筑波大学
University of Tsukuba

Graduate School of Science and Technology
Degree Programs in Systems and Information Engineering
University of Tsukuba

Riemannian Optimization Algorithms for Applications and Their Theoretical Properties

Zhijian Lai

2024/01/22

PhD Thesis Final Defense

Supervisor: Prof. Akiko Yoshise

Committees: Prof. Hiroyuki Sato
Prof. Ying Miao
Prof. Masahiro Hachimori
Prof. Takahito Kuno

Publication List

This talk covers

Published:

1. Lai, Z., Yoshise, A. "Completely positive factorization by a Riemannian smoothing method." *Computational Optimization and Applications*, 83, 933-966 (2022).

Under Review:

2. Lai, Z., Yoshise, A. "Riemannian Interior Point Methods for Constrained Optimization on Manifolds." *Journal of Optimization Theory and Applications.* (2nd round with minor revision)

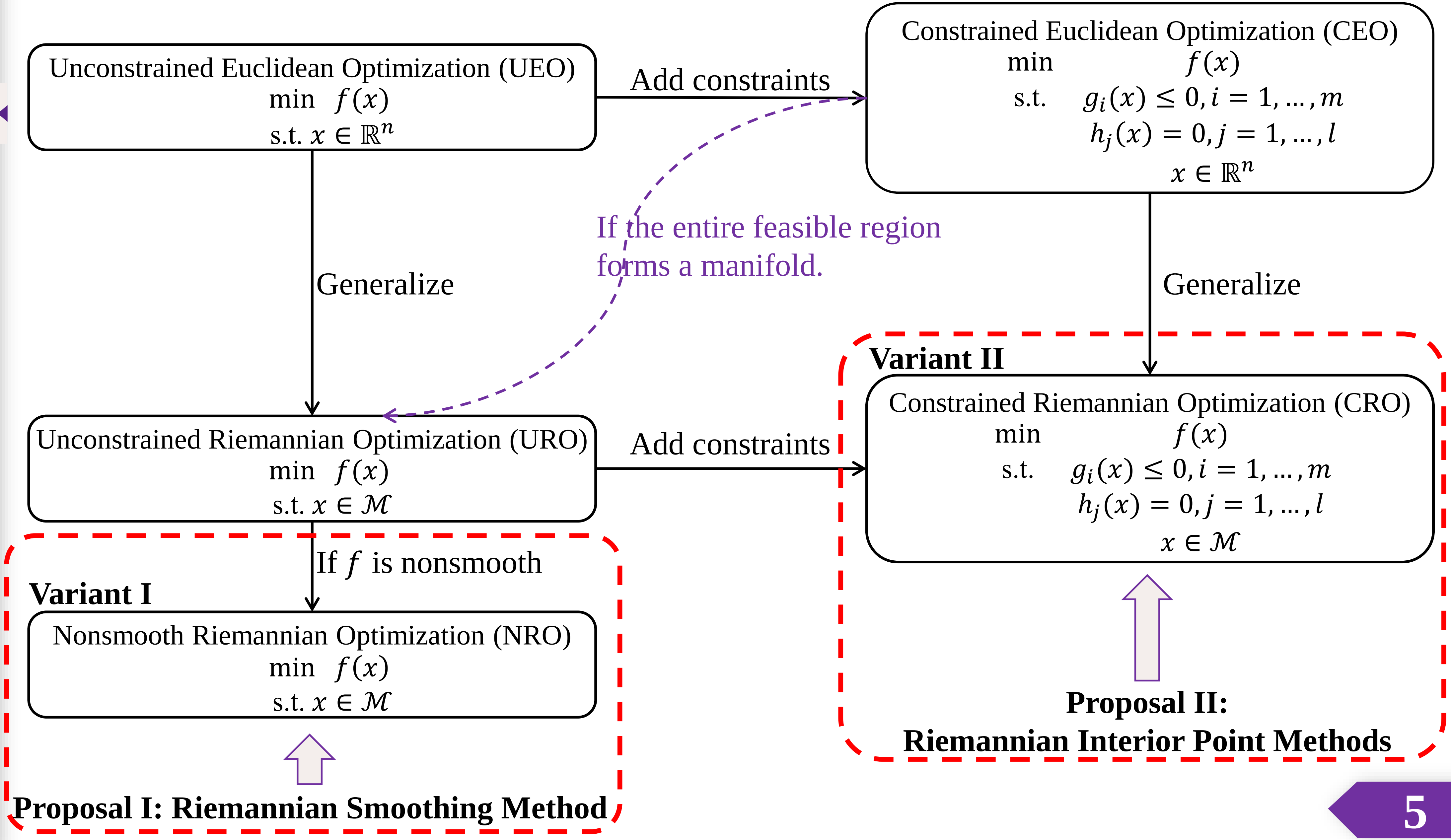
Contents

Click here to Appendix

• Section 1

Background:
What is the Riemannian Optimization?

Goal of This Section: To Show the Position of Our Works



Start from Unconstrained Euclidean Optimization (UEO)

Unconstrained Euclidean Optimization (UEO)

$$\min \ f(x)$$
$$\text{s.t. } x \in \mathbb{R}^n$$

→ Euclidean space $\mathbb{R}^n$

General Line Search Framework for (UEO)

1. Compute a search direction $d_k \in \mathbb{R}^n$;
2. Compute a step length $t_k > 0$;
3. Compute the next point as $x_{k+1} := x_k + t_k d_k$;
4. Set $k \leftarrow k + 1$;

Use local information of f at x_k:

- steepest descent direction: $d_k = -\nabla f(x_k)$
- Newton direction: $d_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$

For arbitrary d_k, t_k, the new point $x_{k+1} \in \mathbb{R}^n$. This is why we say that these types of problems are "unconstrained".

⇨ Can line search framework be used for **constrained optimization**?

Constrained Euclidean Optimization (CEO) is Hard!

Unconstrained Euclidean Optimization (UEO)

$$\min \ f(x) \quad \text{s.t. } x \in \mathbb{R}^n$$

Add constraints →

Constrained Euclidean Optimization (CEO)

$$\min \ f(x)$$
$$\text{s.t. } g_i(x) \le 0, i = 1, \dots, m$$
$$h_j(x) = 0, j = 1, \dots, l$$
$$x \in \mathbb{R}^n$$

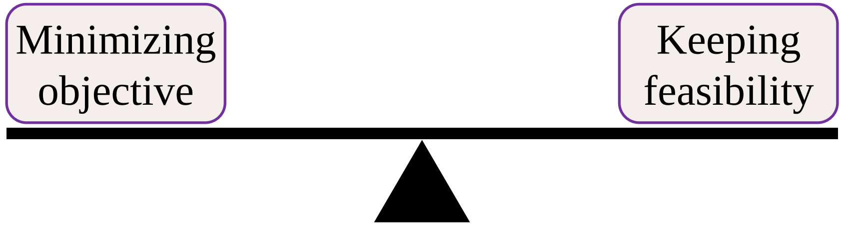
feasible region

General Line Search Framework for (UEO)

1. Compute a search direction $d_k \in \mathbb{R}^n$;
2. Compute a step length $t_k > 0$;
3. Compute the next point as $x_{k+1} := x_k + t_k d_k$;
4. Set $k \leftarrow k + 1$;

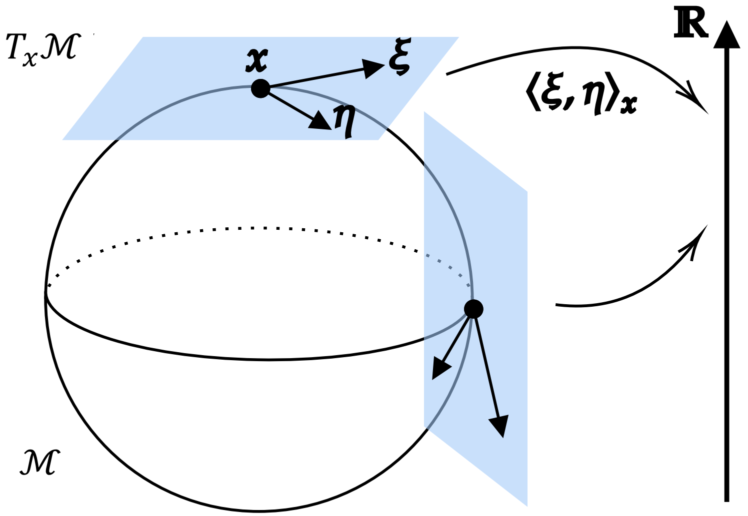
Unable to solve!
Because $x_{k+1} := x_k + t_k d_k$ may not be feasible.

CEO algorithms are often more difficult than UEO:



But if the **feasible region is a manifold** $\mathcal{M}$, we can use line search framework to solve it.
This is the **Riemannian Optimization** --- optimization on manifold $\mathcal{M}$, instead of Euclidean space $\mathbb{R}^n$.

A Glance at Riemannian Manifold/Optimization (Details are presented in the next section)



◆ What is Riemannian manifold?

Riemannian manifold = manifold + Riemannian metric.

A set $\mathcal{M}$ that can be locally linearized.

- $T_x\mathcal{M}$ is tangent space at x.
- $\xi \in T_x\mathcal{M}$ is tangent vector at x.

A Riemannian metric assigns a smooth inner product $\langle\cdot,\cdot\rangle_x$ to each tangent space.

◆ What is Riemannian Optimization?

