A space-decoupling framework for optimization on bounded-rank matrices with orthogonally invariant constraints

Bin Gao

Academy of Mathematics and Systems Science
Chinese Academy of Sciences

Joint work with
Yan Yang (AMSS, CAS)
Ya-xiang Yuan (AMSS, CAS)

Outline

Low-rank optimization

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]
- 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)

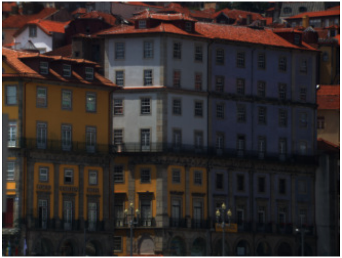
Low-rank matrix factorizations

- 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) $\quad k = 10$

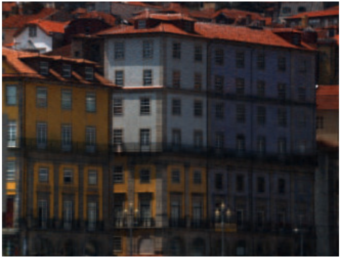
accuracy $\left(1 - \frac{\left\|U_k \Sigma_k V_k^\top - A\right\|_F}{\|A\|_F}\right)$:	$\mathbf{67.6\%}$
storage $\frac{\#\text{parameters } (U_k, \Sigma_k, V_k)}{\#\text{parameters } (A)}$:	1.1% only!

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*

Geometric methods for low-rank problems

I. optimization on manifolds

$$\begin{aligned} \min \quad & f(X) \\ \text{s.t.} \quad & X \in \mathbb{R}_r^{m\times n} \end{aligned}$$

Geometric methods for low-rank problems

Optimization on the set of fixed-rank matrices

$$\boxed{\begin{array}{ll} \displaystyle\min_{X \in \mathbb{R}^{m\times n}} & f(X) \\ \text{s. t.} & X \in \mathbb{R}_r^{m\times n} := \{X \in \mathbb{R}^{m\times n} : \operatorname{rank}(X) = r\} \end{array}}$$

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

Optimization on the set of fixed-rank matrices

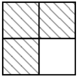
$$\boxed{\begin{array}{ll} \displaystyle\min_{X \in \mathbb{R}^{m \times n}} & f(X) \\ \text{s. t.} & X \in \mathbb{R}_r^{m \times n} := \{X \in \mathbb{R}^{m \times n} : \operatorname{rank}(X) = r\} \end{array}}$$

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

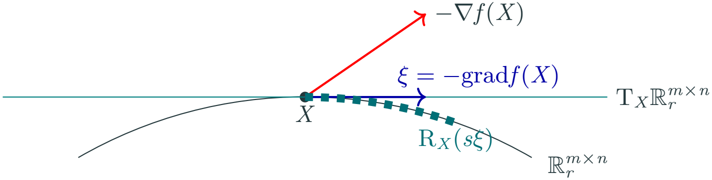
Tangent space [Vandereycken'13]

Given a thin SVD $X = U \Sigma V^\top$ with rank $\underline{r}$

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

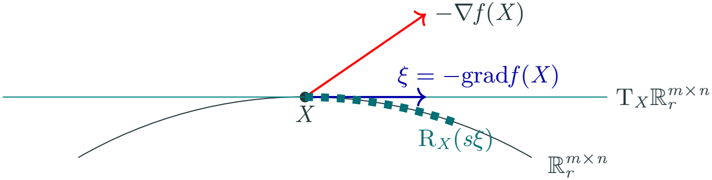
$$= \begin{bmatrix} U & U^\perp \end{bmatrix}$$

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

Optimization on manifold $\mathbb{R}_r^{m\times n}$



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

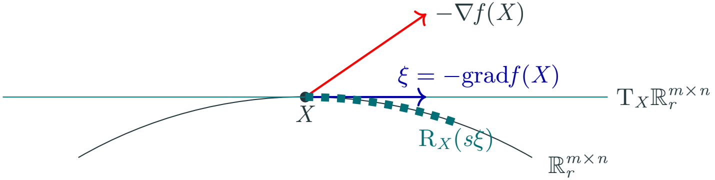
Optimization on manifold $\mathbb{R}_r^{m\times n}$



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

$\rightsquigarrow$ $\mathbb{R}_r^{m\times n}$ **is NOT closed!**

Optimization on manifold $\mathbb{R}_r^{m\times n}$



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

$\rightsquigarrow$ $\mathbb{R}_r^{m\times n}$ is NOT closed!

How to choose a rank parameter?

I. optimization on manifolds

$$\begin{aligned} \min \quad & f(X) \\ \text{s.t.} \quad & X \in \mathbb{R}_r^{m\times n} \end{aligned}$$

Geometric methods for low-rank problems

I. optimization on manifolds

$$\begin{aligned} \min \quad & f(X) \\ \text{s. t.} \quad & X \in \mathbb{R}_r^{m\times n} \end{aligned}$$

II. optimization on varieties

$$\begin{aligned} \min \quad & f(X) \\ \text{s. t.} \quad & X \in \mathbb{R}_{\le r}^{m\times n} \end{aligned}$$

Geometric methods for low-rank problems

$$\boxed{\begin{array}{ll} \min\limits_{X \in \mathbb{R}^{m \times n}} & f(X) \\ \text{s. t.} & X \in \mathbb{R}_{\leq r}^{m \times n} := \{X \in \mathbb{R}^{m \times n} : \operatorname{rank}(X) \leq r\} \end{array}}$$

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

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

$$\boxed{\begin{array}{ll} \min\limits_{X \in \mathbb{R}^{m\times n}} & f(X) \\ \text{s. t.} & X \in \mathbb{R}^{m\times n}_{\leq r} := \{X \in \mathbb{R}^{m\times n} : \operatorname{rank}(X) \leq r\} \end{array}}$$

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

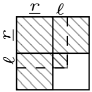
- closure of $\mathbb{R}^{m\times n}_{r}$
- *real-algebraic variety*
- more flexible choices of rank

Tangent cone [Schneider-Uschmajew'15]

Given a thin SVD $X = U\Sigma V^\top$ with rank $\underline{r}$

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

$$= \begin{bmatrix} U & U_1 & U_2 \end{bmatrix}$$



$$\begin{bmatrix} V & V_1 & V_2 \end{bmatrix}^\top$$

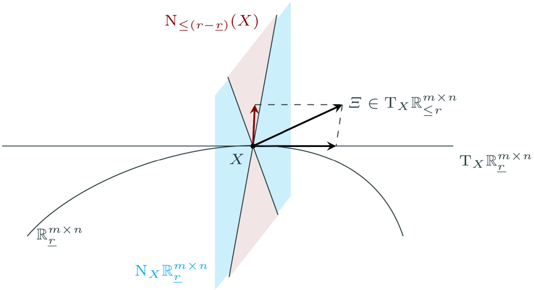
with $[U \;\; U_1 \;\; U_2] \in \mathcal{O}(m)$, $[V \;\; V_1 \;\; V_2] \in \mathcal{O}(n)$, and $\ell = 0, 1, \ldots, r - \underline{r}$

$$\boxed{\begin{array}{ll} \min\limits_{X \in \mathbb{R}^{m\times n}} & f(X) \\ \text{s. t.} & X \in \mathbb{R}_{\leq r}^{m\times n} := \{X \in \mathbb{R}^{m\times n} : \operatorname{rank}(X) \leq r\} \end{array}}$$

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

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

Tangent cone [Schneider-Uschmajew'15]



Low-rank problems: an overview of geometric methods

I. optimization on **manifolds**

$$\begin{aligned} \min \quad & f(X) \\ \mathrm{s.\,t.} \quad & X \in \mathbb{R}_r^{m \times n} \end{aligned}$$

II. optimization on **varieties**

$$\begin{aligned} \min \quad & f(X) \\ \mathrm{s.\,t.} \quad & X \in \mathbb{R}_{\le r}^{m \times n} \end{aligned}$$

Geometric methods for low-rank problems

Low-rank problems: an overview of geometric methods

I. optimization on manifolds

$$\begin{aligned} \min \quad & f(X) \\ \text{s.t.} \quad & X \in \mathbb{R}_r^{m\times n} \end{aligned}$$

