt{\sigma_j^2 \mu_j^2 + \delta}}, \frac{\mu_1(\sigma_1 + \sigma_{m+1})}{\sqrt{\sigma_1^2 + \delta}} \right\}}{\min\left\{ \min\limits_{i,j \in [m], i \neq j} \frac{(\mu_i - \mu_j)(\sigma_i - \sigma_j)}{\sqrt{\sigma_i^2 \mu_i^2 + \delta} + \sqrt{\sigma_j^2 \mu_j^2 + \delta}}, \frac{\mu_m(\sigma_m - \sigma_{m+1})}{\sqrt{\sigma_m^2 + \delta}} \right\}}.$$

Numerical validation

Comparison on different Riemannian metrics

$$g_{(\mathbf{U},\mathbf{V})}(\xi,\eta) := \langle \xi_1, \bar{\mathcal{H}}_1(\mathbf{U},\mathbf{V})[\eta_1]\rangle + \langle \xi_2, \bar{\mathcal{H}}_2(\mathbf{U},\mathbf{V})[\eta_2]\rangle$$

- RGD/RCG (E): Euclidean metric
- RGD/RCG (L1): *Left preconditioning on one block*

$$\bar{\mathcal{H}}_1(\mathbf{U},\mathbf{V})[\eta_1] := \Sigma_{xx}\eta_1, \ \bar{\mathcal{H}}_2(\mathbf{U},\mathbf{V})[\eta_2] := \eta_2$$

- RGD/RCG (L2): *Left preconditioning on one block*

$$\bar{\mathcal{H}}_1(\mathbf{U},\mathbf{V})[\eta_1] := \eta_1, \ \bar{\mathcal{H}}_2(\mathbf{U},\mathbf{V})[\eta_2] := \Sigma_{yy}\eta_2$$

- RGD/RCG (L12): *Left preconditioning* [Shustin-Aeron'23]

$$\bar{\mathcal{H}}_1(\mathbf{U},\mathbf{V})[\eta_1] := \Sigma_{xx}\eta_1, \ \bar{\mathcal{H}}_2(\mathbf{U},\mathbf{V})[\eta_2] := \Sigma_{yy}\eta_2$$

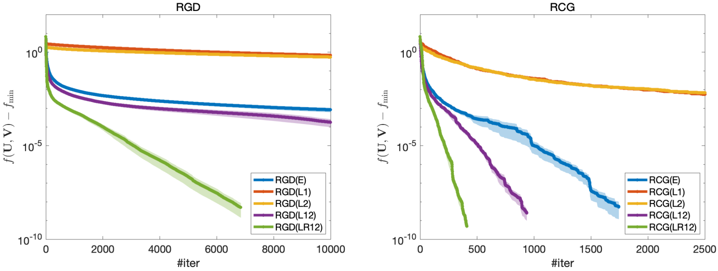
- RGD/RCG (LR12): *Left* and right *preconditioning*

$$\bar{\mathcal{H}}_1(\mathbf{U},\mathbf{V})[\eta_1] := \Sigma_{xx}\eta_1\mathbf{M}_{1,2}, \ \bar{\mathcal{H}}_2(\mathbf{U},\mathbf{V})[\eta_2] := \Sigma_{yy}\eta_2\mathbf{M}_{2,2}$$

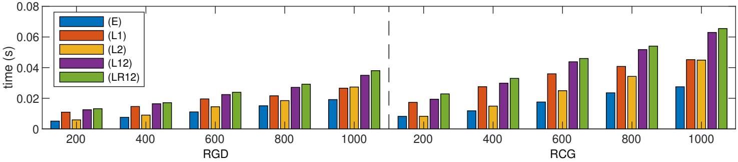
Comparison on different metrics

Errors of function value

- $d_x = 800$, $d_y = 400$, $n = 30000$, $m = 5$, and $\mathbf{N} := \mathrm{diag}(m, m-1, \ldots, 1)$
- $\mathbf{X}$ and $\mathbf{Y}$: unit distribution on $[0, 1]$