Unconstrained Riemannian Optimization (URO)

$$\min \ f(x)$$
$$\text{s.t. } x \in \mathcal{M}$$

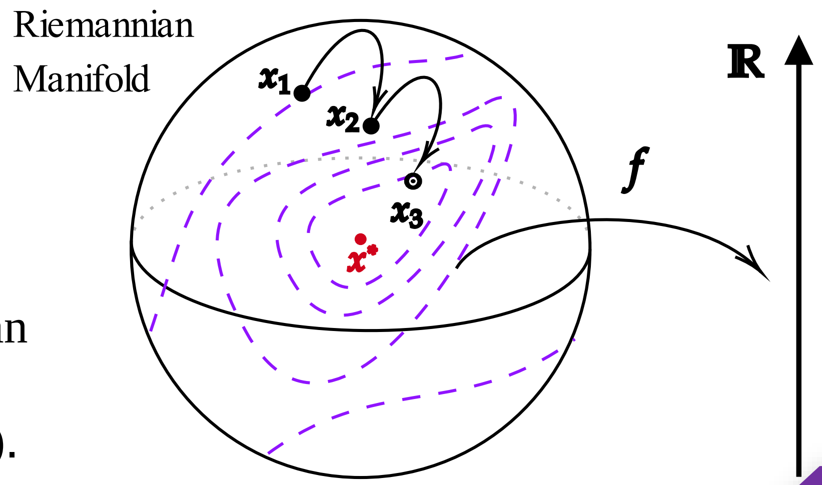
Constrained Euclidean optimization,

$$\min_{x\in\mathbb{R}^n} f(x) \ \text{ s.t. } x^\top x = 1.$$

⇔

Unconstrained Riemannian optimization,

$$\min_{x\in\mathbb{S}^{n-1}=\{x\in\mathbb{R}^n:\|x\|_2=1\}} f(x).$$



Line Search Framework for Riemannian Optimization (Details are presented in the next section)

Unconstrained Euclidean Optimization (UEO)
$\min \ f(x)$
s.t. $x \in \mathbb{R}^n$

⇒

Unconstrained Riemannian Optimization (URO)
$\min \ f(x)$
s.t. $x \in \mathcal{M}$

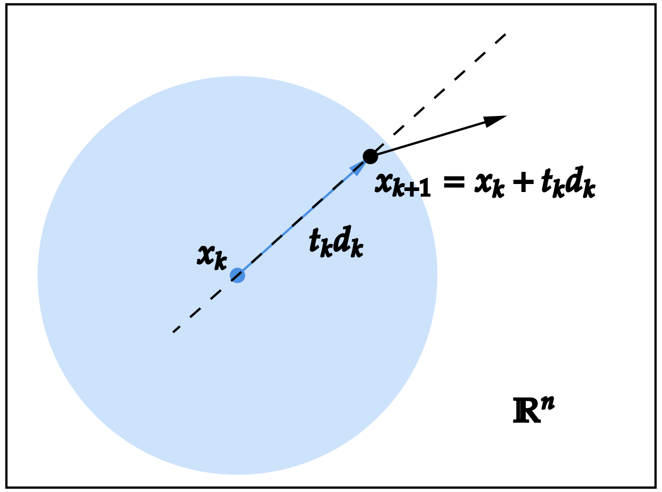
General Line Search Framework for (UEO)

1. Compute a search direction $d_k \in \mathbb{R}^n$;
2. Compute a step length $t_k > 0$;
3. Compute the next point as $x_{k+1} := x_k + t_k d_k$;
4. $k \leftarrow k + 1$;

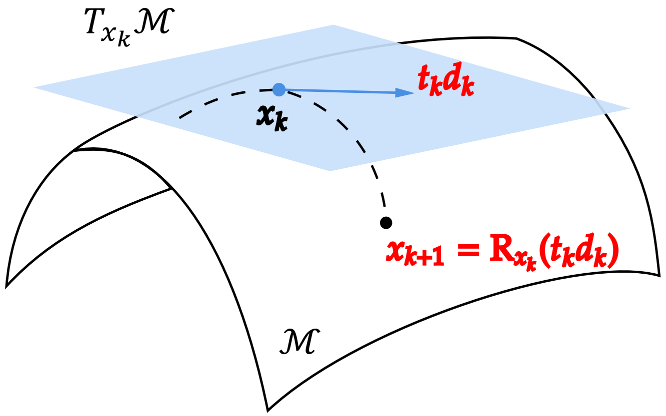
⇒

General Line Search Framework for (URO)

1. Compute a search direction $d_k \in T_{x_k}\mathcal{M}$;
2. Compute a step length $t_k > 0$;
3. Compute the next point $x_{k+1} := \mathrm{R}_{x_k}(t_k d_k)$
4. $k \leftarrow k + 1$;



⇒



Use Retraction to back to $\mathcal{M}$.

Advantages in Comparison to Euclidean Optimization

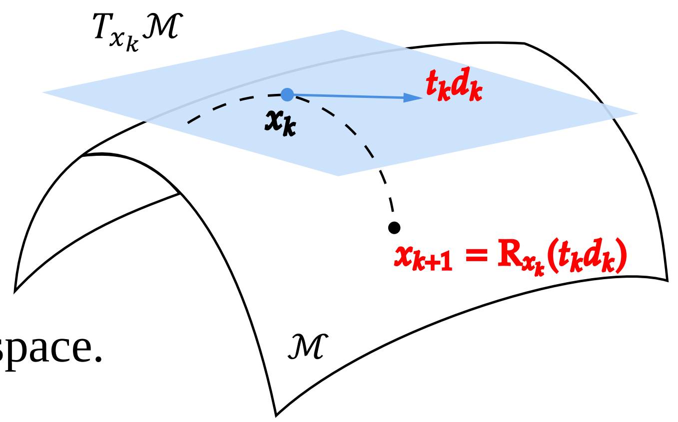
Riemannian version of classical methods (2002-):
Riemannian steepest decent,
Riemannian conjugate gradient,
Riemannian trust region,
Riemannian Newton,
Riemannian BFGS,
Riemannian proximal gradient,
Riemannian stochastic algorithms,
Riemannian ADMM and more.

Unconstrained Riemannian Optimization (URO)

$$\min \; f(x) \quad \text{s.t. } x \in \mathcal{M}$$

General Line Search Framework for (URO)

1. Compute a search direction $d_k \in T_{x_k}\mathcal{M}$;
2. Compute a step length $t_k > 0$;
3. Compute the next point $x_{k+1} := \mathrm{R}_{x_k}(t_k d_k)$
4. $k \leftarrow k + 1$;



Advantages of Riemannian Optimization:
1. All iterates on the manifold.
2. Transform constrained problems into unconstrained ones.
3. Use of the geometric structure of the feasible region.
4. Convergence properties of like optimization on Euclidean space.

Use Retraction to back to $\mathcal{M}$.

Current Manifolds & Applications & Citation

- **40 + available manifolds $\mathcal{M}$ in Riemannian solver "Manopt" (Boumal et al., 2014):**

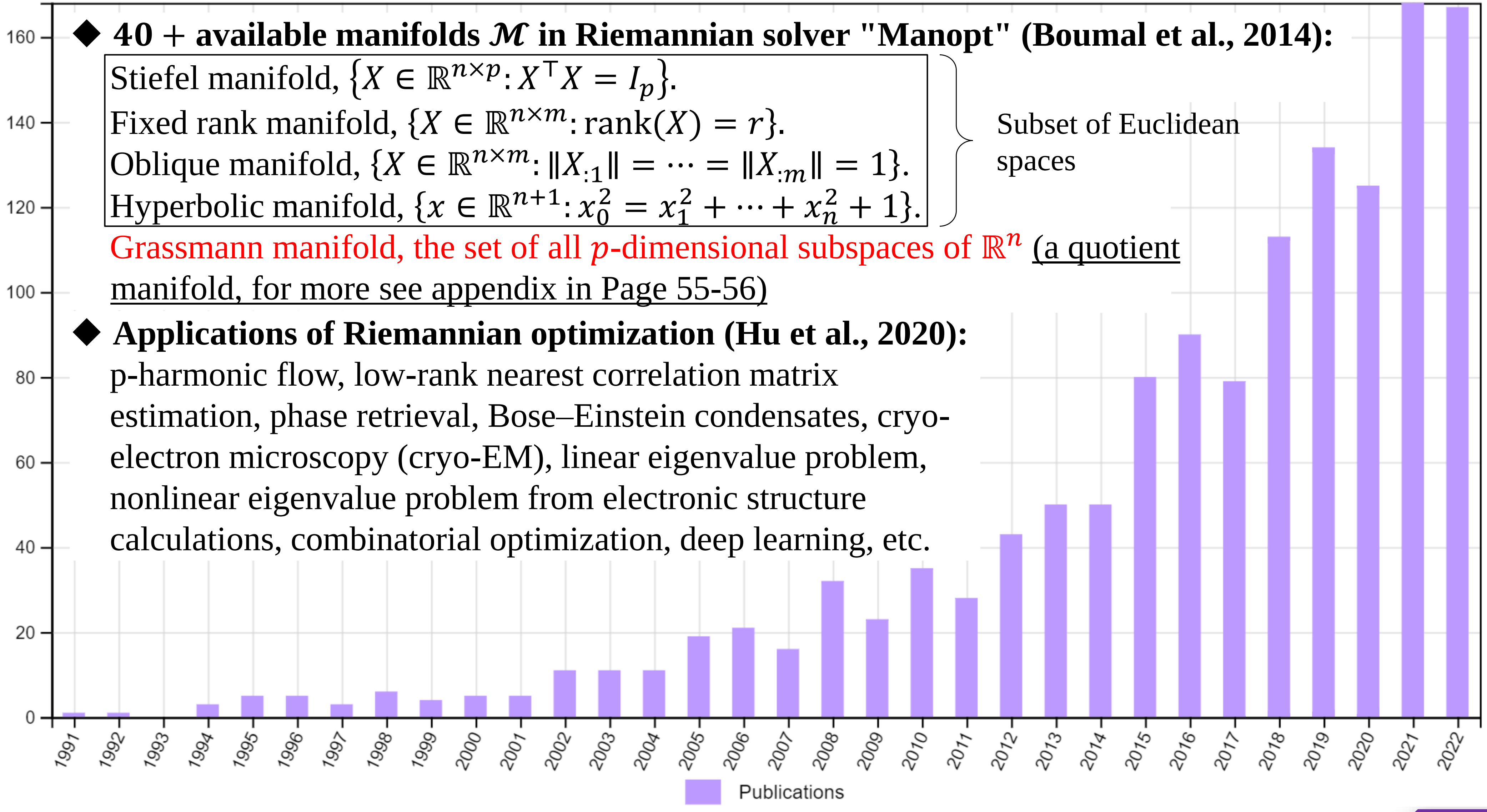
  Stiefel manifold, $\{X \in \mathbb{R}^{n \times p} : X^\top X = I_p\}$.
  Fixed rank manifold, $\{X \in \mathbb{R}^{n \times m} : \text{rank}(X) = r\}$.
  Oblique manifold, $\{X \in \mathbb{R}^{n \times m} : \|X_{:1}\| = \cdots = \|X_{:m}\| = 1\}$.
  Hyperbolic manifold, $\{x \in \mathbb{R}^{n+1} : x_0^2 = x_1^2 + \cdots + x_n^2 + 1\}$.

