Optimization over the intersection of manifolds

Yan Yang

CSIAM Youth Forum, Chengdu, Sichuan

June 27, 2026

State Key Laboratory of Mathematical Sciences
Academy of Mathematics and Systems Science
Chinese Academy of Sciences, China

Joint work with
Dr. Bin Gao (AMSS, CAS)
Prof. Ya-xiang Yuan (AMSS, CAS)

Outline

Background: optimization with geometric constraints

Riemannian optimization problem

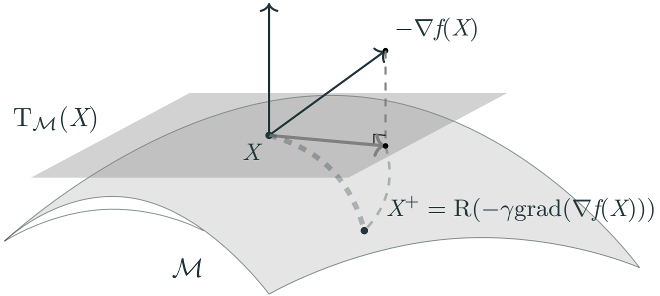
$$\boxed{\min_{X \in \mathcal{M}} f(X)}$$

with $\mathcal{M} \subseteq \mathbb{R}^m$ as a smooth manifold

Geometric ingredients

- Tangent space $\mathrm{T}_{\mathcal{M}}(X)$
- Riemannian gradient $\mathrm{grad}_{\mathcal{M}} f(X)$
- Retraction $\mathrm{R}_x : \mathrm{T}_{\mathcal{M}}(X) \to \mathcal{M}$

Riemannian gradient descent



Specific manifolds: theory and applications

Stiefel and Grassmann manifolds $\{X \in \mathbb{R}^{m\times p} \mid X^\top X = I_p\}$

- Principal component analysis problems [Jolliffe-Trendafilov-Uddin'03; Xu-Jiang-Liu-So'26]
- Computer vision and statistics on manifolds [Turaga-Veeraraghavan-Chellappa'08]
- Electronic structure calculations [Liu-Wang-Wen-Yuan'13; Jiang-Dai'16; Gao-Liu-Yuan'18]

Fixed-rank matrix manifold $\{X \in \mathbb{R}^{m\times n} \mid \mathrm{rank}(X) = r\}$

- Low-rank matrix completion [Vandereycken'13]
- Magnetic resonance imaging (MRI) [Banco-Aeron-Hoge'16; Choi-Bao-Zhang'18; Fessler'20]
- Data analysis, e.g., weather forecast [Loucheur-Absil-Journée'23] and exoplanet detection [Daglayan-Vary-Absil-Cantalloube-Christiaens-Gillis-Jacques-Leplat-Absil'24]

Fixed-rank tensor manifolds $\{\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d} \mid \mathrm{rank}(\mathcal{X}) = \mathbf{r}\}$

- Tucker: tensor completion and parametric PDEs [Kressner–Steinlechner–Vandereycken'14]
- Tensor train: high-dimensional completion [Steinlechner'16; Peng-Zhu-Gao-Wang-Yuan'25]

Additional structures imposed on the manifold

$$\begin{aligned} \min_{X \in \mathcal{E}} \quad & f(X) \\ \text{s.t.} \quad & h(X) = 0 \\ & X \in \mathcal{M} \end{aligned}$$

Blanket assumptions

- Euclidean space $\mathcal{E}$ accommodates $\mathbb{R}^m$, $\mathbb{R}^{m \times n}$, or $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$
- Objective $f \colon \mathcal{E} \to \mathbb{R}$ is continuously differentiable
- Differential $\mathrm{D}h$ has **full rank** in a neighborhood of

$$\mathcal{H} := \{X \in \mathcal{E} \mid h(X) = 0\}$$

Manifold + additional constraints

Additional structures imposed on the manifold

$$\begin{aligned} \min_{X \in \mathcal{E}} \quad & f(X) \\ \text{s. t.} \quad & h(X) = 0 \\ & X \in \mathcal{M} \end{aligned}$$

Blanket assumptions

- Euclidean space $\mathcal{E}$ accommodates $\mathbb{R}^m$, $\mathbb{R}^{m \times n}$, or $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$
- Objective $f \colon \mathcal{E} \to \mathbb{R}$ is continuously differentiable
- Differential $\mathrm{D}h$ has **full rank** in a neighborhood of

$$\mathcal{H} := \{X \in \mathcal{E} \mid h(X) = 0\}$$

$\mathcal{H}$ is a **smooth manifold** embedded in $\mathcal{E}$

Application I. Low rank + additional constraints

Fixed-rank constraint + an additional constraint

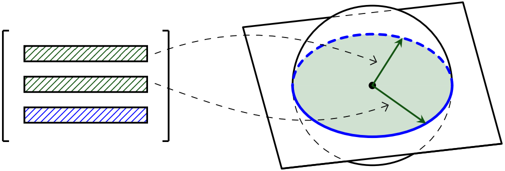
$$
\begin{aligned}
\min_{X \in \mathbb{R}^{m\times n}} \quad & f(X) \\
\text{s. t.} \quad & h(X) = 0 \\
& X \in \mathcal{M}_r = \{X \in \mathbb{R}^{m\times n} \mid \operatorname{rank}(X) = r\}
\end{aligned}
$$

Low rank + oblique manifold

Low-rank data fitting on sphere [Chu-Del Buono-Lopez-Politi'05 SIMAX]

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

$\mathbb{R}_2^{3\times 4} \cap \mathrm{Ob}(3,4) \ni$



Low rank + Frobenius sphere

Approximation of graph similarity matrices [Cason-Absil-Van Dooren'13 LAA]

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

Low rank + hyperbolic manifold

Approximation of hyperbolic embeddings [Jawanpuria-Meghwanshi-Mishra'19 CDC]

$$\mathcal{H} = \mathbb{H}_n^m := \{X \in \mathbb{R}^{m \times n} \mid \mathrm{diag}(X \mathrm{Diag}(-1, 1, \ldots, 1) X^\top) - \mathbf{1} = \mathbf{0}\}$$

Low rank + Frobenius sphere (TT tensors)

Applications in quantum information theory [Peng-Zhu-Gao-Wang-Yuan'25]

$$\mathcal{H} = \{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \mid \|\mathbf{X}\|_{\mathrm{F}}^2 - 1 = 0\}$$