II. optimization on varieties

$$\begin{aligned} \min \quad & f(X) \\ \text{s.t.} \quad & X \in \mathbb{R}_{\le r}^{m\times n} \end{aligned}$$

Geometric methods for low-rank problems

III. parametrization $(\mathcal{M}, \varphi)$

$$\begin{aligned} \min \quad & g(x) = f(\varphi(x)) \\ \text{s.t.} \quad & x \in \mathcal{M} \end{aligned}$$

$$\begin{tikzcd} \mathcal{M} \arrow[d, "\varphi"'] \arrow[dr, "f \circ \varphi"] & \\ \mathbb{R}_{\le r}^{m\times n} \arrow[r, "f"'] & \mathbb{R} \end{tikzcd}$$

Optimization via smooth parametrizations

$$\boxed{\begin{array}{ll} \min & g(x) = f(\varphi(x)) \\ \text{s. t.} & x \in \mathcal{M} \end{array}}$$

$$\begin{tikzcd} \mathcal{M} \arrow[d, "\varphi"'] \arrow[dr, "f \circ \varphi"] & \\ \mathbb{R}^{m \times n}_{\leq r} \arrow[r, "f"'] & \mathbb{R} \end{tikzcd}$$

- search space: $\mathcal{M}$
- “Lift” φ: $\varphi(\mathcal{M}) = \mathbb{R}^{m \times n}_{\leq r}$

Examples

- LR factorization:

$$S = \mathbb{R}^{m \times r} \times \mathbb{R}^{n \times r} \qquad \varphi(L, R) = LR^\top$$

  - Riemannian methods [Mishra et al.’12; Levin et al.’22]
  - Gauss–Southwell type methods [Olikier et al.’23]

- Burer–Monteiro factorization [Burer-Monteiro’03]

$$\mathcal{M} = \mathbb{R}^{m \times r} \qquad \varphi(R) = RR^\top$$

Low rank + additional constraints

Bounded-rank matrices

$$\begin{aligned} \min_{X \in \mathbb{R}^{m \times n}} \quad & f(X) \\ \mathrm{s.\,t.} \quad & \operatorname{rank}(X) \leq r \end{aligned}$$

Low rank + additional constraints

Bounded-rank matrices + orthogonally invariant constraints

$$\boxed{\begin{aligned} \min_{X \in \mathbb{R}^{m \times n}} \quad & f(X) \\ \text{s. t.} \quad & \operatorname{rank}(X) \leq r, \\ & h(X) = 0 \end{aligned}}$$

Bounded-rank matrices + orthogonally invariant constraints

$$\boxed{\begin{aligned} \min_{X \in \mathbb{R}^{m \times n}} \quad & f(X) \\ \text{s. t.} \quad & \operatorname{rank}(X) \leq r, \\ & h(X) = 0 \end{aligned}}$$

Blanket assumptions

- Objective: $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ is twice continuously differentiable
- Smooth and orthogonally invariant constraints:

$$h(XQ) = h(X) \quad \text{for} \quad Q \in \mathcal{O}(n)$$

- The differential $\mathrm{D}h$ has full rank in

$$\mathcal{H} := \{X \in \mathbb{R}^{m \times n} : \ h(X) = 0\}$$

Bounded-rank matrices + orthogonally invariant constraints

$$\boxed{\begin{aligned} \min_{X \in \mathbb{R}^{m\times n}} \quad & f(X) \\ \text{s. t.} \quad & \operatorname{rank}(X) \le r, \\ & h(X) = 0 \end{aligned}}$$

Blanket assumptions

- Objective: $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ is twice continuously differentiable
- Smooth and orthogonally invariant constraints:

$$h(XQ) = h(X) \quad \text{for} \quad Q \in \mathcal{O}(n)$$

- The differential $\mathrm{D}h$ has full rank in

$$\mathcal{H} := \left\{X \in \mathbb{R}^{m\times n} : \ h(X) = 0\right\}$$

$\rightsquigarrow$ $\mathcal{H}$ is a smooth manifold embedded in $\mathbb{R}^{m\times n}$

Feasible region

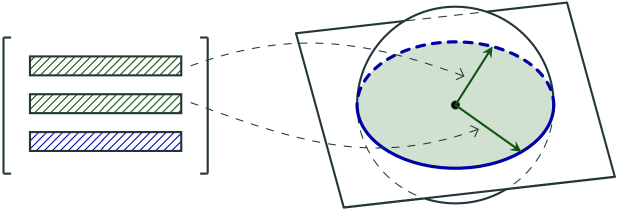
$$\mathbb{R}^{m\times n}_{\le r} \cap \mathcal{H} \quad \text{with} \quad \mathcal{H} = \{X : \; h(X) = 0\}$$

Low rank + oblique manifold

- Low-rank data fitting on sphere [Chu et al.'05 SIMAX]
- Model reduction of Markov processes
- Reinforcement learning (RL)
- Neural network training

$$\mathcal{H} = \mathrm{Ob}(m,n) := \{X \in \mathbb{R}^{m\times n} : \; \mathrm{diag}(XX^\top) - \mathbf{1} = 0\}$$

$\mathbb{R}^{3\times 4}_{\le 2} \cap \mathcal{H} \ni$



Applications (cont'd)

Low rank + Frobenius-sphere

- Low-rank approximation of graph similarity matrices [Cason et al.'13 LAA]

$$\mathcal{H} = \mathrm{S}_{\mathrm{F}}(m, n) := \{X \in \mathbb{R}^{m \times n} : \|X\|_{\mathrm{F}}^2 - 1 = 0\}$$

Low rank + stacked Stiefel manifold

- Low-rank solutions of synchronization problems [Boumal'15]

$$\mathcal{H} = \left\{X = (X_1; X_2; \ldots; X_k) \in \mathbb{R}^{kp \times n} : X_j^\top \in \mathrm{St}(n, p) \text{ for } j = 1, 2, \ldots, k\right\}$$

$$\begin{bmatrix} \boxed{X_1 X_1^\top = I} \\ \boxed{X_2 X_2^\top = I} \\ \boxed{X_3 X_3^\top = I} \end{bmatrix} \in \mathcal{H}$$

Low rank + SDPs

- Low-rank semidefinite programs [[Journée et al.'10; Wang et al.'23; Tang-Toh'23]

$$\min \ \langle C, Y \rangle, \qquad \mathrm{s.t.} \quad Y \in \mathbb{R}^{n \times n}_{\leq r} \cap \mathbb{S}^n_+, \ \mathcal{A}(Y) = b$$

$$\min \ \left\langle C, XX^\top \right\rangle, \quad \mathrm{s.t.} \quad X \in \mathbb{R}^{n \times n}_{\leq r}, \ \mathcal{A}(XX^\top) = b$$

Challenges

Constraint-coupled optimization

$$\boxed{\begin{array}{ll} \min\limits_{X \in \mathbb{R}^{m \times n}} & f(X) \\ \text{s. t.} & X \in \mathbb{R}^{m \times n}_{\leq r} \cap \mathcal{H} \end{array}}$$

Challenges

1. Unknown optimality conditions
   Bouligand-tangent cone of the coupled feasible region
2. Unclear strategy to preserve feasibility on $\mathbb{R}^{m \times n}_{\leq r} \cap \mathcal{H}$
3. Inherently non-smooth structure of $\mathbb{R}^{m \times n}_{\leq r}$

Constraint-coupled optimization

$$\boxed{\begin{array}{ll} \min\limits_{X \in \mathbb{R}^{m \times n}} & f(X) \\ \mathrm{s.\,t.} & X \in \mathbb{R}^{m \times n}_{\leq r} \cap \mathcal{H} \end{array}}$$

Challenges

1. Unknown optimality conditions
   Bouligand-tangent cone of the coupled feasible region
2. Unclear strategy to preserve feasibility on $\mathbb{R}^{m \times n}_{\leq r} \cap \mathcal{H}$
3. Inherently non-smooth structure of $\mathbb{R}^{m \times n}_{\leq r}$

$\rightsquigarrow$ How to tackle the coupled and non-smooth constraints?