  } Subset of Euclidean spaces

  Grassmann manifold, the set of all p-dimensional subspaces of $\mathbb{R}^n$ (a quotient manifold, for more see appendix in Page 55-56)

- **Applications of Riemannian optimization (Hu et al., 2020):**

  p-harmonic flow, low-rank nearest correlation matrix estimation, phase retrieval, Bose–Einstein condensates, cryo-electron microscopy (cryo-EM), linear eigenvalue problem, nonlinear eigenvalue problem from electronic structure calculations, combinatorial optimization, deep learning, etc.



Source: Web of Science Core Collection. Topic : Riemannian optimization. Publication Year: 1990-01-01 to 2022-12-31.

New Challenges and Variant I: Nonsmooth Riemannian Optimization (NRO)

Background

Preliminaries

Proposal - I

Proposal - II

Conclusions

Variant I

Unconstrained Riemannian Optimization (URO)	If f is nonsmooth →	Nonsmooth Riemannian Optimization (NRO)
$\min \ f(x)$ s.t. $x \in \mathcal{M}$		$\min \ f(x)$ s.t. $x \in \mathcal{M}$

Why is (NRO) important?

Case 1: Use nonsmooth objective to replace smooth one for improving robustness.
Case 2: Add nonsmooth regularization terms to objective function.
Case 3: Treat the additional constraints as exact penalty terms.

E.g., Robust Low-Rank Matrix Completion:

$$\min_{X \in \mathcal{M}} \|\mathcal{P}_\Omega(X - A)\|_F^2 \quad \Longrightarrow \quad \min_{X \in \mathcal{M}} \|\mathcal{P}_\Omega(X - A)\|_1$$

E.g., Riemannian Exact Penalty Method (REPM) solve following subproblem at each iteration: $\min_{x \in \mathcal{M}} f(x) + \rho\left(\sum_i \max\{0, g_i(x)\} + \sum_j \left|h_j(x)\right|\right)$

Existing algorithms for (NRO) :

derivative-free techniques; subgradient techniques; smoothing techniques.

Our Contributions: We proposed general framework of Riemannian Smoothing Method!

New Challenges and Variant II: Constrained Riemannian Optimization (CRO)

Unconstrained Riemannian Optimization (URO)

$$\min \ f(x) \quad \text{s.t. } x \in \mathcal{M}$$

Add constraints →

Variant II

Constrained Riemannian Optimization (CRO)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0, i = 1, \dots, m \\ & h_j(x) = 0, j = 1, \dots, l \\ & x \in \mathcal{M} \end{aligned}$$

Resolve!

Weakness of (URO)

1. It requires the entire feasible region to form exactly one manifold.
2. Adding new constraints does not necessarily guarantee that the entire feasible region is still a manifold.
3. Even if the entire feasible region is proven to be a manifold, there are no available software packages.

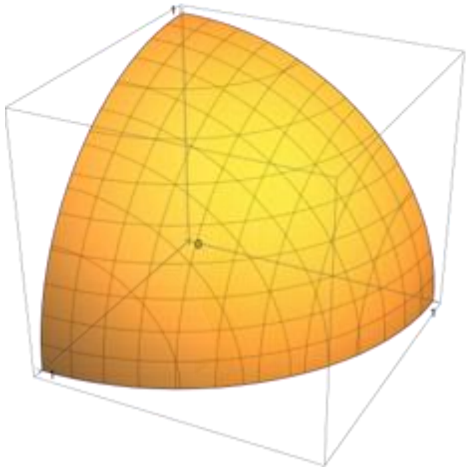
Euclidean optimization,

$$\min_{x \in \mathbb{R}^n} f(x) \ \text{ s.t. } x^\top x = 1, x \ge 0.$$

⇔

Riemannian optimization,

$$\min_{x \in \mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}} f(x) \ \text{ s.t. } x \ge 0.$$



Applications of (CRO)

- nonnegative principal component analysis (Montanari & Richard, 2016);
- orthogonal nonnegative matrix factorization (Jiang et al., 2022);
- minimum balanced cut for graph bisection (Liu & Boumal, 2020);
- subproblem of k-indicators model for data clustering (Chen et al., 2019); etc.

New Challenges and Variant II: Constrained Riemannian Optimization (CRO)

Unconstrained Riemannian Optimization (URO)
$$\min \ f(x) \quad \text{s.t. } x \in \mathcal{M}$$

Add constraints →

Variant II

Constrained Riemannian Optimization (CRO)
$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0, i = 1, \dots, m \\ & h_j(x) = 0, j = 1, \dots, l \\ & x \in \mathcal{M} \end{aligned}$$

Data Collaboration Analysis: (Nosaka & Yoshise, 2023) created collaborative data representation by solving

$$\min_{G_i \in M_i} \frac{1}{2} \sum_{i \neq i'}^{N} \|\tilde{A}_i G_i - \tilde{A}_{i'} G_{i'}\|_F^2$$

where $M_i = \{G \in \mathbb{R}^{\hat{m}_i \times \hat{m}} \mid G \text{ is full rank}\}$ and data matrices $\tilde{A}_i \in \mathbb{R}^{r \times \hat{m}_i}$. Recently, to improve performance, Nosaka is trying to solve the new model:

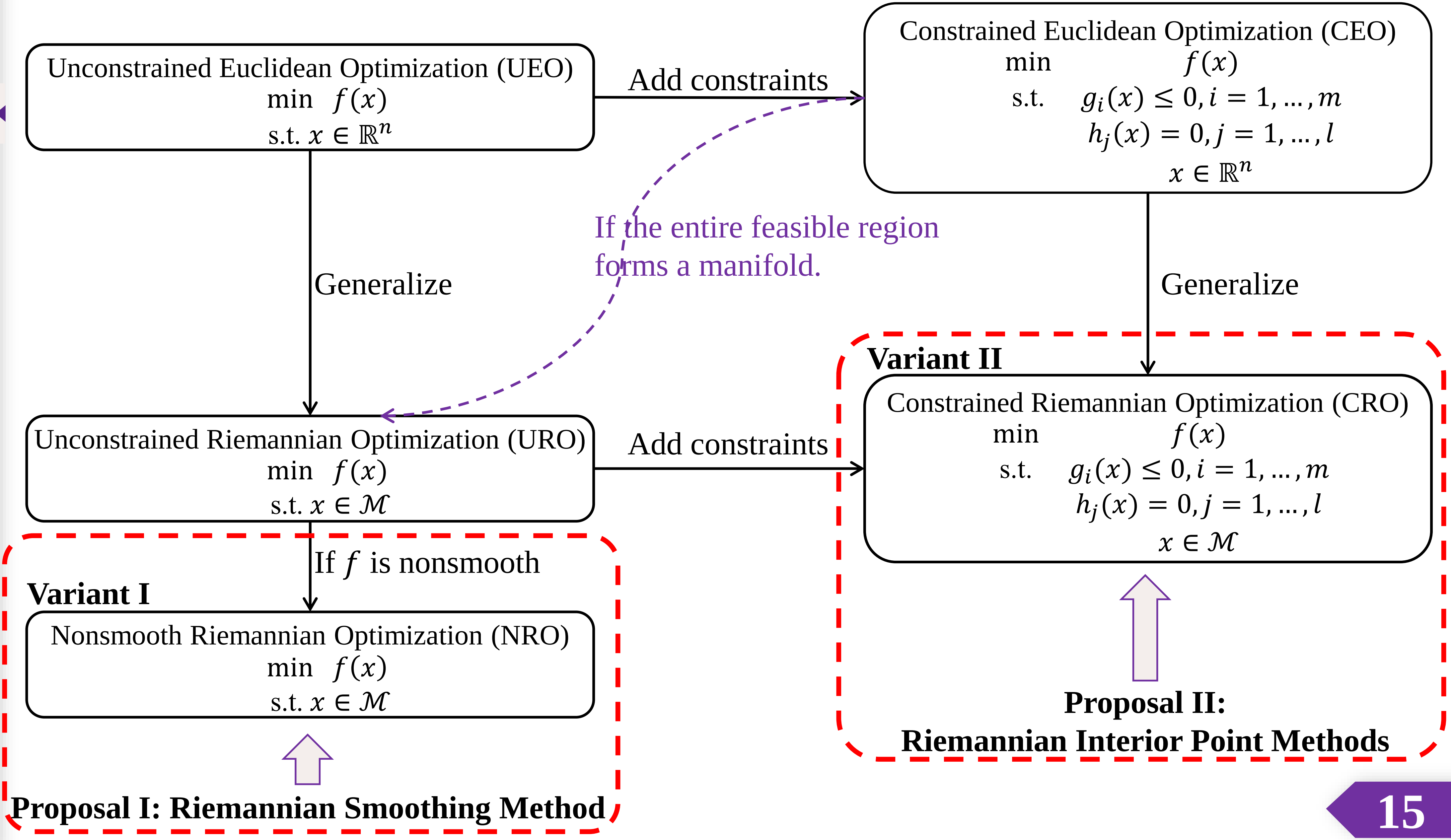
$$\min_{G_i \in M_i} \frac{1}{2} \sum_{i,i'=1}^{N} \|\tilde{A}_i G_i - \tilde{A}_{i'} G_{i'}\|_F^2 \quad \text{s.t. } \|G_i\|_F^2 - \hat{m} = 0, \forall i \in N.$$

Algorithmic research on (CRO) is still in its infancy — Started in 2020:

- Riemannian Augmented Lagrangian Method (Liu & Boumal, 2020)
- Riemannian Exact Penalty Method (Liu & Boumal, 2020)
- Riemannian Sequential Quadratic Programming Method (Schiela & Ortiz, 2021, Obara et al., 2022)

Our Contributions: We proposed Riemannian Interior Point Methods!