Sparse constraint + an additional constraint

$$\begin{aligned} \min_{X \in \mathbb{R}^m} \quad & f(X) \\ \mathrm{s.\,t.} \quad & h(X) = 0 \\ & X \in \mathcal{C}_s = \{X \in \mathbb{R}^m \mid \|X\|_0 = s\} \end{aligned}$$

Sparse + Frobenius sphere

Gene expression patterns and financial applications [Beck-Vaisbourd'16 JOTA]

$$\mathcal{H} = \{X \in \mathbb{R}^m \mid \|X\|_{\mathrm{F}}^2 - 1 = 0\}$$

Sparse + Stiefel manifold

Sparse principal component analysis problems [Chen-Huang'26]

$$\mathcal{H} = \{X \in \mathbb{R}^{n \times p} \mid X^\top X - I_p = 0\}$$

Optimization over intersection of manifolds

Additional structures imposed on the general manifold $\mathcal{M}$

$$\begin{aligned} \min_{X \in \mathcal{E}} \quad & f(X) \\ \mathrm{s.\,t.} \quad & X \in \mathcal{H} \cap \mathcal{M} \end{aligned}$$

where $\mathcal{H} = \{X \in \mathcal{E} \mid h(X) = 0\}$

Challenges introduced by the coupled constraints

- Lack of manifold structure on $\mathcal{H} \cap \mathcal{M}$
- Geometry of the coupled region $\mathcal{H} \cap \mathcal{M}$
  - Characterization of the tangent cone $\mathrm{T}_{\mathcal{H} \cap \mathcal{M}}(X)$
- Difficulty of preserving the feasibility $X \in \mathcal{H} \cap \mathcal{M}$
  - Absence of closed-form $\mathcal{P}_{\mathcal{H} \cap \mathcal{M}}(\cdot)$
  - Iterates of alternating projections [von Neumann'50; Lewis-Malick'08]

Existing algorithms and motivations

Optimization on $\mathcal{M}$ with equality constraints

$$\begin{aligned} \min_{X\in\mathcal{M}} \quad & f(X) \\ \mathrm{s.\,t.} \quad & h(X) = 0 \end{aligned}$$

- Constraint qualifications and optimality conditions [Bergmannand-Herzog'19 SIOPT; Andreani-Couto-Ferreira-Haeser'24 SIOPT; Andreani-Couto-Ferreira-Haeser-Prudente'26 SIOPT]
- Riemannian augmented Lagrangian methods [Liu-Boumal'20; Zhou-Bao-Ding-Zhu'23 MP]
- Riemannian SQP methods [Schiela-Ortiz'21 SIOPT; Obara-Okuno-Takeda'22 SIOPT]
- Riemannian interior point methods [Lai-Yoshise'24 JOTA]
- Retractions by alternating projections [Chen-He-Huang'26]
- Exact penalty methods when $\mathcal{M}$ is convex [Xiao-Liu-Toh'24 MOR; Xiao-Tang-Wang-Toh'25]

Discussions

☹ A sequence of subproblems should be solved, e.g.,

$$\min_{X\in\mathcal{M}} f(X) + \langle \lambda, h(X)\rangle + \frac{\beta}{2}\|h(X)\|^2$$

Two subtasks

Optimization over the intersection

$$\begin{aligned} \min_{X \in \mathcal{E}} \quad & f(X) \\ \text{s.t.} \quad & X \in \mathcal{H} \cap \mathcal{M} \end{aligned}$$

I. Feasibility

Find points $X \in \mathcal{H} \cap \mathcal{M}$

II. Optimality

Decrease f over $\mathcal{H} \cap \mathcal{M}$

Alternating projections

Method of alternating projections [von Neumann'50]

MAP Iteration

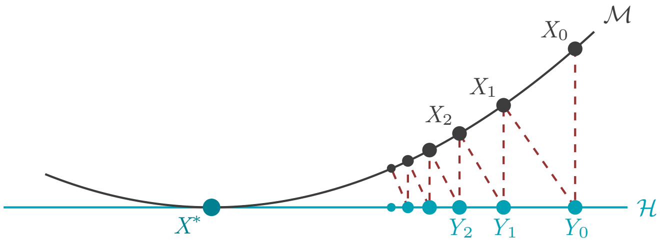
$$X_{k+1} \in \mathcal{P}_{\mathcal{M}}\big(\mathcal{P}_{\mathcal{H}}(X_k)\big)$$

$\Longrightarrow$

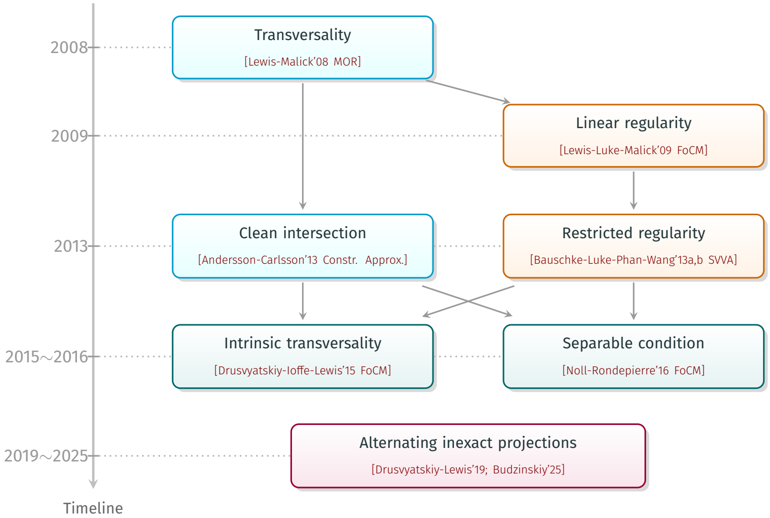
Feasibility Problem

Find a point $X^* \in \mathcal{H} \cap \mathcal{M}$

Illustration



Developments and intersection conditions



Motivation of the notions

Transversality

$\mathrm{T}_{\mathcal{H}}(X) + \mathrm{T}_{\mathcal{M}}(X) \;=\; \mathcal{E}$

→

Clean intersection

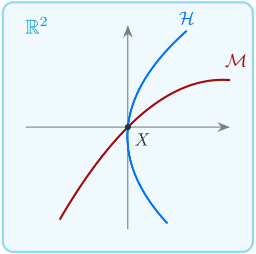
$\mathrm{T}_{\mathcal{H}\cap\mathcal{M}}(X) = \mathrm{T}_{\mathcal{H}}(X) \cap \mathrm{T}_{\mathcal{M}}(X)$