Variational geometry of $\mathbb{R}^{m\times n}_{\leq r}$

- Mordukhovich normal cone [Luke'13]
- Bouligand tangent cone and Fréchet normal cone [Schneider-Uschmajew'15]
- Clarke tangent cone and the corresponding normal cone [Hosseini et al.'19; Li et al.'19]

Bounded-rank optimization: optimality analysis

Variational geometry of $\mathbb{R}^{m\times n}_{\le r}$

- Mordukhovich normal cone [Luke'13]
- Bouligand tangent cone and Fréchet normal cone [Schneider-Uschmajew'15]
- Clarke tangent cone and the corresponding normal cone [Hosseini et al.'19; Li et al.'19]

Variational geometry of $\mathbb{R}^{m\times n}_{\le r} \cap \mathcal{C}$

- Mordukhovich normal cone: $\mathcal{C} = \mathrm{Sym}(n)$ [Tam'17]
- Fréchet normal cone: $\mathcal{C} = \mathrm{Sym}(n) \cap \mathcal{B}_i$ [Li et al.'20]

  $\mathcal{B}_1 = \{X : \|X\|_{\mathrm{F}} \le 1\}$, $\mathcal{B}_2 = \{X : -tI_n \preceq X \preceq tI_n\}$, $\mathcal{B}_3 = \{X : X \succeq 0,\ \mathrm{Tr}(X) = 1\}$
- $\mathcal{C}$ is a spectral set [[Lewis'96; Pan'17; Li-Luo'23]

$m \ne n$ breaks the symmetry!

- Bouligand tangent cone: $\mathcal{C} = \{X \in \mathbb{R}^{m\times n} : \|X\|_{\mathrm{F}} = 1\}$ [Cason et al.'13]
- Fréchet normal cone: $\mathcal{C} = \{X \in \mathbb{R}^{m\times n} : \mathcal{A}(X) = b\}$ [Li-Luo'23]

Bounded-rank optimization: optimality analysis

Variational geometry of $\mathbb{R}^{m\times n}_{\leq r}$

- Mordukhovich normal cone [Luke'13]
- Bouligand tangent cone and Fréchet normal cone [Schneider-Uschmajew'15]
- Clarke tangent cone and the corresponding normal cone [Hosseini et al.'19; Li et al.'19]

Variational geometry of $\mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{C}$

- Mordukhovich normal cone: $\mathcal{C} = \mathrm{Sym}(n)$ [Tam'17]
- Fréchet normal cone: $\mathcal{C} = \mathrm{Sym}(n) \cap \mathcal{B}_i$ [Li et al.'20]

  $\mathcal{B}_1 = \{X : \|X\|_{\mathrm{F}} \leq 1\}$, $\mathcal{B}_2 = \{X : -tI_n \preceq X \preceq tI_n\}$, $\mathcal{B}_3 = \{X : X \succeq 0,\ \mathrm{Tr}(X) = 1\}$
- $\mathcal{C}$ is a spectral set [[Lewis'96; Pan'17; Li-Luo'23]

$m \neq n$ breaks the symmetry!

- Bouligand tangent cone: $\mathcal{C} = \{X \in \mathbb{R}^{m\times n} : \|X\|_{\mathrm{F}} = 1\}$ [Cason et al.'13]
- Fréchet normal cone: $\mathcal{C} = \{X \in \mathbb{R}^{m\times n} : \mathcal{A}(X) = b\}$ [Li-Luo'23]

$\rightsquigarrow$ **Not available in $\mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H}$!**

Bounded-rank optimization: algorithms

- Projected gradient descent framework [Jain et al.'14; Schneider-Uschmajew'15; Olikier et al.'22; 2024]
- Retraction-free methods [Schneider-Uschmajew'15; Olikier et al.'23; 2024]
  line-search along restricted tangent cones
- Optimizing over a smooth parameterization

$$\begin{tikzcd} \mathcal{M} \arrow[d, "\phi"'] \arrow[dr, "g := f \circ \phi"] & \\ \mathbb{R}^{m\times n}_{\le r} \arrow[r, "f"'] & \mathbb{R} \end{tikzcd}$$

$$\min g(Y) \;\; \text{s.t.} \;\; Y \in \mathcal{M}$$

- LR factorization [Mishra et al. 2012; Levin et al. 2022]

$$\mathcal{M} = \mathbb{R}^{m\times r} \times \mathbb{R}^{n\times r}, \; \phi(L, R) = LR^\top$$

- SVD-type lift [Mishra et al.'14; Levin et al.'23]

$$\mathcal{M} = \mathrm{St}(m, r) \times \mathrm{Sym}(r) \times \mathrm{St}(n, r), \; \phi(U, S, V) = USV^\top$$

- Desingularization [Khrulkov-Oseledets'18; Rebjock-Boumal'24]

$$\mathcal{M} = \left\{(X, G) \in \mathbb{R}^{m\times n} \times \mathrm{Gr}(n, n-r) : XG = 0\right\}, \; \phi(X, G) = X$$

Bounded-rank optimization: algorithms

- Projected gradient descent framework [Jain et al.'14; Schneider-Uschmajew'15; Olikier et al.'22; 2024]
- Retraction-free methods [Schneider-Uschmajew'15; Olikier et al.'23; 2024]
  line-search along restricted tangent cones
- Optimizing over a smooth parameterization

$$\begin{tikzcd} \mathcal{M} \arrow[d, "\phi"'] \arrow[dr, "g := f \circ \phi"] & \\ \mathbb{R}^{m\times n}_{\leq r} \arrow[r, "f"'] & \mathbb{R} \end{tikzcd}$$

$$\min g(Y) \;\; \text{s.t.} \;\; Y \in \mathcal{M}$$

- LR factorization [Mishra et al. 2012; Levin et al. 2022]

$$\mathcal{M} = \mathbb{R}^{m\times r} \times \mathbb{R}^{n\times r}, \; \phi(L, R) = LR^\top$$

- SVD-type lift [Mishra et al.'14; Levin et al.'23]

$$\mathcal{M} = \mathrm{St}(m, r) \times \mathrm{Sym}(r) \times \mathrm{St}(n, r), \; \phi(U, S, V) = USV^\top$$

- Desingularization [Khrulkov-Oseledets'18; Rebjock-Boumal'24]

$$\mathcal{M} = \left\{(X, G) \in \mathbb{R}^{m\times n} \times \mathrm{Gr}(n, n-r) : XG = 0\right\}, \; \phi(X, G) = X$$

$\rightsquigarrow$ **Not applicable in coupled constraints $\mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H}$!**

A space-decoupling framework?

$$\boxed{\begin{array}{ll} \min\limits_{X \in \mathbb{R}^{m \times n}} & f(X) \\ \mathrm{s.\,t.} & X \in \mathbb{R}^{m \times n}_{\leq r} \cap \mathcal{H} \end{array}}$$

A space-decoupling framework?

$$\boxed{\begin{array}{ll} \min\limits_{X \in \mathbb{R}^{m\times n}} & f(X) \\ \text{s. t.} & X \in \mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H} \end{array}}$$

Space-decoupling for coupled constraints

$$\mathcal{M}_h \xhookrightarrow{\quad \phi(x) := X \quad} \mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H}$$

A space-decoupling framework?

$$\boxed{\begin{array}{ll} \min\limits_{X\in\mathbb{R}^{m\times n}} & f(X) \\ \text{s. t.} & X \in \mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H} \end{array}}$$

Space-decoupling for coupled constraints

$$\mathcal{M}_h \xhookrightarrow{\phi(x) := X} \mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H}$$

rank information?