Summary: Position of Our Works



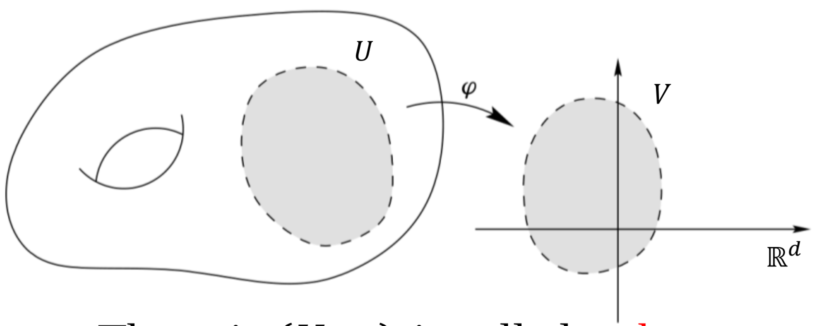
• Section 2

Preliminaries of Riemannian Optimization

What is the Manifold? (Strict definitions)

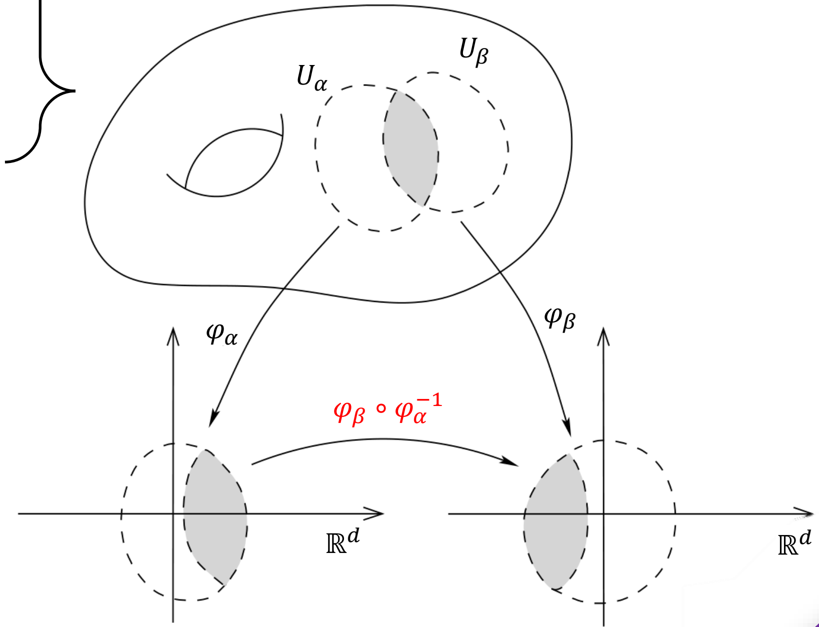
A ***d*-dimensional (smooth) manifold** is a topological space $\mathcal{M}$ satisfying the following three properties:

1. $\mathcal{M}$ is second-countable and Hausdorff.
2. $\mathcal{M}$ is locally Euclidean of dimension d (i.e., each point of $\mathcal{M}$ has a neighborhood U and a homeomorphism $\varphi: U \to V$ from U to an open set V in $\mathbb{R}^d$).



The pair (U, φ) is called a chart.

3. there is a family $\{(U_\lambda, \varphi_\lambda)\}_{\lambda \in \Lambda}$ with $\mathcal{M} = \bigcup_{\lambda \in \Lambda} U_\lambda$ such that for any $\alpha, \beta \in \Lambda$ with $U_\alpha \cap U_\beta \neq \emptyset$, the coordinate transformation
$$\varphi_\beta \circ \varphi_\alpha^{-1}: \varphi_\alpha(U_\alpha \cap U_\beta) \subseteq \mathbb{R}^d \to \varphi_\beta(U_\alpha \cap U_\beta) \subseteq \mathbb{R}^d$$
is of class C^∞.



Make the consistent smoothness across all charts.

$$\underline{f \circ \varphi_\alpha^{-1}} = \left(\underline{f \circ \varphi_\beta^{-1}}\right) \circ \left(\varphi_\beta \circ \varphi_\alpha^{-1}\right).$$

Make sense.

A function $f: \mathcal{M} \to \mathbb{R}$ is smooth at $p \in \mathcal{M}$ if there exists a chart (U, φ) such that $f \circ \varphi^{-1}$ is of class C^∞ at $\varphi(p)$.

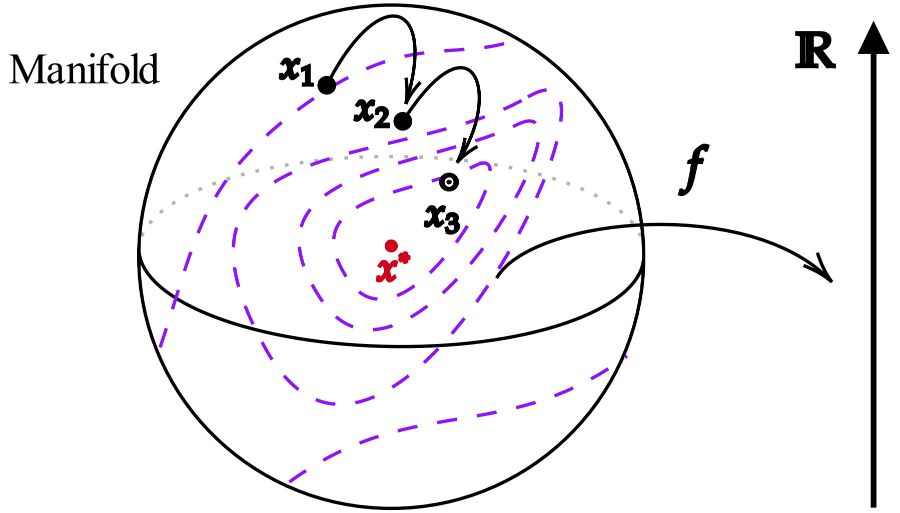
Source: (Lee, 2012) Fig. 1.2 and Fig. 1.6.

How to Optimize a Function on Manifold?

Optimization problem on a manifold $\mathcal{M}$:

$$\min \quad f(x)$$
$$\text{s.t. } x \in \mathcal{M}$$

where $f: \mathcal{M} \to \mathbb{R}$.



Goal:
To find a local optimal solution $x^* \in \mathcal{M}$.
(In general, $\mathcal{M}$ is nonconvex.)

Method:
The iterative methods can still be used.
But there are questions that we need to address:

Q1: What is the direction of movement?	we need	Tangent vector (Page 19)
Q2: What is a good direction?		Riemannian gradient (Page 20)
Q3: What is the optimal condition?		Singularity of gradient vector field (Page 21)
Q4: How to move on manifolds?		Using retraction to create a curve (Page 22)

Q1: What is the Direction of Movement? Tangent Vector

Embedded submanifold $\mathcal{M}$ of $\mathbb{R}^n$ = manifold + subset of $\mathbb{R}^n$.

◆As the velocity of curve:

Consider a curve $\gamma: I \subseteq \mathbb{R} \to \mathcal{M}$ passing through point x with $\gamma(0) = x$. Then

$$\gamma'(0) := \lim_{t \to 0} \frac{\gamma(t) - \gamma(0)}{t} = \frac{d}{dt}\gamma(t)\Big|_{t=0}$$

is a tangent vector at point x.

⇩ But in general, the $\gamma(t) - \gamma(0)$ is not defined.

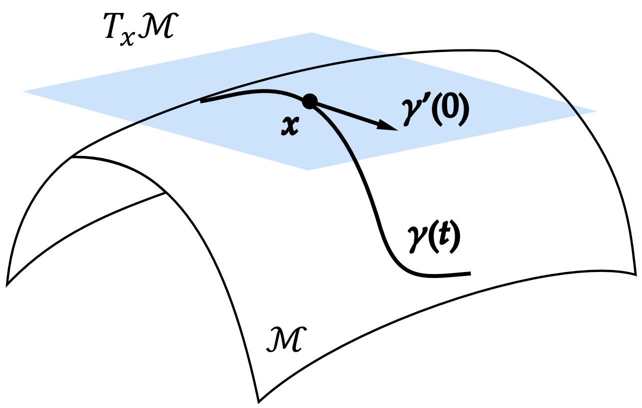
◆As the differential operator:

$\mathfrak{F}_x(\mathcal{M})$: the set of all smooth real-valued functions defined in a neighborhood of $x \in \mathcal{M}$.

A map $\xi: \mathfrak{F}_x(\mathcal{M}) \to \mathbb{R}$ is called a tangent vector at x on $\mathcal{M}$ if there exists a curve $\gamma: I \subseteq \mathbb{R} \to \mathcal{M}$ such that $\gamma(0) = x$ and

$$\xi f = \frac{d}{dt} f(\gamma(t))\Big|_{t=0}, \ \forall f \in \mathfrak{F}_x(\mathcal{M}).$$

We often write $\xi \equiv \gamma'(0)$. The tangent space, $T_x\mathcal{M}$, is the set of all possible tangent vectors at that point.



Notice:

① $T_x\mathcal{M}$ are linear spaces sharing the same dimension.

② In general, $T_x\mathcal{M}$ is determined by x, except for $\mathbb{R}^n \cong T_x\mathbb{R}^n$.

③ For embedded $\mathcal{M}$, $T_x\mathcal{M}$ is a subspace of $\mathbb{R}^n$.

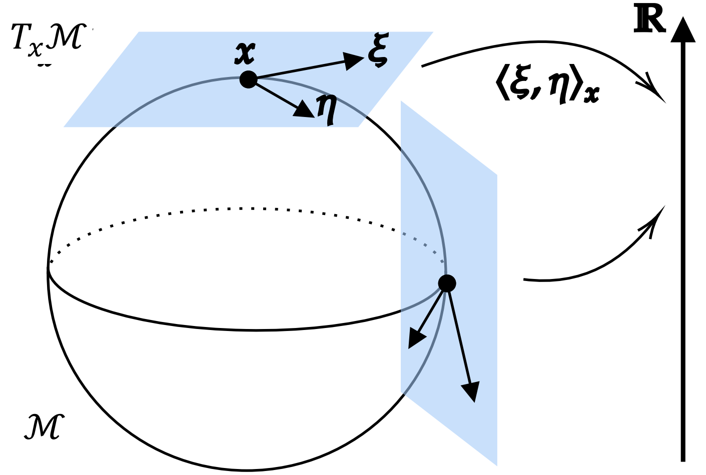
E.g., $T_x\mathbb{S}^{n-1} = \{u \in \mathbb{R}^n : x^\top u = 0\}$.