→

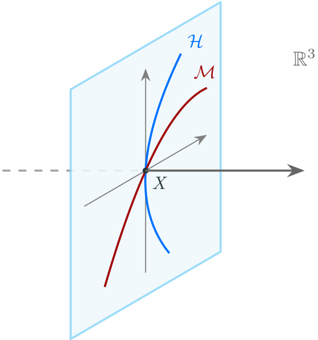
Intrinsic transversality

$\mathrm{N}_{\mathcal{H}}(X)$ and $\mathrm{N}_{\mathcal{M}}(X)$ are "separable"

Open question [Ioffe'17 JOTA]



Trivial lifting



$\mathrm{T}_{\mathcal{H}}(X) + \mathrm{T}_{\mathcal{M}}(X) = \mathbb{R}^2$

Transversality holds

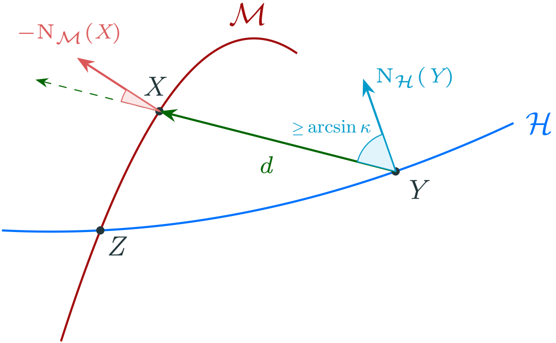
$\mathrm{T}_{\mathcal{H}}(X) + \mathrm{T}_{\mathcal{M}}(X) \subsetneq \mathbb{R}^3$

Transversality fails

The sets $\mathcal{H}$ and $\mathcal{M}$ are ***intrinsically transversal*** at $z \in \mathcal{H} \cap \mathcal{M}$ if there exist $\kappa \in (0, 1]$ and a neighborhood $\mathcal{B}$ of Z such that for all $X \in (\mathcal{M} \cap \mathcal{B}) \setminus \mathcal{H}$ and $Y \in (\mathcal{H} \cap \mathcal{B}) \setminus \mathcal{M}$ with $d = \frac{X - Y}{\|X - Y\|}$, one has

$$\max\left\{ \operatorname{dist}\left(d, -\mathrm{N}_{\mathcal{M}}(X)\right),\ \operatorname{dist}\left(d, \mathrm{N}_{\mathcal{H}}(Y)\right) \right\} \geq \kappa$$

Illustration



A geometric framework

$$\min_{X \in \mathcal{E}} f(X) \quad \text{s.t. } X \in \mathcal{H} \cap \mathcal{M}$$ (Intrinsically Transversal)

$$\min_{X \in \mathcal{M}} f(X) \quad \text{s.t. } h(X) = 0$$

I. Feasibility

Find points $X \in \mathcal{H} \cap \mathcal{M}$

Perspective I on IT

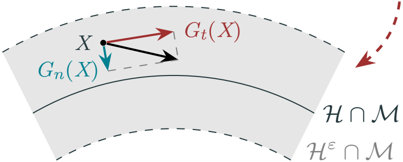
$$G_n(X) \in \mathrm{T}_{\mathcal{M}}(X) \Rightarrow \mathrm{dist}(X, \mathcal{H}) \downarrow$$

II. Optimality

Decrease f over $\mathcal{H} \cap \mathcal{M}$

Perspective II on IT

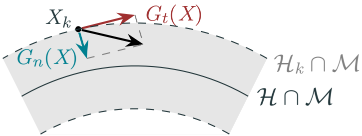
$$G_t(X) \in \mathrm{T}_{\mathcal{M}}(X) \Rightarrow f(X) \downarrow$$



Perturbed manifolds

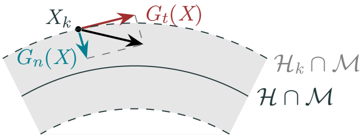
$$\mathcal{H}(Y) := \{X \in \mathcal{E} \mid h(X) = h(Y)\}$$

$$\mathcal{H}_k := \mathcal{H}(X_k)$$



Update rule

$$\boxed{X_{k+1} = \mathrm{R}^{\mathcal{M}}_{X_k}\left(\alpha_k G_n(X_k) + \beta_k G_t(X_k)\right)}$$



Perturbed manifolds

$$\mathcal{H}(Y) := \{X \in \mathcal{E} \mid h(X) = h(Y)\}$$
$$\mathcal{H}_k := \mathcal{H}(X_k)$$

Update rule

$$\boxed{X_{k+1} = \mathrm{R}^{\mathcal{M}}_{X_k}\left(\alpha_k G_n(X_k) + \beta_k G_t(X_k)\right)}$$

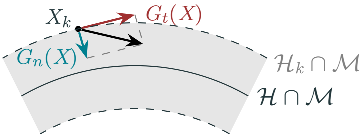
Important observation

Given two manifolds $\mathcal{X}_1, \mathcal{X}_2 \subseteq \mathcal{E}$ with a point $X \in \mathcal{X}_1 \cap \mathcal{X}_2$. Then for any $d \in \mathrm{N}_{\mathcal{X}_2}(X)$ and $\eta \in \mathcal{E}$, it holds that

$$\left\langle \mathcal{P}_{\mathrm{T}_{\mathcal{X}_1}(X)}(d), \mathcal{P}_{\mathrm{T}_{\mathcal{X}_1 \cap \mathcal{X}_2}(X)}(\eta) \right\rangle = 0$$

A geometric framework (cont'd)

Perturbed manifolds



$$\mathcal{H}(Y) := \{X \in \mathcal{E} \mid h(X) = h(Y)\}$$
$$\mathcal{H}_k := \mathcal{H}(X_k)$$

Update rule

$$\boxed{X_{k+1} = \mathrm{R}^{\mathcal{M}}_{X_k}\left(\alpha_k G_n(X_k) + \beta_k G_t(X_k)\right)}$$

Important observation

Given two manifolds $\mathcal{X}_1, \mathcal{X}_2 \subseteq \mathcal{E}$ with a point $X \in \mathcal{X}_1 \cap \mathcal{X}_2$. Then for any $d \in \mathrm{N}_{\mathcal{X}_2}(X)$ and $\eta \in \mathcal{E}$, it holds that

$$\left\langle \mathcal{P}_{\mathrm{T}_{\mathcal{X}_1}(X)}(d), \mathcal{P}_{\mathrm{T}_{\mathcal{X}_1 \cap \mathcal{X}_2}(X)}(\eta) \right\rangle = 0$$