A space-decoupling framework?

$$\boxed{\begin{array}{ll} \min\limits_{X\in\mathbb{R}^{m\times n}} & f(X) \\ \text{s. t.} & X \in \mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H} \end{array}}$$

Space-decoupling for coupled constraints

$$\begin{tikzcd}
 & \mathcal{M}_h \arrow[dl] \arrow[d] \arrow[r, hook, "\phi(x) := X"] & \mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H} \\
\boxed{\text{rank information?}} & \boxed{h(X) = 0?} &
\end{tikzcd}$$

$$\boxed{\begin{array}{ll} \min\limits_{X \in \mathbb{R}^{m\times n}} & f(X) \\ \text{s. t.} & X \in \mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H} \end{array}}$$

Space-decoupling for coupled constraints

$$\begin{tikzcd} \mathcal{M}_h \arrow[rr, hook, "\phi(x) := X"] \arrow[d] \arrow[dr] & & \mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H} \\ \boxed{\text{rank information?}} & \boxed{h(X) = 0?} & \end{tikzcd}$$

Goal:

- a smooth parametrization
- optimality analysis
- convergence analysis
- equivalence of two problems

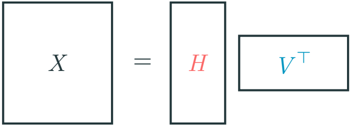
Orthogonally invariant functions and optimality

Orthogonally invariant "meets" low rank

Low rank decomposition

$$\mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H} = \bigcup_{s=0}^{s=r} (\mathbb{R}^{m\times n}_{s} \cap \mathcal{H})$$

- $\mathcal{H} = \{X \in \mathbb{R}^{m\times n} : \ h(X) = 0\}$ with orthogonally invariant h
- $X \in \mathbb{R}^{m\times n}_{s} \cap \mathcal{H}$ extracts low-rank (H) and orthogonally invariant (V)

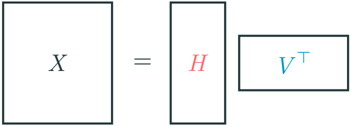


Orthogonally invariant “meets” low rank

Low rank decomposition

$$\mathbb{R}^{m\times n}_{\le r} \cap \mathcal{H} = \bigcup_{s=0}^{s=r} (\mathbb{R}^{m\times n}_{s} \cap \mathcal{H})$$

- $\mathcal{H} = \{X \in \mathbb{R}^{m\times n} : \ h(X) = 0\}$ with orthogonally invariant h
- $X \in \mathbb{R}^{m\times n}_{s} \cap \mathcal{H}$ extracts low-rank (H) and orthogonally invariant (V)



Submanifold inherites orthogonal invariance

$$X = HV^\top, \quad V^\top V = I$$

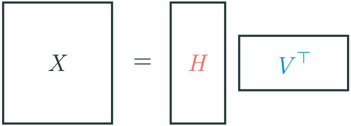
- Natural embeddings: $i^s(\cdot) : \ \mathbb{R}^{m\times s} \longrightarrow \mathbb{R}^{m\times n}, \ H \longmapsto [H \ 0^{m\times(n-s)}]$
- $h^s := h \circ i^s$ is orthogonally invariant
- $\mathcal{H}^s := \left\{H \in \mathbb{R}^{m\times s} : h^s(H) = 0\right\}$ is an embedded submanifold of $\mathbb{R}^{m\times s}$

Orthogonally invariant “meets” low rank

Low rank decomposition

$$\mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H} = \bigcup_{s=0}^{s=r} (\mathbb{R}^{m\times n}_{s} \cap \mathcal{H})$$

- $\mathcal{H} = \{X \in \mathbb{R}^{m\times n} : \; h(X) = 0\}$ with orthogonally invariant h
- $X \in \mathbb{R}^{m\times n}_{s} \cap \mathcal{H}$ extracts low-rank (H) and orthogonally invariant (V)



Submanifold inherites orthogonal invariance

$$X = HV^\top, \quad V^\top V = I$$

- Natural embeddings: $i^s(\cdot) : \; \mathbb{R}^{m\times s} \longrightarrow \mathbb{R}^{m\times n}, \; H \longmapsto [H \; 0^{m\times(n-s)}]$
- $h^s := h \circ i^s$ is orthogonally invariant
- $\mathcal{H}^s := \left\{H \in \mathbb{R}^{m\times s} : h^s(H) = 0\right\}$ is an embedded submanifold of $\mathbb{R}^{m\times s}$

$$\rightsquigarrow \quad \boxed{X \in \mathcal{H} \iff H \in \mathcal{H}^s}$$

Space decoupling - level 1: tangent space

$$\mathcal{H} = \{X \in \mathbb{R}^{m\times n} : h(X) = 0\} \quad \text{and} \quad h(XQ) = h(X) \text{ for } Q \in \mathcal{O}(n)$$

Definition

- *Bouligand tangent cone*:

$$\mathrm{T}_X\mathcal{X} := \left\{\eta = \lim_{k\to\infty} \frac{(X_k - X)}{t_k} : X_k \in \mathcal{X}, t_k > 0 \text{ for all } k, \lim_{k\to\infty} t_k = 0\right\}.$$

- *Fréchet normal cone*: $\mathrm{N}_X\mathcal{X} := (\mathrm{T}_X\mathcal{X})^\circ$

Tangent space decoupling at $X = HV^\top \in \mathcal{H}$

$$\begin{aligned}\mathrm{T}_X\mathcal{H} &= \left\{KV^\top + BV_\perp^\top : K \in \mathrm{T}_H\mathcal{H}^s, B \in \mathbb{R}^{m\times(n-s)}\right\} \\ &= (\mathrm{T}_H\mathcal{H}^s)\, V^\top \oplus \left(\mathbb{R}^{m\times(n-s)}\right) V_\perp^\top\end{aligned}$$

$$\mathcal{H} = \{X \in \mathbb{R}^{m\times n} : h(X) = 0\} \quad \text{and} \quad h(XQ) = h(X) \text{ for } Q \in \mathcal{O}(n)$$

Definition

- *Bouligand tangent cone*:

$$\mathrm{T}_X\mathcal{X} := \left\{\eta = \lim_{k\to\infty} \frac{(X_k - X)}{t_k} : X_k \in \mathcal{X}, t_k > 0 \text{ for all } k, \lim_{k\to\infty} t_k = 0\right\}.$$

- *Fréchet normal cone*: $\mathrm{N}_X\mathcal{X} := (\mathrm{T}_X\mathcal{X})^\circ$

Tangent space decoupling at $X = HV^\top \in \mathcal{H}$

$$\begin{aligned}\mathrm{T}_X\mathcal{H} &= \left\{KV^\top + BV_\perp^\top : K \in \mathrm{T}_H\mathcal{H}^s, B \in \mathbb{R}^{m\times(n-s)}\right\}\\ &= (\mathrm{T}_H\mathcal{H}^s)\, V^\top \oplus \left(\mathbb{R}^{m\times(n-s)}\right) V_\perp^\top\end{aligned}$$

$$\rightsquigarrow \quad \mathrm{T}_X\left(\mathbb{R}^{m\times n}_{\le r} \cap \mathcal{H}\right)?$$

Smooth layers

$$\mathbb{R}_{\leq r}^{m\times n} \cap \mathcal{H} = \bigcup_{s=0}^{s=r} (\mathbb{R}_s^{m\times n} \cap \mathcal{H}) : \quad X = HV^\top = U\Sigma V^\top$$

- $\mathbb{R}_s^{m\times n} \cap \mathcal{H}$ is a smooth manifold with

$$\begin{aligned} \mathrm{T}_X \left(\mathbb{R}_s^{m\times n} \cap \mathcal{H}\right) &= \left\{ KV^\top + UJV_\perp^\top : \ K \in \mathrm{T}_H \mathcal{H}^s, \ J \in \mathbb{R}^{s\times(n-s)} \right\} \\ &= \mathrm{T}_X \mathbb{R}_s^{m\times n} \cap \mathrm{T}_X \mathcal{H} \end{aligned}$$

Space decoupling - level 2: tangent cone

Smooth layers

$$\mathbb{R}^{m\times n}_{\le r} \cap \mathcal{H} = \bigcup_{s=0}^{s=r} (\mathbb{R}^{m\times n}_{s} \cap \mathcal{H}) : \quad X = HV^\top = U\Sigma V^\top$$