Average time per iteration



Condition numbers

metric	method	#iter	time (s)	gnorm	$D(\mathbf{U}, \mathbf{U}^*)$	$D(\mathbf{V}, \mathbf{V}^*)$	κ_g
(E)	RGD	10000	249.11	5.95e-02	2.69e-05	2.66e-05	2.10e+04
	RCG	1745	31.03	1.70e-05	4.01e-10	3.89e-10	
(L1)	RGD	10000	255.33	1.02e+00	4.12e-04	4.07e-04	1.43e+07
	RCG	2500	74.13	4.94e-02	2.85e-04	2.79e-04	
(L2)	RGD	10000	245.81	8.20e-01	4.13e-04	4.05e-04	1.52e+07
	RCG	2500	56.16	6.90e-02	2.93e-04	2.90e-04	
(L12)	RGD	10000	274.91	4.67e-04	9.68e-07	9.57e-07	1.12e+04
	RCG	937	30.39	8.82e-07	1.68e-09	1.65e-09	
(LR12)	RGD	6607	195.03	1.34e-06	7.47e-16	7.46e-16	2.38e+03
	RCG	410	15.38	8.49e-07	4.63e-09	4.53e-09	

$D(\mathbf{U}, \mathbf{U}^*) := \|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^*(\mathbf{U}^*)^\top\|_2$

Example: singular value decomposition

Singular value decomposition

Leading p singular values of a matrix

$$\begin{aligned} \min_{\mathbf{U}\in\mathbb{R}^{m\times p},\mathbf{V}\in\mathbb{R}^{n\times p}} \quad & -\mathrm{tr}(\mathbf{U}^\top \mathbf{A}\mathbf{V}\mathbf{N}) \\ \text{s. t.} \quad & (\mathbf{U},\mathbf{V}) \in \mathcal{M} := \mathrm{St}(p,m) \times \mathrm{St}(p,n) \end{aligned}$$

- $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m \geq n$ and $p < n$
- $\mathbf{N} := \mathrm{diag}\{\mu_1, \ldots, \mu_p\}$ with $\mu_1 > \cdots > \mu_p > 0$

Compared (Riemannian) methods

- RGD, RCG (E): Euclidean metric [Sato-Iwai'13]
- RGD, RCG (R12): **new** metric with right preconditioning effect

$$g_{\mathrm{new},(\mathbf{U},\mathbf{V})}(\xi,\eta) := \langle \xi_1, \eta_1 \mathbf{M}_{1,2}\rangle + \langle \xi_2, \eta_2 \mathbf{M}_{2,2}\rangle,$$

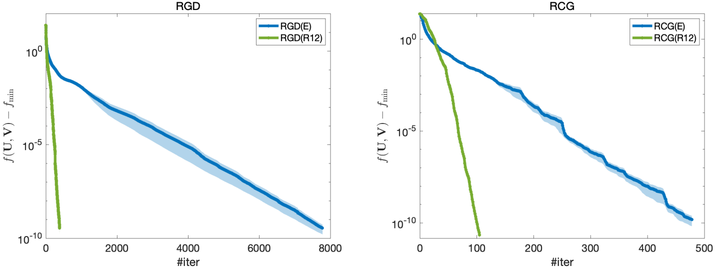
for $\xi, \eta \in \mathrm{T}_{(\mathbf{U},\mathbf{V})}\mathcal{M}$, where

$$\mathbf{M}_{1,2} = (\mathrm{sym}(\mathbf{U}^\top \mathbf{A}\mathbf{V}\mathbf{N})^2 + \delta \mathbf{I}_p)^{\frac{1}{2}}, \ \mathbf{M}_{2,2} = (\mathrm{sym}(\mathbf{V}^\top \mathbf{A}^\top \mathbf{U}\mathbf{N})^2 + \delta \mathbf{I}_p)^{\frac{1}{2}},$$

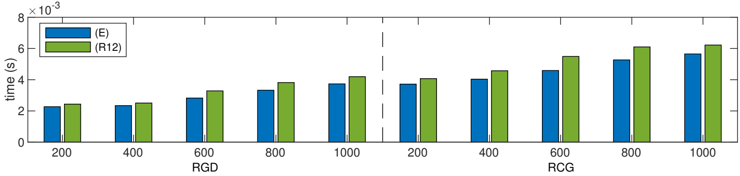
and $\delta > 0$.