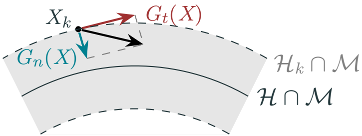
$$\boxed{G_n(X_k) = \mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X_k)}(d_k)}$$

$$\boxed{G_t(X_k) = \mathcal{P}_{\mathrm{T}_{\mathcal{H}_k \cap \mathcal{M}}(X_k)}(\eta_k)}$$

Perturbed manifolds

$$\mathcal{H}(Y) := \{X \in \mathcal{E} \mid h(X) = h(Y)\}$$

$$\mathcal{H}_k := \mathcal{H}(X_k)$$



Update rule

$$X_{k+1} = \mathrm{R}^{\mathcal{M}}_{X_k}\left(\alpha_k \mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X_k)}(d_k) + \beta_k \mathcal{P}_{\mathrm{T}_{\mathcal{H}_k \cap \mathcal{M}}(X_k)}(\eta_k)\right)$$

Important observation

Given two manifolds $\mathcal{X}_1, \mathcal{X}_2 \subseteq \mathcal{E}$ with a point $X \in \mathcal{X}_1 \cap \mathcal{X}_2$. Then for any $d \in \mathrm{N}_{\mathcal{X}_2}(X)$ and $\eta \in \mathcal{E}$, it holds that

$$\left\langle \mathcal{P}_{\mathrm{T}_{\mathcal{X}_1}(X)}(d), \mathcal{P}_{\mathrm{T}_{\mathcal{X}_1 \cap \mathcal{X}_2}(X)}(\eta) \right\rangle = 0$$

$$G_n(X_k) = \mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X_k)}(d_k)$$

$$G_t(X_k) = \mathcal{P}_{\mathrm{T}_{\mathcal{H}_k \cap \mathcal{M}}(X_k)}(\eta_k)$$

Orthogonal tangent directions

Bouliganded tangent cone tangent space for a manifold

$$\mathrm{T}_{\mathcal{X}}(X) = \{\eta \in \mathcal{E} \mid \exists t_i \to 0, \text{ s. t. } \operatorname{dist}(X + t_i\eta, \mathcal{X}) = o(t_i)\}$$

Given $X \in \mathcal{X}_1 \cap \mathcal{X}_2$, it holds that

$$\mathrm{T}_{\mathcal{X}_1 \cap \mathcal{X}_2}(X) \subseteq \mathrm{T}_{\mathcal{X}_1}(X) \cap \mathrm{T}_{\mathcal{X}_2}(X) \quad \text{and} \quad \mathrm{N}_{\mathcal{X}_1 \cap \mathcal{X}_2}(X) \supseteq \mathrm{N}_{\mathcal{X}_1}(X) + \mathrm{N}_{\mathcal{X}_2}(X)$$

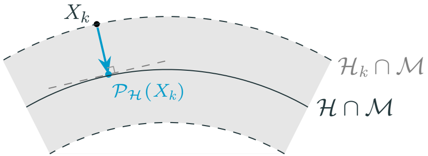
$\mathcal{M}_r$ as a smooth manifold based on the SVD $X = U\Sigma V^\top$

$$\mathrm{T}_{\mathcal{M}_r}(X) = \left\{[U\ U_\perp] \begin{bmatrix} \mathbb{R}^{r\times r} & \mathbb{R}^{r\times(n-r)} \\ \mathbb{R}^{(m-r)\times r} & 0^{(m-r)\times(n-r)} \end{bmatrix} [V\ V_\perp]^\top \right\}$$

$$\mathrm{N}_{\mathcal{M}_r}(X) = \left\{[U\ U_\perp] \begin{bmatrix} 0^{r\times r} & 0^{r\times(n-r)} \\ 0^{(m-r)\times r} & \mathbb{R}^{(m-r)\times(n-r)} \end{bmatrix} [V\ V_\perp]^\top \right\}$$

Tangent directon toward feasibility

$$G_n(X_k) = \mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X_k)}(d_k)$$
$$d_k \in \mathrm{N}_{\mathcal{H}_k}(X_k)$$



Construction of $d_k \in \mathrm{N}_{\mathcal{H}_k}(X_k) = \mathrm{span}\{\nabla h(X_k)\}$

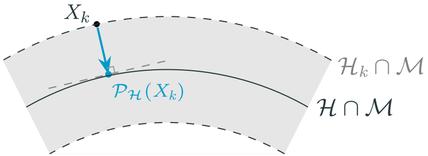
- Projection direction $\tilde{d}_k = X_k - \mathcal{P}_{\mathcal{H}}(X_k)$
- Gauss–Newton direction $d_k = -\nabla h(X_k)(\nabla h(X_k)^\top \nabla h(X_k))^{-1} h(X_k)$

$$d_k = \min_{d \in \mathrm{N}_{\mathcal{H}_k}(X_k)} \|h(X_k) + \nabla h(X_k)^\top d\|^2$$

- First-order surrogate $\|\tilde{d}_k - d_k\| = \mathcal{O}(\|\tilde{d}_k\|^2)$

Tangent directon toward feasibility

$$G_n(X_k) = \mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X_k)}(d_k)$$
$$d_k \in \mathrm{N}_{\mathcal{H}_k}(X_k)$$



Construction of $d_k \in \mathrm{N}_{\mathcal{H}_k}(X_k) = \mathrm{span}\{\nabla h(X_k)\}$

- Projection direction $\tilde{d}_k = X_k - \mathcal{P}_{\mathcal{H}}(X_k)$
- Gauss–Newton direction $d_k = -\nabla h(X_k)(\nabla h(X_k)^\top \nabla h(X_k))^{-1} h(X_k)$

$$d_k = \min_{d \in \mathrm{N}_{\mathcal{H}_k}(X_k)} \|h(X_k) + \nabla h(X_k)^\top d\|^2$$

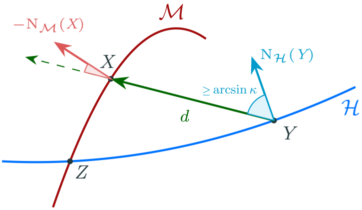
- First-order surrogate $\|\tilde{d}_k - d_k\| = \mathcal{O}(\|\tilde{d}_k\|^2)$

Can $\mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X_k)}(d_k)$ still guarantee $\mathrm{dist}(X_k, \mathcal{H}) \downarrow$?