- $\mathbb{R}^{m\times n}_{s} \cap \mathcal{H}$ is a smooth manifold with

$$\begin{aligned} \mathrm{T}_X \left(\mathbb{R}^{m\times n}_{s} \cap \mathcal{H}\right) &= \left\{ KV^\top + UJV_\perp^\top : \; K \in \mathrm{T}_H \mathcal{H}^s, \; J \in \mathbb{R}^{s\times(n-s)} \right\} \\ &= \mathrm{T}_X \mathbb{R}^{m\times n}_{s} \cap \mathrm{T}_X \mathcal{H} \end{aligned}$$

Closed-form expression

$$\mathrm{T}_X \left(\mathbb{R}^{m\times n}_{\le r} \cap \mathcal{H}\right) = \left\{ \begin{array}{c} KV^\top + UJV_\perp^\top + U_\perp R V_\perp^\top : \; K \in \mathrm{T}_H \mathcal{H}^s \\ J \in \mathbb{R}^{s\times(n-s)}, \; R \in \mathbb{R}^{(m-s)\times(n-s)}, \; \operatorname{rank}(R) \le r - s \end{array} \right\}$$

Space decoupling - level 2: tangent cone

Smooth layers

$$\mathbb{R}^{m\times n}_{\leq r}\cap\mathcal{H}=\bigcup_{s=0}^{s=r}(\mathbb{R}^{m\times n}_{s}\cap\mathcal{H}):\quad X=HV^\top=U\Sigma V^\top$$

- $\mathbb{R}^{m\times n}_{s}\cap\mathcal{H}$ is a smooth manifold with

$$\begin{aligned}\mathrm{T}_X\left(\mathbb{R}^{m\times n}_{s}\cap\mathcal{H}\right)&=\left\{KV^\top+UJV_\perp^\top:\ K\in\mathrm{T}_H\mathcal{H}^s,\ J\in\mathbb{R}^{s\times(n-s)}\right\}\\&=\mathrm{T}_X\mathbb{R}^{m\times n}_{s}\cap\mathrm{T}_X\mathcal{H}\end{aligned}$$

Closed-form expression

$$\mathrm{T}_X\left(\mathbb{R}^{m\times n}_{\leq r}\cap\mathcal{H}\right)=\left\{\begin{array}{c}KV^\top+UJV_\perp^\top+U_\perp RV_\perp^\top:\ K\in\mathrm{T}_H\mathcal{H}^s\\ J\in\mathbb{R}^{s\times(n-s)},\ R\in\mathbb{R}^{(m-s)\times(n-s)},\ \mathrm{rank}(R)\leq r-s\end{array}\right\}$$

Intersection rules for tangent and normal cones

$$\begin{aligned}\mathrm{T}_X\left(\mathbb{R}^{m\times n}_{\leq r}\cap\mathcal{H}\right)&=\mathrm{T}_X\mathbb{R}^{m\times n}_{\leq r}\cap\mathrm{T}_X\mathcal{H}\\\mathrm{N}_X\left(\mathbb{R}^{m\times n}_{\leq r}\cap\mathcal{H}\right)&=\mathrm{N}_X\mathbb{R}^{m\times n}_{\leq r}\oplus\mathrm{N}_X\mathcal{H}\end{aligned}$$

- Recover the tangent cone in [Cason-Absil-Dooren'13, Theorem 6.1] when $\mathcal{H}=\{X\in\mathbb{R}^{m\times n}:\|X\|_\mathrm{F}=1\}$

A space-decoupling framework

Space decoupling - level 3: a smooth parametrization

$$\mathcal{M}_h \xhookrightarrow{\phi(x) := X} \mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H}$$

rank information?

$h(X) = 0?$

$$\begin{tikzcd}
& \mathcal{M}_h \arrow[dl] \arrow[d] \arrow[r, hook, "\phi(x) := X"] & \mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H} \\
\boxed{\text{rank information?}} & \boxed{h(X) = 0?} &
\end{tikzcd}$$

A space-decoupling parameterization $\mathcal{M}_h$

- Borrow ideas from desingularization [Khrulkov-Oseledets'18; Rebjock-Boumal'24]
- Parameterize the coupled and non-smooth $\mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H}$

$$\mathcal{M}_h := \left\{(X, G) \in \mathbb{R}^{m\times n} \times \mathrm{Gr}(n, n-r) : XG = 0,\ h(X) = 0\right\}$$

where $\mathrm{Gr}(n, n-r) = \{P \in \mathrm{Sym}(n) : P^2 = P,\ \mathrm{rank}(P) = n - r\}$

Space decoupling - level 3: a smooth parametrization

$$\begin{tikzcd} & \mathcal{M}_h \arrow[dl] \arrow[d] \arrow[r, hook, "\phi(X,\,G)\,:=\,X"] & \mathbb{R}^{m\times n}_{\le r} \cap \mathcal{H} \\ \boxed{G \in \mathrm{Gr}(n, n-r)} & \boxed{h(X) = 0?} & \end{tikzcd}$$

A space-decoupling parameterization $\mathcal{M}_h$

- Borrow ideas from desingularization [Khrulkov-Oseledets'18; Rebjock-Boumal'24]
- Parameterize the coupled and non-smooth $\mathbb{R}^{m\times n}_{\le r} \cap \mathcal{H}$

$$\mathcal{M}_h := \left\{ (X, G) \in \mathbb{R}^{m\times n} \times \mathrm{Gr}(n, n-r) : XG = 0,\ h(X) = 0 \right\}$$

where $\mathrm{Gr}(n, n-r) = \{P \in \mathrm{Sym}(n) : P^2 = P,\ \mathrm{rank}(P) = n - r\}$

Space decoupling - level 3: a smooth parametrization

$$\begin{tikzcd} & \mathcal{M}_h \arrow[r, hook, "\phi(X,\,G) := X"] \arrow[dl] \arrow[d] & \mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H} \\ \boxed{G \in \mathrm{Gr}(n, n-r)} & \boxed{h(X) = 0} & \end{tikzcd}$$

A space-decoupling parameterization $\mathcal{M}_h$

- Borrow ideas from desingularization [Khrulkov-Oseledets'18; Rebjock-Boumal'24]
- Parameterize the coupled and non-smooth $\mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H}$

$$\mathcal{M}_h := \left\{(X, G) \in \mathbb{R}^{m\times n} \times \mathrm{Gr}(n, n-r) : XG = 0,\ h(X) = 0\right\}$$

where $\mathrm{Gr}(n, n-r) = \{P \in \mathrm{Sym}(n) : P^2 = P,\ \mathrm{rank}(P) = n - r\}$

$$\begin{tikzcd} & \mathcal{M}_h \arrow[r, hook, "\phi(X,G) := X"] \arrow[dl] \arrow[d] & \mathbb{R}^{m\times n}_{\le r}\cap\mathcal{H} \\ \boxed{G \in \mathrm{Gr}(n, n-r)} & \boxed{h(X) = 0} & \end{tikzcd}$$

A space-decoupling parameterization $\mathcal{M}_h$

- Borrow ideas from desingularization [Khrulkov-Oseledets'18; Rebjock-Boumal'24]
- Parameterize the coupled and non-smooth $\mathbb{R}^{m\times n}_{\le r}\cap\mathcal{H}$

$$\mathcal{M}_h := \left\{(X, G) \in \mathbb{R}^{m\times n} \times \mathrm{Gr}(n, n-r) : XG = 0,\ h(X) = 0\right\}$$

where $\mathrm{Gr}(n, n-r) = \{P \in \mathrm{Sym}(n) : P^2 = P, \mathrm{rank}(P) = n - r\}$

$\rightsquigarrow$ $\mathcal{M}_h$ is a smooth manifold embedded in $\mathbb{R}^{m\times n} \times \mathrm{Sym}(n)$!

$$
\begin{tikzcd}
\mathcal{M}_h \arrow[d, "\phi"'] \arrow[dr, "\bar{f} := f \circ \phi"] & \\
\mathbb{R}^{m \times n}_{\leq r} \cap \mathcal{H} \arrow[r, "f"'] & \mathbb{R}
\end{tikzcd}
$$

$$\min_{(X,G) \in \mathcal{M}_h} \bar{f}\,(X, G) \qquad (\mathbb{M})$$

$$\min_{X \in \mathbb{R}^{m \times n}_{\leq r} \cap \mathcal{H}} f(X) \qquad (\mathbb{O})$$