Errors of function value

- $m = 1000$, $n = 500$, $p = 10$, $\mathbf{N} := \operatorname{diag}(p, p-1, \ldots, 1)$
- $\mathbf{A} = \mathbf{U}^* \Sigma (\mathbf{V}^*)^\top$: $\Sigma = \operatorname{diag}(1, \rho, \rho^2, \ldots, \rho^{p-1})$ with $\rho := 1/1.5$



Average time per iteration



Condition numbers

metric	method	#iter	time (s)	gnorm	$D(\mathbf{U}, \mathbf{U}^*)$	$D(\mathbf{V}, \mathbf{V}^*)$	κ_g
(E)	RGD	7781	117.29	9.64e-07	4.53e-05	4.53e-05	2.43e+03
	RCG	478	5.44	8.54e-07	2.00e-05	2.00e-05	
(R12)	RGD	387	3.41	8.72e-07	2.38e-15	1.38e-15	9.50e+01
	RCG	105	1.45	7.88e-07	3.26e-07	3.83e-07	

$D(\mathbf{U}, \mathbf{U}^*) := \|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^*(\mathbf{U}^*)^\top\|_2$

- Condition numbers:

$$\kappa(\mathrm{Hess}_{\mathrm{e}} f(\mathbf{U}^*, \mathbf{V}^*)) = \frac{(\mu_1 + \mu_2)(\gamma + 1)}{(\mu_{p-1} - \mu_p)(\gamma^{p-2} - \gamma^{p-1})} = \frac{153389}{63} \approx \mathbf{2.43 \times 10^3},$$

$$\kappa(\mathrm{Hess}_{\mathrm{new}} f(\mathbf{U}^*, \mathbf{V}^*)) = \frac{(\mu_1 + \mu_2)(1 + \gamma)}{(\mu_1 - \mu_2)(1 - \gamma)} = \mathbf{95}$$

$\rightsquigarrow$ “Numerical” coincides with “theoretical”!

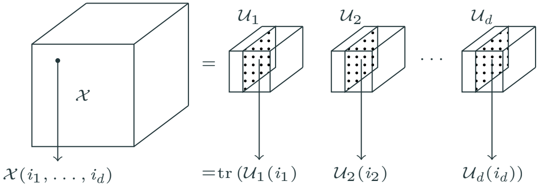
Example: tensor ring completion

TR-based model

$$\min_{(\mathcal{U}_1,\ldots,\mathcal{U}_d)\in\mathcal{M}_{\mathcal{U}}} \frac{1}{2}\left\|\mathrm{Proj}_{\Omega}\left([\![\mathcal{U}_1,\ldots,\mathcal{U}_d]\!]_{\mathrm{TR}}\right)-\mathrm{Proj}_{\Omega}(\mathcal{A})\right\|_{\mathrm{F}}^2$$

- $\mathcal{M}_{\mathcal{U}}$: search space via TR decomposition

$$\mathcal{M}_{\mathcal{U}} = \mathbb{R}^{r_1\times n_1\times r_2}\times\mathbb{R}^{r_2\times n_2\times r_3}\times\cdots\times\mathbb{R}^{r_d\times n_d\times r_1}$$



matricization

$(\mathcal{U}_1,\ldots,\mathcal{U}_d)\in\mathcal{M}_{\mathcal{U}}$ $\quad \xrightarrow{\mathbf{W}_k=(\mathcal{U}_k)_{(2)}}$ $\quad \mathcal{M}\ni(\mathbf{W}_1,\ldots,\mathbf{W}_d)$