Perspective I on intrinsic transversality

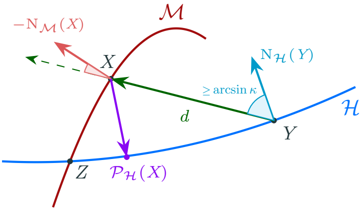
Intrinsic transversality is defined via all (X, Y) pairs around Z

$$\max\{\mathrm{dist}(d, -\mathrm{N}_{\mathcal{M}}(X)),\ \mathrm{dist}(d, \mathrm{N}_{\mathcal{H}}(Y))\} \geq \kappa$$



Intrinsic transversality is defined via **all** (X, Y) pairs around Z

$$\max\{\mathrm{dist}(d, -\mathrm{N}_{\mathcal{M}}(X)),\ \mathrm{dist}(d, \mathrm{N}_{\mathcal{H}}(Y))\} \geq \kappa$$



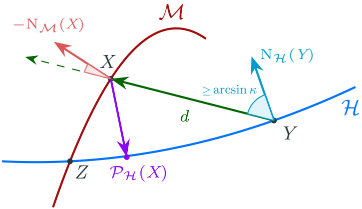
Projection-based transversality $\mathcal{H}, \mathcal{M}$ are C^2 submanifolds

There exists $\kappa' \in (0, 1]$ such that for the **specific** pairs, it holds that

$$(X, \mathcal{P}_{\mathcal{H}}(X)) \Rightarrow \mathrm{dist}\left(d, -\mathrm{N}_{\mathcal{M}}(X)\right) \geq \kappa' \quad (Y, \mathcal{P}_{\mathcal{M}}(Y)) \Rightarrow \mathrm{dist}\left(d, \mathrm{N}_{\mathcal{H}}(Y)\right) \geq \kappa'$$

Intrinsic transversality is defined via
all (X, Y) pairs around Z

$$\max\{\mathrm{dist}(d, -\mathrm{N}_{\mathcal{M}}(X)),\ \mathrm{dist}(d, \mathrm{N}_{\mathcal{H}}(Y))\} \geq \kappa$$



Projection-based transversality $\mathcal{H}, \mathcal{M}$ are C^2 submanifolds

There exists $\kappa' \in (0, 1]$ such that for the specific pairs, it holds that

$$(X, \mathcal{P}_{\mathcal{H}}(X)) \Rightarrow \mathrm{dist}\left(d, -\mathrm{N}_{\mathcal{M}}(X)\right) \geq \kappa' \quad (Y, \mathcal{P}_{\mathcal{M}}(Y)) \Rightarrow \mathrm{dist}\left(d, \mathrm{N}_{\mathcal{H}}(Y)\right) \geq \kappa'$$

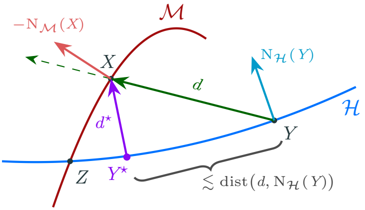
Theorem

Intrinsic transversality $\Longleftrightarrow$ Projection-based transversality

Theorem $\mathcal{H}, \mathcal{M}$ are C^2 submanifolds

Intrinsic transversality $\iff$ Projection-based transversality

Nontrivial direction "$\Longleftarrow$"



- Assume IT fails: (X, Y) with d nearly normal to both $\mathcal{H}, \mathcal{M}$; $Y^\star := \mathcal{P}_{\mathcal{H}}(X)$
- Prox-regularity of manifolds gives $\|Y - Y^\star\| \lesssim \mathrm{dist}(d, \mathrm{N}_{\mathcal{H}}(Y))$, and thus

$$\begin{aligned}\mathrm{dist}\big(d^\star, \mathrm{N}_{\mathcal{M}}(X)\big) &\le \mathrm{dist}\big(d, \mathrm{N}_{\mathcal{M}}(X)\big) + C_1\|d - d^\star\| \\ &\le \mathrm{dist}\big(d, \mathrm{N}_{\mathcal{M}}(X)\big) + C_2\ \mathrm{dist}\big(d, \mathrm{N}_{\mathcal{H}}(Y)\big)\end{aligned}$$

- IT fails $\iff$ both terms are small $\implies$ $(X, Y^\star)$ leads to a contradiction

Tangent direction I: projected Gauss–Newton direction

$$\begin{array}{ll} \min\limits_{X \in \mathcal{E}} & f(X) \\ \text{s.t.} & X \in \mathcal{H} \cap \mathcal{M} \end{array}$$

$$G_n(X_k) = \mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X_k)}(d_k)$$
$$d_k = -\nabla h(X_k)(\nabla h(X_k)^\top \nabla h(X_k))^{-1} h(X_k)$$

Projection-based transversality

$$\|\mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X_k)}(d_k)\| \geq \kappa' \|d_k\|$$

Non-degeneracy of ∇h

$$\|d_k\| \geq \sigma_{\max}^{-1}(\nabla h(X_k)) \|h(X_k)\|$$

Optimality analysis

$$\begin{aligned} G_n(X_k) = 0 \iff \mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X_k)}(d_k) = 0 \\ \implies h(X_k) = 0 \end{aligned}$$

Perspective II on intrinsic transversality

$$
\begin{tikzcd}
\begin{array}{c}\text{Transversality}\\ \mathrm{T}_{\mathcal{H}}(X) + \mathrm{T}_{\mathcal{M}}(X) \;=\; \mathcal{E}\end{array} \arrow[r] & \begin{array}{c}\text{Clean intersection}\\ \mathrm{T}_{\mathcal{H}\cap\mathcal{M}}(X) = \mathrm{T}_{\mathcal{H}}(X) \cap \mathrm{T}_{\mathcal{M}}(X)\end{array} \arrow[r] & \begin{array}{c}\text{Intrinsic transversality}\\ \mathrm{N}_{\mathcal{H}}(X) \text{ and } \mathrm{N}_{\mathcal{M}}(X) \text{ are ``separable''}\end{array} \arrow[l, bend left, dashed, "\text{Open question [Ioffe'17 JOTA]}"]
\end{tikzcd}
$$