- $\phi : \mathcal{M}_h \to \mathbb{R}^{m \times n}_{\leq r} \cap \mathcal{H} : (X, G) \mapsto X$ is a surjection
- $(\mathbb{M})$ is a smooth Riemannian optimization problem
  - Riemannian derivatives
  - retractions
  - vector transports

Space decoupling - level 4: optimization

Riemannian gradient $\nabla_{\mathcal{M}_h}\bar{f}(X,G)$

$$\bar{K} = \mathcal{P}_{\mathrm{T}_H\mathcal{H}^r}\left(\nabla f(X)\,V\right), \qquad \bar{V}_p = G\nabla f(X)^\top H M_{H,\omega}^{-1}$$

Riemannian Hessian $\nabla^2_{\mathcal{M}_h}\bar{f}(X,G)[\eta,\zeta]$

$$\begin{cases} \bar{K} = \nabla^2_{\mathcal{H}} f(X)[\eta]\,V + \left(I - HM_{H,\omega}^{-1}H^\top\right)\nabla f(X)\,V_p, \\ \bar{V}_p = G\left((\nabla^2 f(X)[\eta])^\top H + \nabla f(X)^\top (I - HM_{H,\omega}^{-1}H^\top)K\right)M_{H,\omega}^{-1} \end{cases}$$

First-order retraction

$$\mathrm{R}_{(X,G)}(\eta,\zeta) := (\mathrm{R}_H^{\mathcal{H}^r}(K)(\mathrm{R}_V^{\mathrm{St}}(V_p))^\top,\ I - \mathrm{R}_V^{\mathrm{St}}(V_p)(\mathrm{R}_V^{\mathrm{St}}(V_p))^\top)$$

Second-order retraction

$$\mathrm{R}_{(X,G)}(\eta,\zeta) := \left(\left[\mathcal{P}_{\mathcal{H}^r}\left((X+\eta)\,W\right)\right]W^\top, I - WW^\top\right)$$

$$W = L(L^\top L)^{-\frac{1}{2}},\quad L = V + V_p(I - K^\top HM_{H,\omega}^{-1})$$

Vector transport

$$\bar{K} = \mathcal{T}_{\tilde{K}}^{\mathcal{H}}(K) \text{ and } \bar{V}_p = (I - \tilde{V}\tilde{V}^\top)\mathcal{T}_{\tilde{V}_p}^{\mathrm{St}}(V_p)$$

Space decoupling - level 4: optimization

Riemannian gradient $\nabla_{\mathcal{M}_h}\bar{f}(X,G)$

$$\bar{K} = \mathcal{P}_{\mathrm{T}_H\mathcal{H}^r}\left(\nabla f(X)V\right), \qquad \bar{V}_p = G\nabla f(X)^\top HM_{H,\omega}^{-1}$$

Riemannian Hessian $\nabla^2_{\mathcal{M}_h}\bar{f}(X,G)[\eta,\zeta]$

$$\begin{cases} \bar{K} = \nabla^2_{\mathcal{H}} f(X)[\eta]\,V + \left(I - HM_{H,\omega}^{-1}H^\top\right)\nabla f(X)\,V_p, \\ \bar{V}_p = G\left((\nabla^2 f(X)[\eta])^\top H + \nabla f(X)^\top (I - HM_{H,\omega}^{-1}H^\top)K\right)M_{H,\omega}^{-1} \end{cases}$$

First-order retraction

$$\mathrm{R}_{(X,G)}(\eta,\zeta) := (\mathrm{R}_H^{\mathcal{H}^r}(K)(\mathrm{R}_V^{\mathrm{St}}(V_p))^\top,\ I - \mathrm{R}_V^{\mathrm{St}}(V_p)(\mathrm{R}_V^{\mathrm{St}}(V_p))^\top)$$

Second-order retraction

$$\mathrm{R}_{(X,G)}(\eta,\zeta) := ([\mathcal{P}_{\mathcal{H}^r}((X+\eta)W)]\,W^\top, I - WW^\top)$$

$$W = L(L^\top L)^{-\frac{1}{2}}, \quad L = V + V_p(I - K^\top HM_{H,\omega}^{-1})$$

Vector transport

$$\bar{K} = \mathcal{T}_{\tilde{K}}^{\mathcal{H}}(K) \text{ and } \bar{V}_p = (I - \tilde{V}\tilde{V}^\top)\mathcal{T}_{\tilde{V}_p}^{\mathrm{St}}(V_p)$$

Optimization on $\mathcal{M}_h$ is painless if geometry of $\mathcal{H}$ is developed!

$$
\begin{tikzcd}
\mathcal{M}_h \arrow[d, "\phi"'] \arrow[dr, "\bar{f} := f \circ \phi"] & \\
\mathbb{R}^{m \times n}_{\leq r} \cap \mathcal{H} \arrow[r, "f"'] & \mathbb{R}
\end{tikzcd}
$$

$$\min_{(X,G) \in \mathcal{M}_h} \bar{f}\,(X, G) \qquad (\mathbb{M})$$

$$\min_{X \in \mathbb{R}^{m \times n}_{\leq r} \cap \mathcal{H}} f(X) \qquad (\mathbb{O})$$

$$\begin{tikzcd}
\mathcal{M}_h \arrow[d, "\phi"'] \arrow[dr, "\bar{f} := f \circ \phi"] & \\
\mathbb{R}^{m \times n}_{\leq r} \cap \mathcal{H} \arrow[r, "f"'] & \mathbb{R}
\end{tikzcd}$$

$$\min_{(X,G) \in \mathcal{M}_h} \bar{f}\,(X, G) \qquad (\mathbb{M})$$

$$\min_{X \in \mathbb{R}^{m \times n}_{\leq r} \cap \mathcal{H}} f(X) \qquad (\mathbb{O})$$

$\rightsquigarrow$ Parameterization may introduce spurious stationarity!

Equivalence of problems

$$\begin{tikzcd}
\mathcal{M}_h \arrow[d, "\phi"'] \arrow[dr, "\bar{f} := f \circ \phi"] & \\
\mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H} \arrow[r, "f"'] & \mathbb{R}
\end{tikzcd}$$

$$\min_{(X,G)\in\mathcal{M}_h} \bar{f}\,(X, G) \qquad (\mathbb{M})$$

$$\min_{X\in\mathbb{R}^{m\times n}_{\leq r}\cap\mathcal{H}} f(X) \qquad (\mathbb{O})$$

$\rightsquigarrow$ Parameterization may introduce spurious stationarity!

Equivalence

The space-decoupling parametrization $\mathcal{M}_h$ of $\mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H}$ satisfies

1. "$1 \Rightarrow 1$" holds at (X, G) if and only if $\operatorname{rank}(X) = r$
2. "$2 \Rightarrow 1$" holds everywhere on $\mathcal{M}_h$

Convergence results

Important ingredients

- $\mathcal{M}_h$ is a complete manifold
- Compact sublevel sets of $f \implies$ compact sublevel sets of $\bar{f}$

General convergence theorem

- The sequence $\{(X_k, G_k)\}$ generated by monotone algorithms has at least one accumulation point
- If the accumulation point (X^*, G^*) is second-order stationary for ($\mathbb{M}$), then X^* is first-order stationary for ($\mathbb{O}$)

Specific algorithms

- RGD accumulates to first-order critical points
- RTR accumulates to second-order critical points

Numerical experiments

1. Low-rank approximation of spherical data

Approximate $A \in \mathcal{H} = \mathrm{Ob}(m,n)$ with low-rank matrix [Chu et al.'05]

$$\boxed{\begin{aligned} \min_{X \in \mathbb{R}^{m\times n}} \quad & \frac{1}{2}\left\|\mathcal{P}_\Omega(X - A)\right\|^2 \\ \text{s. t.} \quad & \operatorname{rank}(X) \le r \\ & \operatorname{diag}(XX^\top) - \mathbf{1} = 0 \end{aligned}} \xrightarrow{\mathcal{H}=\mathrm{Ob}(m,n)} \boxed{\min_{x \in \mathcal{M}_h} \quad \bar{f}(x) := \frac{1}{2}\left\|\mathcal{P}_\Omega(\phi(x) - A)\right\|_{\mathrm{F}}^2}$$