Q2: What is a Good Direction? Riemannian Gradient

Definition (Riemannian manifold)
A Riemannian metric $\langle\cdot,\cdot\rangle$ assigns an inner product

$$\langle\cdot,\cdot\rangle_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$$

to each tangent space of the manifold in a way that varies smoothly from point to point.
Then, $(\mathcal{M}, \langle\cdot,\cdot\rangle)$ is called a Riemannian manifold.

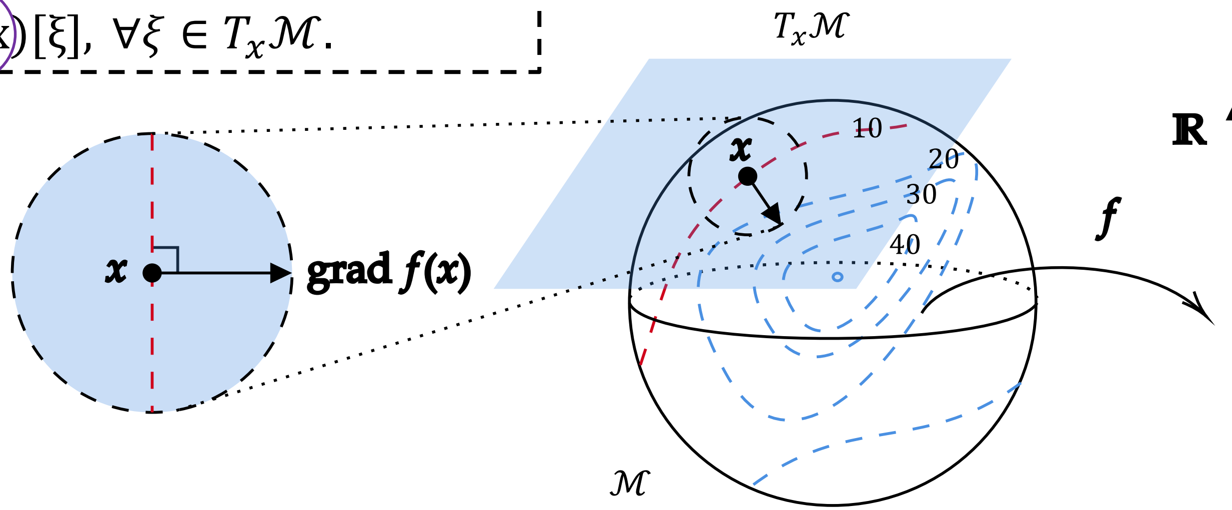


Definition (Riemannian gradient)
Riemannian gradient of a function $f: \mathcal{M} \to \mathbb{R}$ at a point $x \in \mathcal{M}$, denoted as $\mathrm{grad} f(x)$, is a unique element* in $T_x\mathcal{M}$ that satisfies

$$\langle \mathrm{grad} f(x), \xi\rangle_x = \mathrm{D}f(x)[\xi],\ \forall \xi \in T_x\mathcal{M}.$$

Assume $\mathrm{D}f(x): T_x\mathcal{M} \to \mathbb{R}$ is well-defined.
(See appendix in Page 58)

grad $f(x)$ is the direction of fastest increase at x.



(*) Riesz Theorem: For an inner product space $(V, \langle\cdot,\cdot\rangle)$, if $T: V \to \mathbb{R}$ is a linear, then there is a unique $y \in V$ such that $T(x) = \langle v, x\rangle$ for all $x \in V$.

Q3: What is the Optimal Condition? Singularity of Gradient Vector Field

Recall that $T\mathcal{M} = \{(x, v): x \in \mathcal{M} \text{ and } v \in T_x\mathcal{M}\}$ is called the tangent bundle of $\mathcal{M}$.

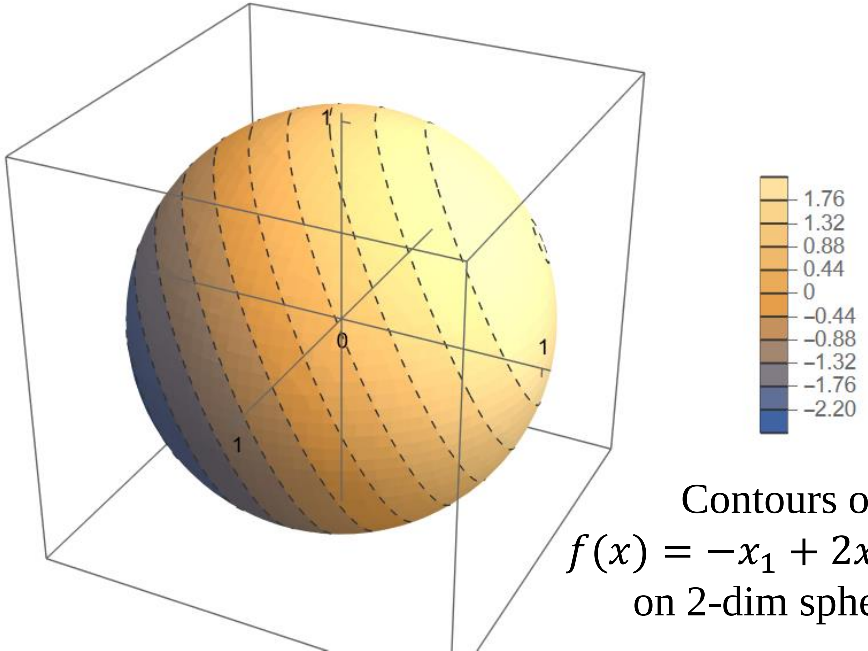
Definition (vector field)
A vector field on $\mathcal{M}$ is a map $V: \mathcal{M} \to T\mathcal{M}$ such that $V(x) \in T_x\mathcal{M}$ for all $x \in \mathcal{M}$.

Riemannian gradient,

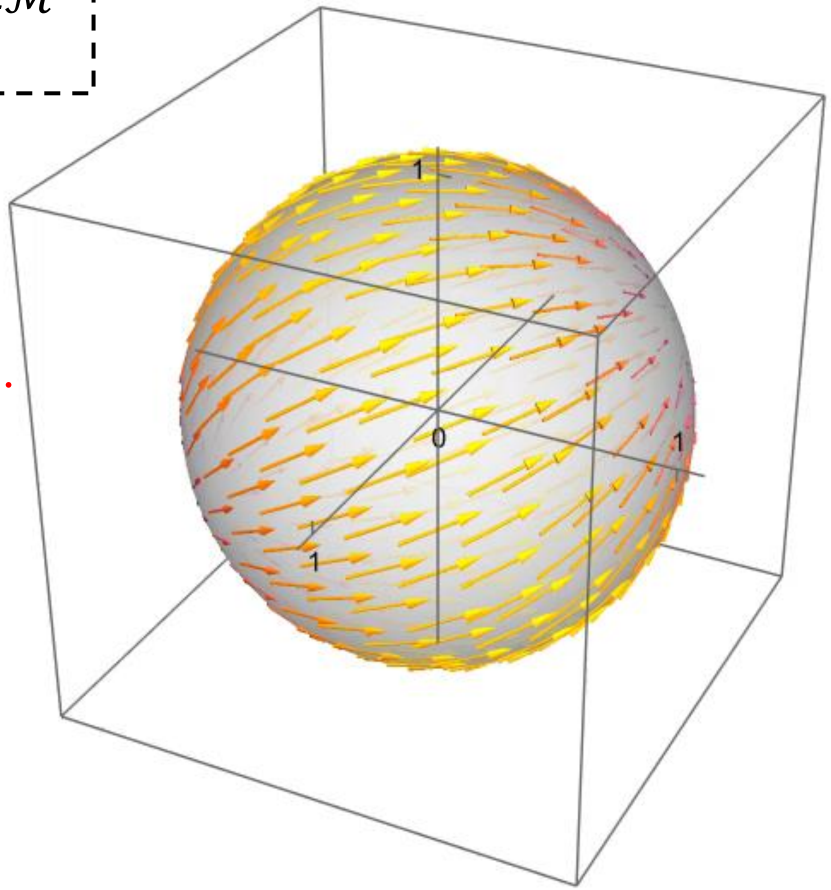
$$x \mapsto \operatorname{grad} f(x),$$

is a special vector field generated by a real-valued function f.

- $-\operatorname{grad} f(x)$ is the direction of steepest descent at x.
- If x^* is a local minimizer/maximizer, then $\operatorname{grad} f(x^*) = 0_{x^*}$.



Contours of
$f(x) = -x_1 + 2x_2 + x_3$
on 2-dim sphere.



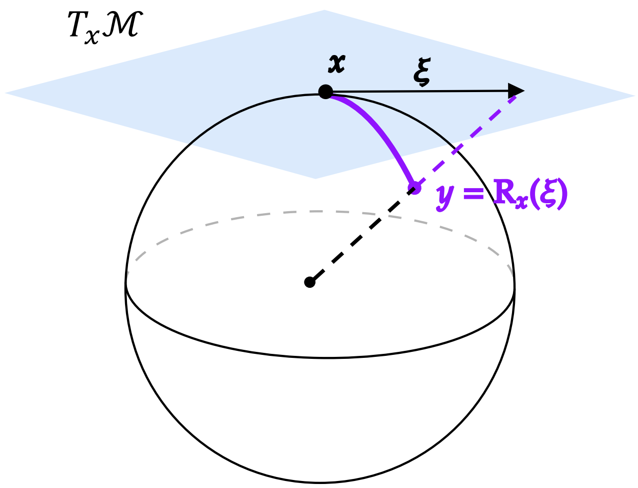
Riemannian gradient field of
$f(x) = -x_1 + 2x_2 + x_3$
on 2-dim sphere.

Q4: How to Move on Manifolds? Using Retraction to Create a Curve

Definition (Retraction). A retraction on a manifold $\mathcal{M}$ is a smooth map

$$\mathrm{R}: T\mathcal{M} \to \mathcal{M}: (x, \xi) \mapsto \mathrm{R}_x(\xi)$$

such that for each $(x, \xi) \in T\mathcal{M}$ the curve $\gamma(t) := \mathrm{R}_x(t\xi)$ has $\gamma'(0) = \xi$.