$\mathbb{R}^{r_1\times n_1\times r_2}\times\cdots\times\mathbb{R}^{r_d\times n_d\times r_1}$ $\quad \xleftarrow[\mathcal{U}_k=\mathrm{ten}_{(2)}(\mathbf{W}_k)]{}$ $\quad \mathbb{R}^{n_1\times r_1 r_2}\times\cdots\times\mathbb{R}^{n_d\times r_d r_1}$

tensorization

Preconditioned metric via approximation of diagonal blocks

$$g_{\vec{\mathbf{W}}}(\vec{\boldsymbol{\xi}}, \vec{\boldsymbol{\eta}}) := \sum_{k=1}^{d} \operatorname{tr}\left(\boldsymbol{\xi}_k^\top \bar{\mathcal{H}}_k(\vec{\mathbf{W}})[\boldsymbol{\eta}_k]\right) \text{ for } \vec{\boldsymbol{\xi}}, \vec{\boldsymbol{\eta}} \in \mathrm{T}_{\vec{\mathbf{W}}}\,\mathcal{M},$$

where

$$\bar{\mathcal{H}}_k(\vec{\mathbf{W}})[\boldsymbol{\eta}_k] := \boldsymbol{\eta}_k\left(\mathbf{W}_{\neq k}^\top \mathbf{W}_{\neq k} + \underbrace{\delta \mathbf{I}_{r_{k+1} r_k}}_{\text{Shifting}}\right) \approx \partial_{kk}^2 f_\Omega(\vec{\mathbf{W}})$$

Riemannian gradient

$$\operatorname{grad} f(\vec{\mathbf{W}}) = \left(\bar{\mathcal{H}}_1^{-1}(\vec{\mathbf{W}})[\boldsymbol{\eta}_1], \ldots, \bar{\mathcal{H}}_d^{-1}(\vec{\mathbf{W}})[\boldsymbol{\eta}_d]\right)$$

Numerical experiments

Running Platform

- Personal computer with an Intel Core i9 CPU @ 2.4GHz×8 and 32GB of RAM
- Matlab R2020b under MacOS Ventura 13.1
- Code is publicly available from https://github.com/JimmyPeng1998

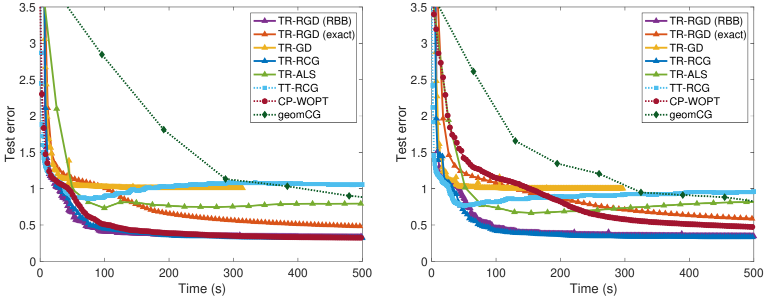
Test methods

- **TR-RGD:** Preconditioned Riemannian gradient descent *TR*
- **TR-RCG:** Preconditioned Riemannian CG *TR*
- **TR-GD:** Euclidean gradient descent *TR* [Zhao-Sugiyama-Yuan-Cichocki'19]
- **TR-ALS:** Alterating least squares *TR* [Wang-Aggarwal-Aeron'17]
- **TT-RCG:** Riemannian conjugate gradient *TT* [Steinlechner'15]
- **CP-WOPT:** limited memory BFGS *CP* [Acar-Dunlavy-Kolda-Mørup'11]
- **GeomCG:** Riemannian conjugate gradient *Tucker* [Kressner-Steinlechner-Vandereycken'14]
- **HaLRTC:** High accuracy low rank tensor completion *Nuclear-norm-based* [Liu-Musialski-Wonka-Ye'13]