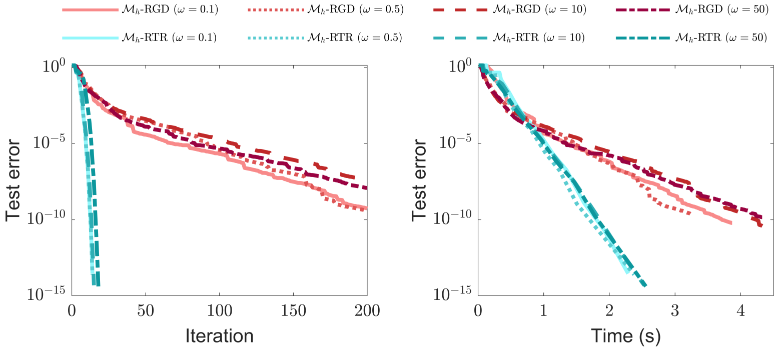
Methods *based on Manopt v8.0*

- Riemannian gradient descent method (RGD)
- Riemannian trust region method (RTR)

Running platform

- CPU: two Intel(R) Xeon(R) Processors Gold 6330 @ 2.00GHz×28
- GPU: one NVIDIA A800 (80GB memory) GPU

Test with the unbiased rank parameter $r = r^* = 6$



Test with over-estimated rank parameters

Algorithm	$r = 7$		$r = 8$		$r = 9$		$r = 10$	
	Test err.	Time	Test err.	Time	Test err.	Time	Test err.	Time
$\mathcal{M}_h$-RGD	4.88e−12	11.87	5.12e−13	6.55	1.11e−12	8.13	4.16e−12	16.98
$\mathcal{M}_h$-RTR	3.34e−15	14.93	5.27e−15	8.74	5.81e−15	7.19	2.14e−15	7.68

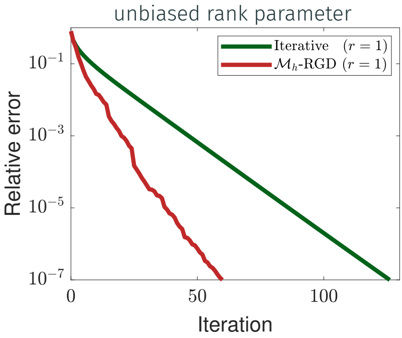
Measuring problem [Cason et al.'13]

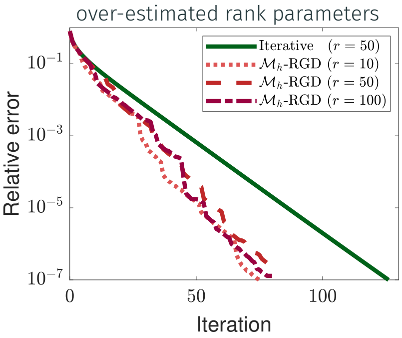
$X \in \mathcal{H} = \mathrm{S_F}(m, n)$ measures the node-to-node similarity between graphs

$$\boxed{\begin{array}{cl} \min\limits_{X \in \mathbb{R}^{m \times n}} & -\operatorname{tr}(X^\top \mathcal{L} \circ \mathcal{L}(X)) \\ \text{s. t.} & \operatorname{rank}(X) \leq r \\ & \|X\|_\mathrm{F}^2 - 1 = 0 \end{array}} \xrightarrow{\mathcal{H} = \mathrm{S_F}(m,n)} \boxed{\min\limits_{x \in \mathcal{M}_h} \; -\operatorname{tr}\left(\phi(x)^\top \mathcal{L} \circ \mathcal{L}(\phi(x))\right)}$$

Problem setting

- G_A of $m = 2000$ vertices: a single cycle
- G_B of $n = 3000$ vertices: binomial random graph
- Solution of low rank: $r^* = \mathrm{rank}(X^*) = 1$



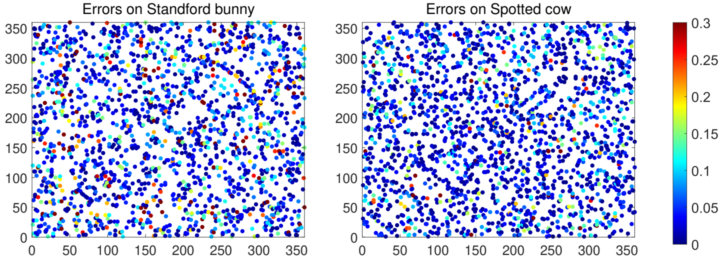


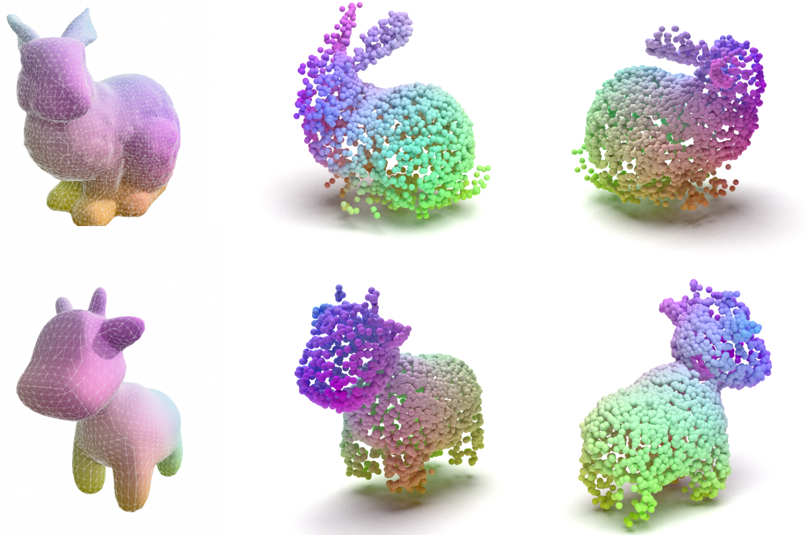
Synchronization problem

- n cameras with a set of relative rotations $\{\hat{R}_{ij} : (i,j) \in E\}$
- reconstruct the absolute rotations $\{R_i\}_{i=1}^n$ such that $\hat{R}_{ij} \approx R_j R_i^\top$
- C is the measurement matrix defined by $\hat{R}_{ij}$

$$\boxed{\begin{array}{ll} \min\limits_{X \in \mathbb{R}^{n \times n}} & \langle C, XX^\top \rangle \\ \mathrm{s.\,t.} & \mathrm{rank}(X) \le r, \quad X \in \mathrm{St}(3n,3)^n \end{array}} \xrightarrow{\mathcal{H}=\mathrm{St}(3n,3)^n} \boxed{\min_{x \in \mathcal{M}_h} \bar{f}(x) = \langle C, \phi(x)\phi(x)^\top \rangle}$$

Reconstructed errors evaluated by $\|R_i R_j^\top - \hat{R}_{ij}\|_{\mathrm{F}}$





Problem formulation

- State space $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ and Markov model $\mathrm{P} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$
- Hadamard parameterization: $P = X \odot X$, $X \in \mathrm{Ob}(|\mathcal{S}|, |\mathcal{S}|)$
- Proposed model

$$\boxed{\begin{array}{ll} \min\limits_{X\in\mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}} & \frac{1}{2}\|X \odot X - \hat{\mathrm{P}}\|_{\mathrm{F}}^2 \\ \text{s. t.} & \operatorname{rank}(X) \leq r \\ & \operatorname{diag}(XX^\top) - \mathbf{1} = 0 \end{array}} \xrightarrow{\mathcal{H}=\mathrm{Ob}(|\mathcal{S}|,|\mathcal{S}|)} \boxed{\min_{x\in\mathcal{M}_h} \ \frac{1}{2}\|\phi(x) \odot \phi(x) - \hat{\mathrm{P}}\|_{\mathrm{F}}^2}$$