Transversality	Clean intersection	Intrinsic transversality
$\mathrm{T}_{\mathcal{H}}(X) + \mathrm{T}_{\mathcal{M}}(X) = \mathcal{E}$	$\mathrm{T}_{\mathcal{H}\cap\mathcal{M}}(X) = \mathrm{T}_{\mathcal{H}}(X) \cap \mathrm{T}_{\mathcal{M}}(X)$	$\mathrm{N}_{\mathcal{H}}(X)$ and $\mathrm{N}_{\mathcal{M}}(X)$ are "separable"

Open question [Ioffe'17 JOTA]

Theorem answers the open question

Let $\mathcal{H}, \mathcal{M} \subseteq \mathcal{E}$ be C^1 manifolds, and they are intrinsically transversal at X. Then they intersect *cleanly* near X: there exists a neighborhood $\mathcal{B}$ of X such that $\mathcal{H} \cap \mathcal{M} \cap \mathcal{B}$ is a C^1 manifold and, for all $Y \in \mathcal{B}$, it holds that

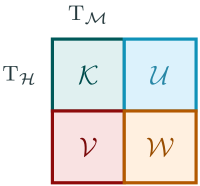
$$\boxed{\mathrm{T}_{\mathcal{H}\cap\mathcal{M}}(Y) = \mathrm{T}_{\mathcal{H}}(Y) \cap \mathrm{T}_{\mathcal{M}}(Y)}$$

Proof sketch

Theorem $\mathcal{H}, \mathcal{M}$ are C^1 submanifolds, intrinsically transversal at Z

$\mathcal{H} \cap \mathcal{M}$ is a C^1 submanifold near Z with $\mathrm{T}_{\mathcal{H}\cap\mathcal{M}} = \mathrm{T}_{\mathcal{H}} \cap \mathrm{T}_{\mathcal{M}}$

- Intersection regularities are invariant under diffeomorphisms
- Flatten $\mathcal{H}$ into a linear space; orthogonally decompose $\mathcal{E}$ along $(\mathrm{T}_{\mathcal{H}}, \mathrm{T}_{\mathcal{M}})$



$\mathcal{E} = \mathcal{K} \oplus \mathcal{U} \oplus \mathcal{V} \oplus \mathcal{W}$

$\mathcal{K} = \mathrm{T}_{\mathcal{H}} \cap \mathrm{T}_{\mathcal{M}}$ and $\mathcal{W} = (\mathrm{T}_{\mathcal{H}} + \mathrm{T}_{\mathcal{M}})^{\perp}$

Graph representation $\begin{cases} \mathcal{M} = \{(k, a(k,v), v, b(k,v)) \mid (k,v) \in \mathcal{K} \oplus \mathcal{V}\} \\ \mathcal{H} = \{(k, u, 0, 0) \mid (k,u) \in \mathcal{K} \oplus \mathcal{U}\} \end{cases}$

★ Intrinsic transversality implies that $b(k,0) \equiv 0$ for all $k \in \mathcal{K}$ small enough

- $\mathcal{H} \cap \mathcal{M} = \{(k, a(k,0), 0, 0) \mid k \in \mathcal{K}\}$ is a graph on $\mathcal{K}$, thus C^1 **embedded**

Tangent direction II: projected anti-gradient direction

$$\begin{array}{ll} \min\limits_{X \in \mathcal{E}} & f(X) \\ \mathrm{s.\,t.} & X \in \mathcal{H} \cap \mathcal{M} \end{array}$$

$$G_t(X_k) = \mathcal{P}_{\mathrm{T}_{\mathcal{H}_k \cap \mathcal{M}}(X_k)}\left(-\nabla f(X_k)\right)$$

Optimality analysis

$$\begin{aligned} G_t(X_k) = 0 &\iff \mathcal{P}_{\mathrm{T}_{\mathcal{H}_k \cap \mathcal{M}}(X_k)}\left(-\nabla f(X_k)\right) = 0 \\ &\iff -\nabla f(X_k) \in \mathrm{N}_{\mathcal{H}_k \cap \mathcal{M}}(X_k) \end{aligned}$$

Descent property

$$\left\langle \mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X)}\left(\nabla f(X)\right), \mathcal{P}_{\mathrm{T}_{\mathcal{H} \cap \mathcal{M}}(X)}\left(-\nabla f(X)\right) \right\rangle = -\left\| \mathcal{P}_{\mathrm{T}_{\mathcal{H} \cap \mathcal{M}}(X)}\left(-\nabla f(X)\right) \right\|^2$$

Projection onto the tangent space

Proposition $S(X) = \mathrm{T}_{\mathcal{H}(X)}(X) \cap \mathrm{T}_{\mathcal{M}}(X)$

Assume that $\mathcal{H}(X)$ and $\mathcal{M}$ are intrinsically transversal at X, and let $\Phi_{\mathcal{M}}(X) := \mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X)} \circ \mathrm{D}h_X^*$. Then for any $\xi \in \mathcal{E}$,

$$\mathcal{P}_{S(X)}(\xi) = \mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X)}(\xi) - \Phi_{\mathcal{M}}(X)\big(\mathrm{D}h_X \circ \Phi_{\mathcal{M}}(X)\big)^{\dagger}\mathrm{D}h_X\big(\mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X)}(\xi)\big)$$

When they are transversal, the pseudoinverse reduces to the ordinary inverse

☺ Computing $G_t(X) = \mathcal{P}_{\mathrm{T}_{\mathcal{H}\cap\mathcal{M}}(X)}(-\nabla f(X))$ is tractable

Projection onto the tangent space (cont'd)

Manifold	Level set of h	$\mathcal{P}_{\mathrm{T}_{\mathcal{H}\cap\mathcal{M}}(X)}$	Format
$\mathcal{M}_r$	$\{X \in \mathbb{R}^{m\times n} \mid \mathcal{A}(X) = b\}$	[Yang-Gao-Yuan'25]	
$\mathcal{M}_r$	$\mathcal{H}$ is orthogonally invariant	[Yang-Gao-Yuan'26 MP]	
$\mathcal{M}_r$	$\mathcal{H}$ is hyperbolic	Appendix A	low-rank
$\mathcal{S}_r(n)$	$\{X \in \mathcal{S}(n) \mid \Vert X\Vert_{\mathrm{F}}^2 = 1\}$	[Li-Xiu-Zhou'20 JOTA]	
$\mathcal{M}_{\mathbf{r}}^{\mathrm{ht}}$	$\{\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d} \mid \Vert\mathbf{X}\Vert_{\mathrm{F}}^2 = 1\}$	[Yang-Gao-Yuan'25]	
$\mathcal{M}_{\mathbf{r}}^{\mathrm{tt}}$	Reshaped Stiefel manifold	Appendix B	
$\mathcal{C}_s$	$\{X \in \mathbb{R}^{n} \mid \Vert X\Vert_{\mathrm{F}}^2 = 1\}$	[Beck-Hallak'16 MOR]	sparse
$\mathcal{C}_s$	$\{X \in \mathbb{R}^{n\times p} \mid X^\top X = I_p\}$	[Chen-Huang'26]	
$\mathcal{M}$	$\mathcal{H}\cap\mathcal{M}$ is intrinsically transversal	**Our proposition**	general