Retractions are not uniquely determined.
E.g., on the unit sphere $\mathbb{S}^{n-1}$,

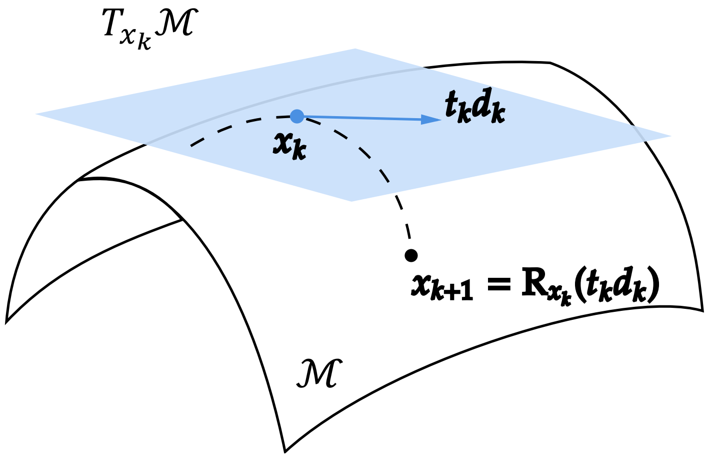
$$\mathrm{R}_x(\xi) = \frac{x + \xi}{\| x + \xi \|}, \text{ or } \mathrm{R}_x(\xi) = \cos(\| \xi \|)x + \frac{\sin(\| \xi \|)}{\| \xi \|}\xi.$$

General Line Search Framework for solve $\min_{x \in \mathcal{M}} f(x)$.

Choose an initial point $x_0 \in \mathcal{M}$, a retraction R, and $k \leftarrow 0$;

while $\|\mathrm{grad} f(x_k)\|_{x_k}$ is not close to 0 **do:**

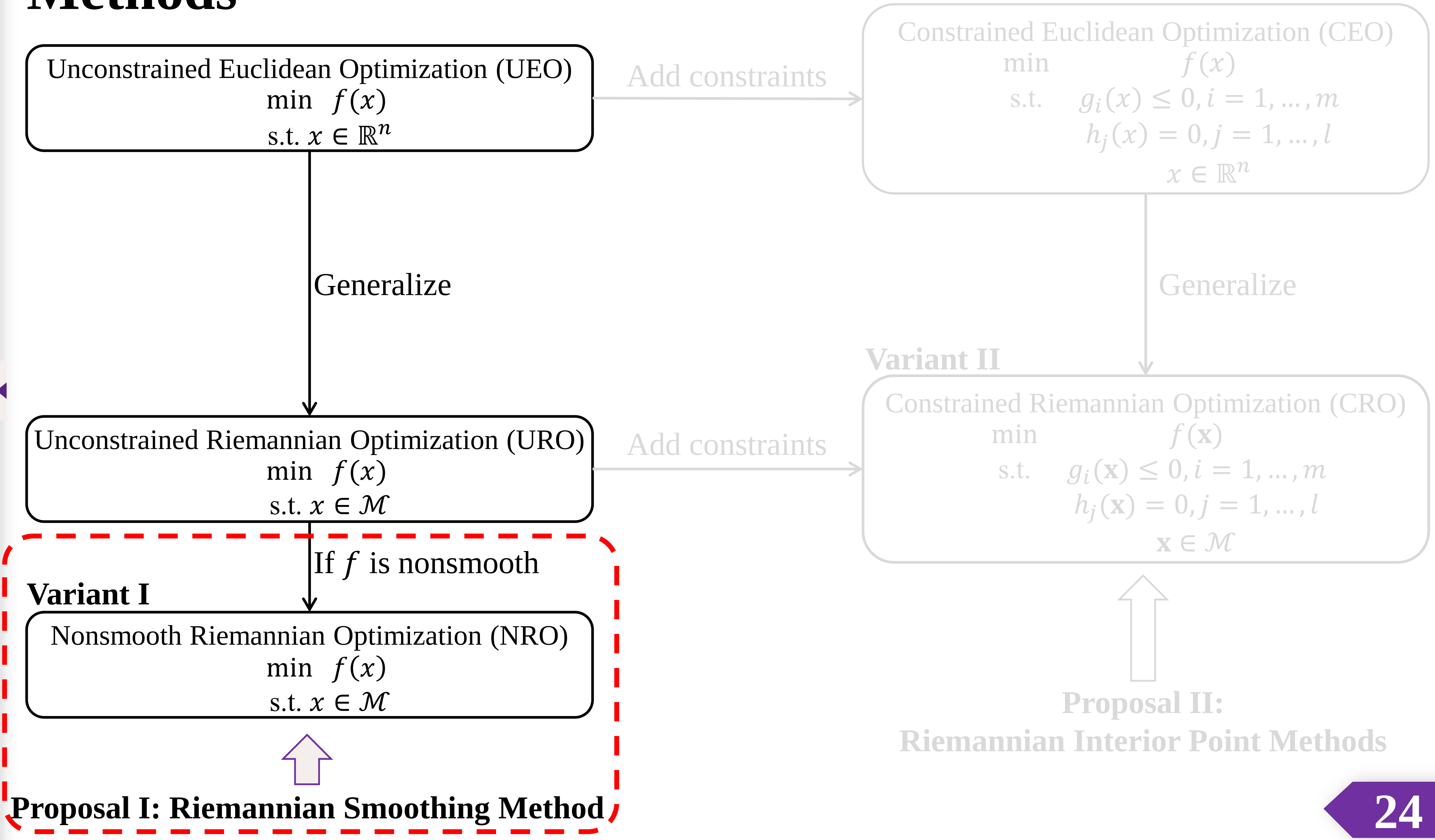
1. Compute a direction $d_k \in T_{x_k}\mathcal{M}$, e.g., $d_k = -\mathrm{grad} f(x)$;
2. Compute a step length $t_k > 0$, e.g., Armijo condition;
3. Compute the next point $x_{k+1} := \mathrm{R}_{x_k}(t_k d_k)$;
4. Set $k \leftarrow k + 1$;



- Section 3

Our Proposal I
- Riemannian Smoothing Methods

Position of Our Proposal I - Riemannian Smoothing Methods



Smoothing Function

Variant I

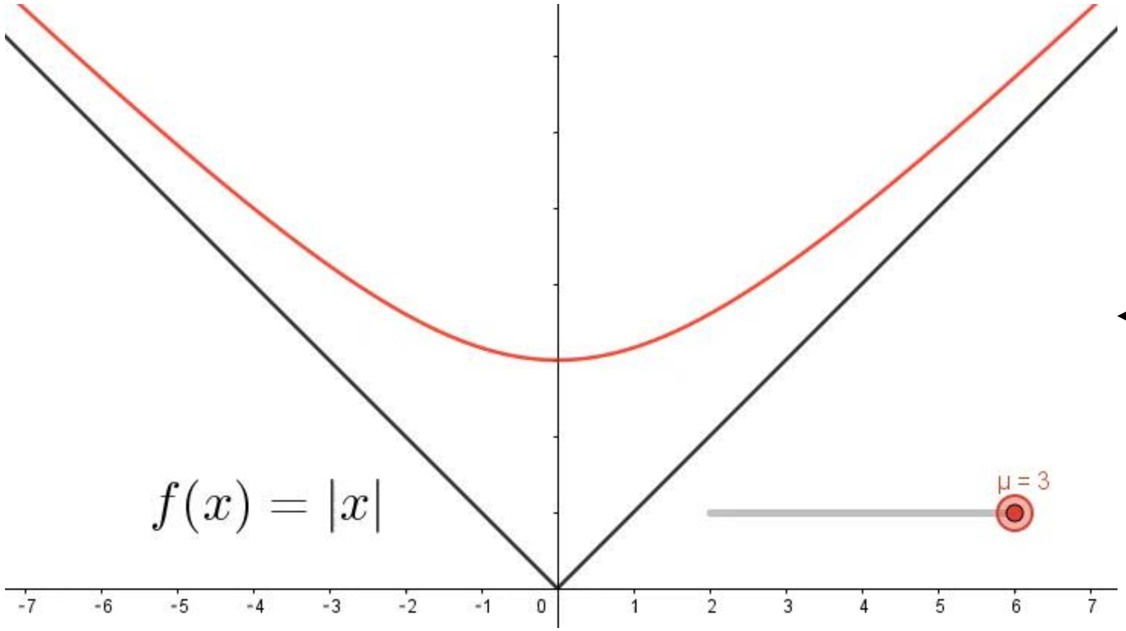
Unconstrained Riemannian Optimization (URO)
$\min \; f(x)$
s.t. $x \in \mathcal{M}$

→ If f is nonsmooth →

Nonsmooth Riemannian Optimization (NRO)
$\min \; f(x)$
s.t. $x \in \mathcal{M}$

Choose a smoothing function $\tilde{f}$ for the nonsmooth f such that

$$\lim_{\mu \to 0^+} |\tilde{f}(x,\mu) - f(x)| = 0 \text{ for any } x.$$



$$\tilde{f}_1(x,\mu) = \begin{cases} |x| & \text{if } |x| > \frac{\mu}{2} \\ \frac{x^2}{\mu} + \frac{\mu}{4} & \text{if } |x| \le \frac{\mu}{2}, \end{cases}$$

$$\tilde{f}_2(x,\mu) = \sqrt{\mu^2 + x^2},$$

$$\tilde{f}_3(x,\mu) = 2\mu \log\left(1 + \exp\left(\frac{x}{\mu}\right)\right) - x,$$

$$\tilde{f}_4(x,\mu) = x \tanh\left(\frac{x}{\mu}\right).$$

General Riemannian Smoothing Method (RSM)

Variant I

Unconstrained Riemannian Optimization (URO)		Nonsmooth Riemannian Optimization (NRO)
$\min\ f(x)$ s.t. $x \in \mathcal{M}$	If f is nonsmooth →	$\min\ f(x)$ s.t. $x \in \mathcal{M}$

Algorithm 1. General Riemannian Smoothing Method for (NRO)

Take $x_0 \in \mathcal{M}$ and set $k = 0$. A nonnegative sequence $\{\delta_k\} \to 0$.

Choose a smoothing function $\tilde{f}$.

Choose an arbitrary Riemannian algorithm.