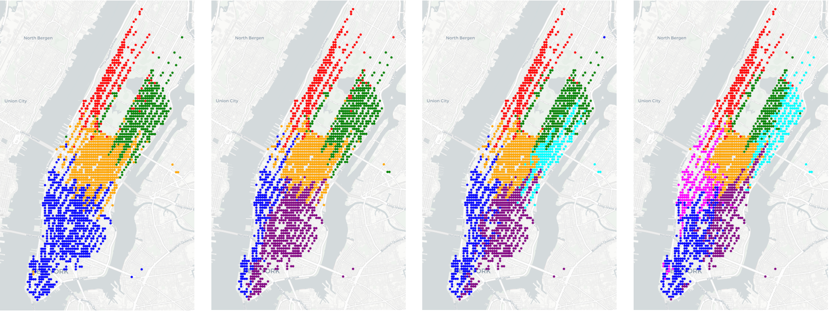
4. Test on the Manhattan transportation network

Problem setting

- Dataset of 1.1×10^7 NYC Yellow cab trips
- Model the transportation dynamics as a Markov proces

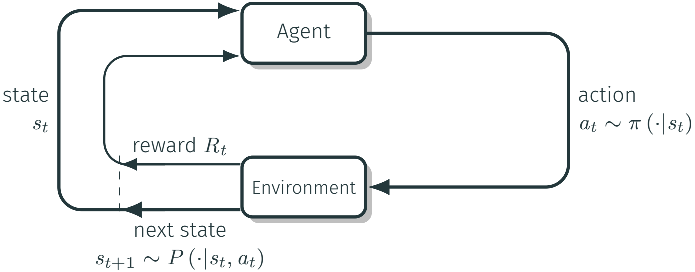
$$\hat{\mathrm{P}}(i,j) = \frac{\sum_{t=1}^{T} \mathbb{I}\left(s_t^{\text{pickup}} = i,\, s_t^{\text{dropoff}} = j\right)}{\sum_t^T \mathbb{I}\left(s_t^{\text{pickup}} = i\right)}, \quad \text{for } \; i,j \in \mathcal{S}$$

State compression



$r = 4$ $\quad$ $r = 5$ $\quad$ $r = 6$ $\quad$ $r = 7$

5. Low-rank reinforcement learning

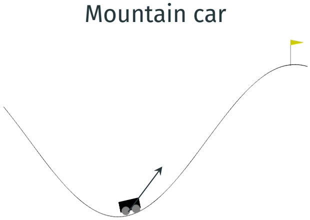


Low-rank strategy

- Hadamard parameterization for policy: $\pi(X) := X \odot X$, $X \in \mathrm{Ob}(|\mathcal{S}|, |\mathcal{A}|)$
- Impose low rank constraint: $\mathrm{rank}(X) \leq r$

$$\boxed{\begin{array}{cl} \min\limits_{X \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{A}|}} & -J(\pi(X)) \\ \mathrm{s.t.} & \mathrm{rank}(X) \leq r, \\ & \mathrm{diag}(XX^\top) - \mathbf{1} = 0 \end{array}} \xrightarrow{\mathcal{H} = \mathrm{Ob}(|\mathcal{S}|, |\mathcal{S}|)} \boxed{\min_{x \in \mathcal{M}_h} -J(\pi(\phi(x)))}$$

5. Test in two environments



Parameter efficiency

Algorithm	Pendulum			Mountain car		
	Parameters	ρ_{storage}	Reward	Parameters	ρ_{storage}	Reward
REINFORCE	$86,961$	1.00	-0.94	$582,096$	1.00	96.99
Q-learning	$86,961$	1.00	-0.75	$582,096$	1.00	83.95
LRRPG ($r = 5$)	$10,810$	0.12	-0.73	$15,485$	0.03	88.31
LRRPG ($r = 10$)	$21,620$	0.25	-0.58	$30,970$	0.05	86.44
LRRPG ($r = 15$)	$32,430$	0.37	-0.83	$46,455$	0.08	78.23

Motivation

- Weight normalization [Salimans-Kingma'16]

$$W \in \mathbb{R}^{m\times n} \Longleftarrow W = (g\mathbf{1}^\top) \odot X \text{ with } g \in \mathbb{R}^m, X \in \mathrm{Ob}(m,n)$$

- Low-rank structure [Idelbayev-Perpinan'22; Qu et al.'24]

Proposed model

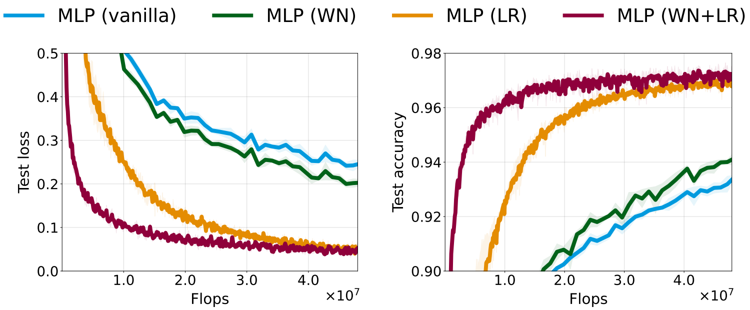
$$\boxed{\begin{aligned} \min_{\{X_i, g_i\}_{i=1}^l} \quad & f(\{X_i, g_i\}_{i=1}^l) \\ \text{s.t.} \quad & \mathrm{rank}(X_i) \le r_i \\ & \mathrm{diag}(X_i X_i^\top) - \mathbf{1} = 0 \end{aligned}} \xrightarrow{\mathcal{H}=\mathrm{Ob}(m,n)^l} \boxed{\min_{x_i \in \mathcal{M}_{h_i},\, g_i} \quad f(\{\phi(x_i), g_i\}_{i=1}^l)}$$

Compared models

Model	Weight norm.	Low rank	Net. param.	Search space	Remark
Vanilla	-	-	$\{W_i\}$	$\mathbb{R}^{m_i\times n_i}$	-
WN	✓	-	$\{X_i, g_i\}$	$\mathrm{Ob}(m_i, n_i) \times \mathbb{R}^{m_i}$	-
LR	-	✓	$\{x_i, g_i\}$	$\mathcal{M}_{h_i} \times \mathbb{R}^{m_i}$	$h_i(\cdot) \equiv 0$
WN+LR	✓	✓	$\{x_i, g_i\}$	$\mathcal{M}_{h_i} \times \mathbb{R}^{m_i}$	$h_i(X_i) = X_i X_i^\top - \mathbf{1}$

6. Test on deep learning tasks

Test on MNIST dataset



Test on CIFAR-10 dataset

Algorithm	Parameters	ρ_{storage}	Infer. flops	ρ_{infer}	Train. acc. (%)	Test acc. (%)
CNN (vanilla)	2.62M	1.00	565.50M	1.00	96.67	87.80
CNN (WN)	2.62M	1.00	565.50M	1.00	100.00	90.31
CNN (LR)	0.38M	0.14	179.02M	0.32	90.01	83.07
CNN (WN+LR)	0.38M	0.14	179.02M	0.32	95.39	87.30

Take-home notes

A space-decoupling framework

Level 1: Tangent space
bridge the rank information and orthogonal invariance

Level 2: Tangent cone
identify tangent and normal cones to $\mathbb{R}^{m\times n}_{\leq r} \cap \mathcal{H}$

Level 3: Parametrization
a space-decoupling parametrization $\mathcal{M}_h$ *and its geometry*

Level 4: Optimization
Riemannian optimization on $\mathcal{M}_h$

References

1. Geometric methods in matrix spaces:
   Yan Yang, **Bin Gao**, Ya-xiang Yuan. *A space-decoupling framework for optimization on bounded-rank matrices with orthogonally invariant constraints.*
   arXiv:2501.13830, (2025)
2. Geometric methods in tensor spaces:
   Bin Gao, Renfeng Peng, Ya-xiang Yuan. *Desingularization of bounded-rank tensor sets.*
   arXiv:2411.14093, (2024)

- Variational analysis of determinantal varieties:
  Yan Yang, **Bin Gao**, Ya-xiang Yuan. *Variational analysis of determinantal varieties*
  arXiv:2511.22613, (2025).

Thanks for your attention!

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

homepage

group blog