Algorithm and convergence analysis

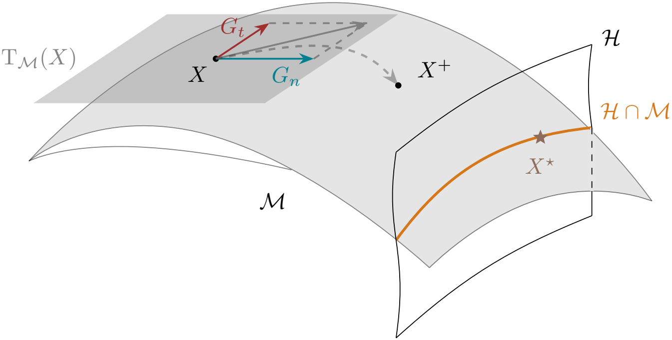
A Geometric method via Orthogonal Tangent Directions (GOTD)

$$d_k = -\nabla h(X_k)(\nabla h(X_k)^\top \nabla h(X_k))^{-1} h(X_k)$$

$$\eta_k = -\nabla f(X_k)$$

$$X_{k+1} = \mathrm{R}^{\mathcal{M}}_{X_k}\left(\alpha_k \mathcal{P}_{\mathrm{T}_{\mathcal{M}}(X_k)}(d_k) + \beta_k \mathcal{P}_{\mathrm{T}_{\mathcal{H}_k \cap \mathcal{M}}(X_k)}(\eta_k)\right)$$

Illustration



Assumptions

- There exists a constant $\rho > 0$ such that ∇h has full rank on
$$\mathcal{K}(\rho) := \{X \in \mathbb{R}^m \mid \|h(X)\| \leq \rho\}$$
- The intersection $\mathcal{H}(Y) \cap \mathcal{M}$ is intrinsically transversal for $Y \in \mathcal{K}(\rho)$
- Riemannian L-smoothness of f on $\mathcal{M}$

Optimality

- The condition $\|G_t(X)\| \leq \varepsilon$ reveals $\|\mathcal{P}_{\mathrm{T}_{\mathcal{H}(X)\cap\mathcal{M}}}(-\nabla f(X))\| \leq \varepsilon$
- When $\varepsilon = 0$ and $h(X) = 0$, the first-order optimality holds
$$\mathcal{P}_{\mathrm{T}_{\mathcal{H}\cap\mathcal{M}(X)}}(-\nabla f(X)) = 0$$

Theorem

In GOTD, with appropriate constant step sizes α, β, the iterates $\{X_k\}$ satisfy
$$\frac{1}{K}\sum_{k=1}^{K}\Big(\|G_t(X_k)\|^2 + \mathrm{dist}(X_k, \mathcal{H})\Big) = \mathcal{O}(\frac{1}{K})$$

Numerical experiments

Low-rank approximation of spherical data

Approximate $A \in \mathcal{H} = \mathrm{Ob}(m, n)$ with a low-rank matrix [Chu et al., 2005 SIAMX]

$$
\begin{aligned}
\min_{X \in \mathbb{R}^{m \times n}} \quad & \frac{1}{2} \left\| \mathcal{P}_\Omega (X - A) \right\|^2 \\
\text{s. t.} \quad & \mathrm{diag}(XX^\top) - \mathbf{1} = 0 \\
& \mathrm{rank}(X) = r
\end{aligned}
$$

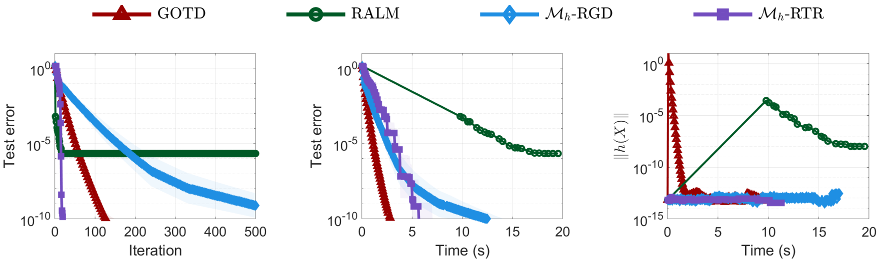
Methods based on Manopt v8.0

- GOTD (ours)
- Riemannian augmented Lagrangian method (RALM) [Liu-Boumal'20]
- Space-decoupling parameterization $\mathcal{M}_h$ [Yang-Gao-Yuan'26 MP]

Settings

- Synthetic $A \in \mathcal{M}_{\leq r} \cap \mathrm{Ob}(m, n)$, $(m, n) = (5000, 6000)$
- rank parameter $r = r^* = 10$, oversampling factor $\mathrm{OS} = 6$

Test on synthetic data



- ☺ RALM and $\mathcal{M}_h$-RTR exhibit better iteration complexity, but they should solve a subproblem for each iterate
- ☺ GOTD recovers the matrix faster than RALM and methods on $\mathcal{M}_h$
- ☺ Geometric methods achieve better precision

Low-rank approximation of hyperbolic embeddings

Approximate $A \in \mathbb{H}_n^m$ under hyperbolic distance [Nickel-Kiela'17]

$$\begin{aligned} \min_{X \in \mathbb{R}^{(n+1)\times m}} \quad & \sum_{i=1}^{m} \operatorname{arccosh}(-\langle X_i, A_i \rangle_{\mathcal{L}})^2 \\ \text{s.t.} \quad & X \in \mathbb{H}_n^m \\ & \operatorname{rank}(X) = r+1 \end{aligned}$$

$$\mathcal{H} = \mathbb{H}_n^m := \{X \in \mathbb{R}^{(n+1)\times m} \mid \operatorname{diag}(X^\top \operatorname{Diag}(-1, 1, \ldots, 1)X) - \mathbf{1} = \mathbf{0}\}$$