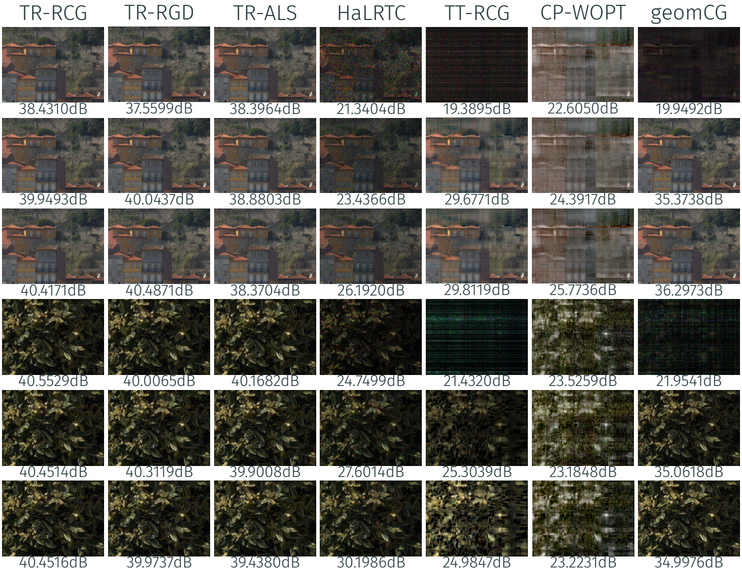
Movie ratings

MovieLens 1M dataset https://grouplens.org/datasets/movielens/1m/

- 6040 users, 3952 movies, 150 periods, 1M ratings
- $|\Omega| = 8 \times 10^5$, $|\Gamma| = 2 \times 10^5$, $\lambda = 1$, and $p = 2.23 \times 10^{-4}$
- Test errors. Left: $\mathbf{r} = (6, 6, 6)$; right: $\mathbf{r} = (6, 10, 3)$



Reconstruction of hyperspectral images



Function-related tensor completion

- $\mathcal{A}$: function-related tensor

$$\mathcal{A}(i_1, i_2, \ldots, i_d) = h\left(\frac{i_1 - 1}{n_1 - 1}, \frac{i_2 - 1}{n_2 - 1}, \ldots, \frac{i_d - 1}{n_d - 1}\right)$$

- $d = 4$, $n_1 = n_2 = n_3 = n_4 = 20$, $|\Gamma| = 100$

Test errors for high-dimensional functions

p	$\exp(-\|\mathbf{x}\|)$				$1/\|\mathbf{x}\|$			
	TR-RGD	TR-RCG	TR-ALS	TT-RCG	TR-RGD	TR-RCG	TR-ALS	TT-RCG
0.001	8.0884e-2	7.4157e-2	7.4161e-2	1.3445e-1	1.7531e-1	1.8106e-1	1.8081e-1	2.6876e-1
0.005	7.3505e-3	8.7366e-3	9.2121e-3	1.5904e-2	3.4428e-2	2.9218e-2	3.2090e-2	1.2899e-1
0.01	6.2650e-3	9.7247e-4	1.8737e-3	4.1233e-3	2.5230e-2	1.7676e-2	1.8697e-2	3.4675e-2
0.05	3.8862e-4	1.5019e-4	1.8218e-4	2.2991e-4	3.8510e-3	3.6002e-3	5.2173e-3	3.5697e-3
0.1	1.2251e-4	5.9871e-5	6.8898e-5	8.2512e-5	7.7886e-4	2.9423e-4	6.0423e-4	7.4727e-4

Take-home notes

- Preconditioned metric on product manifold
- Applications in
  - canonical correlation analysis
  - singular value decomposition
  - tensor ring completion [**G.**-Peng-Yuan'24]

References

- **Bin Gao**, Renfeng Peng, Ya-xiang Yuan. *Optimization on product manifolds under a preconditioned metric*. arxiv.2306.08873, (2023)
- **Bin Gao**, Renfeng Peng, Ya-xiang Yuan. *Riemannian preconditioned algorithms for tensor completion via tensor ring decomposition*. Computational Optimization and Applications, 88 (2024), 443–468
- 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

Thanks for your attention!

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