Many options:
- Steepest Descent (SD)
- Conjugate Gradient (CG)
- Trust Regions (TR)

While stopping criterion not satisfied **do:**

Solve

$$x^k := \arg\min_{x \in \mathcal{M}} \tilde{f}(x, \mu_k)$$

approximately by using the chosen algorithm, starting at x^{k-1}, such that $\|\mathrm{grad}\ \tilde{f}(x^k, \mu_k)\| < \delta_k$;

Choose $0 < \mu_{k+1} < \mu_k$;

$k \leftarrow k + 1$;

End

Our contributions: The first study on a **general** smoothing framework for (NRO).
(Zhang et al., 2023) only used Steepest Descent (SD) method for subproblem.

Application: Completely Positive (CP) Factorization

CP factorization problem:
Given a completely positive matrix A.
Find $B \in \mathbb{R}^{n \times r}$ s.t. $A = BB^\top$ and $B \geq 0$.

An open problem in conic optimization theory (Berman et al., 2015).

Reformulation of CP factorization problem (Groetzner & Dür, 2020):
Given a completely positive matrix A.
Find X s.t. $\bar{B}X \geq 0$ and $X^T X = I$,
where $\bar{B}$ is an initial factorization of A.

Our proposal: Transform it into a Riemannian model:

$$\min_{X \in \mathrm{St}(r,r)} \{\max(-\bar{B}X)\},$$

where orthogonal group $\mathrm{St}(r,r) = \{X \in \mathbb{R}^{r \times r} \mid X^T X = I\}$.

Use LogSumExp $LSE_\mu(x): \mathbb{R}^n \to \mathbb{R}$,

$$LSE_\mu(x) = \mu \log\left(\sum_{i=1}^{n} \exp(x_i/\mu)\right)$$

to replace max function.

Our contributions: The first study to apply Riemannian algorithms to CP factorization problem.
The numerical experiments showed that our method is better than Euclidean methods.

Background

Preliminaries

Proposal - I

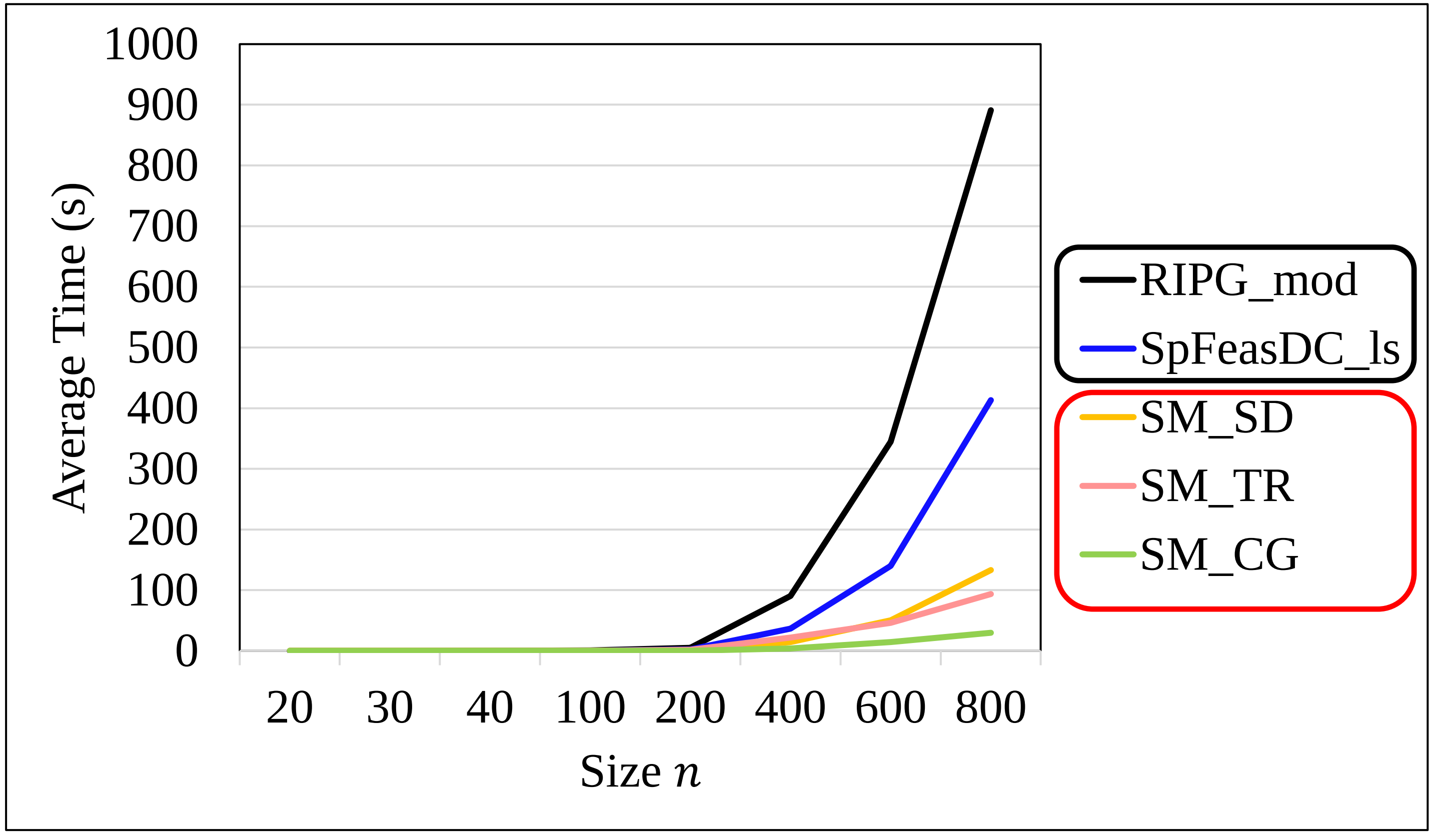
Proposal - II

Conclusions

Experiment 1 - Randomly Generated Instances

CP factorization problem:
Given a completely positive matrix A.
Find $B \in \mathbb{R}^{n\times r}$ s.t. $A = BB^\top$ and $B \geq 0$.

Experiments Settings:
Let $A := HH^T$, where $H \in \mathbb{R}^{n\times n}$ with entries randomly generated by MATLAB command `rand`.
We take $n \in \{20,30,40,100,200,400,600,800\}$ and set $r = 1.5n$.
For each n, we generated 10 instances to examine.



Euclidean methods.

Our Smoothing Methods (SM) with different sub-algorithms:
- Steepest Descent (SD)
- Trust Regions (TR)
- Conjugate Gradient (CG)

Experiment 2 - Specifical Structured Instances

Background

Preliminaries

Proposal - I

Proposal - II

Conclusions

CP factorization problem:
Given a completely positive matrix A.
Find $B \in \mathbb{R}^{n\times r}$ s.t. $A = BB^\top$ and $B \geq 0$.

Experiments Settings:
Let $\mathbf{1}_n$ denote the all-ones vector in $\mathbb{R}^n$ and consider the matrix

$$A = \begin{bmatrix} 0 & \mathbf{1}_{n-1}^T \\ \mathbf{1}_{n-1} & I_{n-1} \end{bmatrix}^T \begin{bmatrix} 0 & \mathbf{1}_{n-1}^T \\ \mathbf{1}_{n-1} & I_{n-1} \end{bmatrix}.$$

We take $n \in \{10,20,50,75,100,150\}$ and set $r = n$.
For each n, we generated 50 starting points to examine.

Our Smoothing Methods (SM) | Euclidean Methods

Size n	SM_SD	SM_CG	SM_TR	SpFeasDC_ls	RIPG_mod	APM_mod
10	**1**	**1**	**1**	**1**	**1**	0.8
20	**1**	**1**	**1**	0.98	0.74	0.9
50	**1**	**1**	**1**	0.98	0	0.76
75	**1**	**1**	**1**	0.98	0	0.64
100	**1**	**1**	**1**	0.8	0	0.6
150	**1**	**1**	**1**	0.7	0	0.35

Success Rate = 1 in each row.

The above table shows that in all cases our methods are always successful; whereas the Success Rates of the Euclidean methods decreased as n increased.

• Section 4

Our Proposal II
- Riemannian Interior Point Methods

Position of Our Proposal II - Riemannian Interior Point Methods

Unconstrained Euclidean Optimization (UEO)
min $f(\mathbf{x})$
s.t. $\mathbf{x} \in \mathbb{R}^n$

Add constraints

Constrained Euclidean Optimization (CEO)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, \dots, m \\ & h_j(x) = 0, j = 1, \dots, l \\ & x \in \mathbb{R}^n \end{aligned}$$

Generalize

Generalize

Unconstrained Riemannian Optimization (URO)

Add constraints

But before that, we need
2nd order geometry on manifolds.

If f is nonsmooth

Variant I

Nonsmooth Riemannian Optimization (NRO)
min $f(\mathbf{x})$
s.t. $\mathbf{x} \in \mathcal{M}$

Proposal I: Riemannian Smoothing Method

Variant II

Constrained Riemannian Optimization (CRO)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, \dots, m \\ & h_j(x) = 0, j = 1, \dots, l \\ & x \in \mathcal{M} \end{aligned}$$

Proposal II:
Riemannian Interior Point Methods

2nd Order Geometry: Differentiating Vector Fields

Background

Preliminaries

Proposal - I

Proposal - II

Conclusions

Recall that:

- $\mathfrak{X}(\mathcal{M})$ denotes the set of all smooth vector fields on $\mathcal{M}$.
- $T\mathcal{M} = \{(x, v): x \in \mathcal{M} \text{ and } v \in T_x\mathcal{M}\}$ is tangent bundle.

Definition (Riemannian connection ∇)
A Riemannian connection on $\mathcal{M}$ is the **unique** function

$$\nabla: T\mathcal{M} \times \mathfrak{X}(\mathcal{M}) \to T\mathcal{M}, (u, V) \mapsto \nabla_u V$$

such that $\nabla_u V \in T_x\mathcal{M}$ whenever $u \in T_x\mathcal{M}$ and satisfies other six conditions*.

Definition (covariant derivative of V at $x \in \mathcal{M}$)
The covariant derivative of $V \in \mathfrak{X}(\mathcal{M})$ at x is a linear operator defined as

$$\nabla V(x): T_x\mathcal{M} \to T_x\mathcal{M}, u \mapsto \nabla_u V.$$

Generalized Jacobian

When $V = \operatorname{grad} f$.