Methods

- GOTD (ours)
- Riemannian augmented Lagrangian method (RALM) [Liu-Boumal'20]
- Riemannian parameterization (RGD, RTR) [Jawanpuria-Meghwanshi-Mishra'19 CDC]

Settings

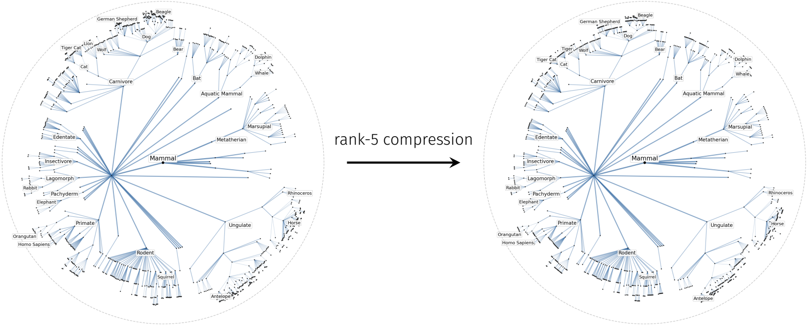
- The mammal subtree of the WordNet database
- $(n, m) = (300, 1170)$, rank parameter $r = 5, 10, 20$

Results of low-rank hyperbolic embeddings

Evaluation on mean average precision (original embeddings: 0.9385)

Rank	GOTD (Ours)	RALM	RGD	RTR
$r = 5$	0.8869	0.6637	0.8881	**0.8883**
$r = 10$	**0.9039**	0.8117	0.8961	0.9024
$r = 20$	**0.9003**	0.8463	0.8989	0.8981

Illustration on 2-dimensional Poincaré ball



Spatially localized solutions of the Schrödinger's equation

$$\begin{aligned} \min_{X \in \mathbb{R}^{n \times p}} \quad & \operatorname{tr}(X^\top H X) \\ \text{s.t.} \quad & X \in \operatorname{St}(n, p) \\ & X \in \mathcal{M} = \{ Y \in \mathbb{R}^{n \times p} \mid \|Y\|_0 = s \} \end{aligned} \xrightarrow{\text{penalty}} \begin{aligned} \min_{X \in \mathbb{R}^{n \times p}} \quad & \operatorname{tr}(X^\top H X) + \mu \|X\|_1 \\ \text{s.t.} \quad & X \in \operatorname{St}(n, p) \end{aligned}$$

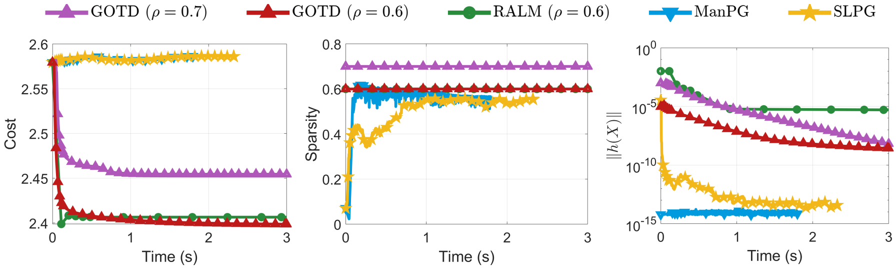
Methods

- GOTD (ours)
- Riemannian augmented Lagrangian method (RALM) [Liu-Boumal'20]
- Manifold proximal gradient method (ManPG) [Chen-Ma-So-Zhang'20]
- Sequential linearized proximal gradient (SLPG) [Liu-Xiao-Yuan'24]

Settings

- H is symmetric, discretizing the Hamiltonian $-\frac{1}{2}\partial_x^2$
- $(n, p) = (256, 15)$, sparsity $\rho := (np - \|X\|_0)/np \in \{0.6,\, 0.7\}$

Comparison on cost, sparsity, and feasibility



- ☺ GOTD finds lower cost with lower density than ManPG and SLPG
- ☺ GOTD achieves better feasibility than RALM

Conclusion

$$\min_{X\in\mathcal{E}} f(X) \quad \text{s.t. } X \in \mathcal{H}\cap\mathcal{M}$$ (Intrinsically Transversal)

$$\min_{X\in\mathcal{M}} f(X) \quad \text{s.t. } h(X) = 0$$

I. Feasibility

Find points $X \in \mathcal{H} \cap \mathcal{M}$

Perspective I on IT

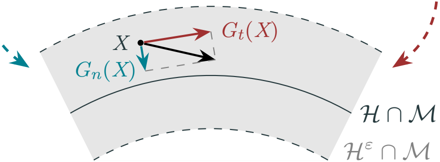
$$G_n(X) \in \mathrm{T}_{\mathcal{M}}(X) \Rightarrow \mathrm{dist}(X, \mathcal{H}) \downarrow$$

II. Optimality

Decrease f over $\mathcal{H} \cap \mathcal{M}$

Perspective II on IT

$$G_t(X) \in \mathrm{T}_{\mathcal{M}}(X) \Rightarrow f(X) \downarrow$$



Conclusion and future work

- ☺ Efficient algorithm exploiting the geometry [Yang-Gao-Yuan'25]

$$\mathrm{T}_{\mathcal{H}\cap\mathcal{M}}(X) = \mathrm{T}_{\mathcal{H}}(X) \cap \mathrm{T}_{\mathcal{M}}(X)$$

- ? Second-order development beneficial from the property [Yang-Gao-Yuan'25]

$$\mathrm{T}^2_{\mathcal{H}\cap\mathcal{M}}(X;\eta) = \mathrm{T}^2_{\mathcal{H}}(X;\eta) \cap \mathrm{T}^2_{\mathcal{M}}(X;\eta)$$

- ? Reasonable extension to multiple manifolds $\mathcal{M}$, $\mathcal{H}_1, \ldots, \mathcal{H}_N$
- ? Technical adaptation to non-smooth sets, e.g., $\mathcal{M} = \mathcal{M}_{\leq r}$, $\mathcal{C}_{\leq s}$
- ? Stochastic oracles and variance-reduction algorithms

References

Yan Yang, Bin Gao, Ya-xiang Yuan. *A space-decoupling framework for optimization on bounded-rank matrices with orthogonally invariant constraints.* Mathematical Programming (2026)

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

Thanks for your attention!

Email: yangyan@amss.ac.cn
Homepage: https://ucas-yanyang.github.io/