Definition (Riemannian Hessian of f at x)
Given a smooth function $f: \mathcal{M} \to \mathbb{R}$.

$$\operatorname{Hess} f(x) \triangleq \nabla \operatorname{grad} f(x): T_x\mathcal{M} \to T_x\mathcal{M}$$

is called Riemannian Hessian of f at $x \in \mathcal{M}$. (self-adjoint!)

Generalized Hessian

(*): Smoothness; Linearity in u; Linearity in V; Leibniz rule; Symmetry; Compatibility.

2nd Order Geometry: Riemannian Newton Method

Definition (Singularity)
Let $F: \mathcal{M} \to T\mathcal{M}$ be a smooth vector field. A point $p \in \mathcal{M}$ is call singularity of F if

$$F(p) = 0_p \in T_p\mathcal{M}$$

where 0_p is the zero element of $T_p\mathcal{M}$.

Recall: the optimal condition of
$$\min_{x \in \mathcal{M}} f(x)$$
is
$$\text{grad}\, f(x^*) = 0_{x^*}.$$

Algorithm 2 Riemannian Newton method

Goal: To find the singularity of the given vector field F.

Take $x_0 \in \mathcal{M}$ and set $k = 0$.

While stopping criterion not satisfied **do:**

Solve the Newton equation

$$\nabla F(x_k) v_k = -F(x_k),$$

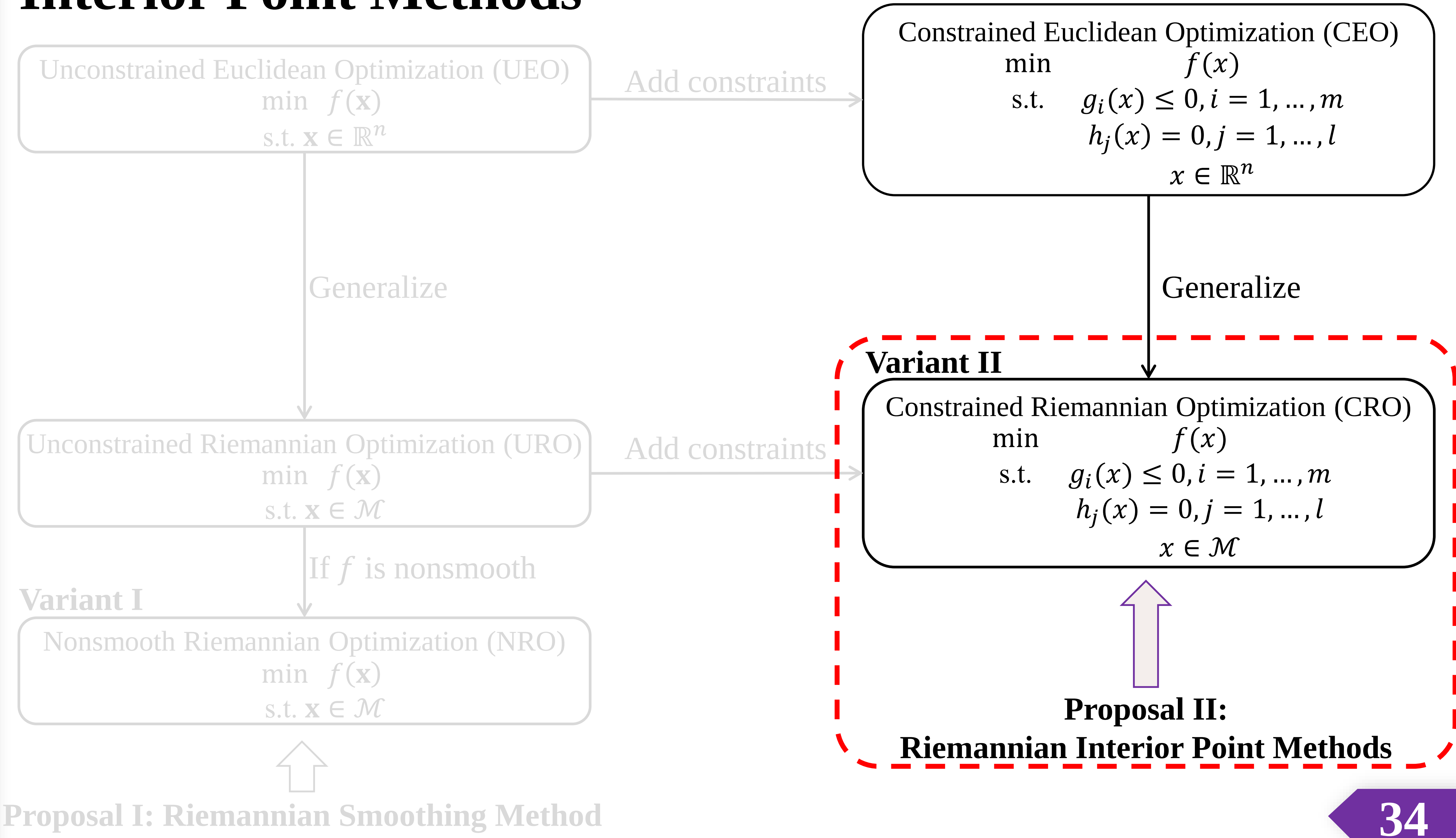
Update $x_{k+1} := \mathrm{R}_{x_k}(v_k)$;

$k \leftarrow k + 1$;

End

- It is a natural extension of the famous Newton method.
- Well-known convergence: the local superlinear/quadratic convergence also hold. (See appendix in Page 64)

We are Ready to the Our Proposal II - Riemannian Interior Point Methods



A New Concept: KKT Vector Field

Constrained Riemannian Optimization (CRO)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, \dots, m \\ & h_j(x) = 0, j = 1, \dots, l \\ & x \in \mathcal{M}, \end{aligned}$$

where $f: \mathcal{M} \to \mathbb{R}, h: \mathcal{M} \to \mathbb{R}^l$, and $g: \mathcal{M} \to \mathbb{R}^m$.

Lagrangian function of (CRO) is

$$\mathcal{L}(x, y, z) := f(x) + \sum_{j=1}^{l} y_j h_j(x) + \sum_{i=1}^{m} z_i g_i(x),$$

where $y \in \mathbb{R}^l$ and $z \in \mathbb{R}^m$ are Lagrange multipliers. Then we have

$$\text{grad}_x \mathcal{L}(x, y, z) = \text{grad} f(x) + \sum_{i=1}^{l} y_i \text{grad} h_i(x) + \sum_{i=1}^{m} z_i \text{grad} g_i(x),$$

$$\text{Hess}_x \mathcal{L}(x, y, z) = \text{Hess} f(x) + \sum_{i=1}^{l} y_i \text{Hess} h_i(x) + \sum_{i=1}^{m} z_i \text{Hess} g_i(x).$$

First order optimal condition (Yang et al., 2014)

If x is a local minimizer of (CRO) and Linear Independence Constraint Qualification (LICQ) holds at x, then x satisfies Riemannian KKT conditions:

$$\begin{cases} \text{grad}_x \mathcal{L}(x, y, z) & = 0_x, \\ h(x) & = 0, \\ g(x) & \leq 0, \\ Zg(x) & = 0, \\ z & \geq 0. \end{cases}$$

Using $s := -g(x)$, the KKT conditions becomes

$$F(w) \triangleq \begin{pmatrix} \text{grad}_x \mathcal{L}(x, y, z) \\ h(x) \\ g(x) + s \\ ZSe \end{pmatrix} = 0 := \begin{pmatrix} 0_x \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{ and } (z, s) \geq 0,$$

$$Z = \begin{pmatrix} z_1 & & \\ & \ddots & \\ & & z_n \end{pmatrix}$$

where $w := (x, y, z, s) \in \bar{\mathcal{M}} \triangleq \mathcal{M} \times \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^m$. We called F the KKT Vector Field defined on $\bar{\mathcal{M}}$ with $T_w \bar{\mathcal{M}} \equiv T_x \mathcal{M} \times \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^m$.

Our goal: Just to find singularity w such that $F(w) = 0$, and $(z, s) \geq 0$.
Newton method is a good idea, but we need $\nabla F(w)$!

Covariant Derivative of KKT Vector Field

Constrained Riemannian Optimization (CRO)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0, i = 1, \dots, m \\ & h_j(x) = 0, j = 1, \dots, l \\ & x \in \mathcal{M}, \end{aligned}$$

where $f: \mathcal{M} \to \mathbb{R}, h: \mathcal{M} \to \mathbb{R}^l$, and $g: \mathcal{M} \to \mathbb{R}^m$.

Lagrangian function of (CRO) is

$$\mathcal{L}(x, y, z) := f(x) + \sum_{j=1}^{l} y_j h_j(x) + \sum_{i=1}^{m} z_i g_i(x),$$

where $y \in \mathbb{R}^l$ and $z \in \mathbb{R}^m$ are Lagrange multipliers. Then we have

$$\mathrm{grad}_x \mathcal{L}(x, y, z) = \mathrm{grad} f(x) + \sum_{i=1}^{l} y_i \mathrm{grad} h_i(x) + \sum_{i=1}^{m} z_i \mathrm{grad} g_i(x),$$

$$\mathrm{Hess}_x \mathcal{L}(x, y, z) = \mathrm{Hess} f(x) + \sum_{i=1}^{l} y_i \mathrm{Hess} h_i(x) + \sum_{i=1}^{m} z_i \mathrm{Hess} g_i(x).$$

For each $x \in \mathcal{M}$, we define

$$H_x: \mathbb{R}^l \to T_x\mathcal{M}, \; H_x v \triangleq \sum_i v_i \mathrm{grad} h_i(x).$$

Hence, the adjoint operator is

$$H_x^*: T_x\mathcal{M} \to \mathbb{R}^l, \; H_x^* \xi = [\langle \mathrm{grad} h_1(x), \xi \rangle_x, \cdots, \langle \mathrm{grad} h_l(x), \xi \rangle_x]^T.$$

Similarly, there are G_x, G_x^*.

Then, the covariant derivative of KKT vector field is a linear operator $\nabla F(w): T_w\bar{\mathcal{M}} \to T_w\bar{\mathcal{M}}$ is given by

$$\nabla F(w)\Delta w = \begin{pmatrix} \mathrm{Hess}_x \mathcal{L}(w) \Delta x + H_x \Delta y + G_x \Delta z \\ H_x^* \Delta x \\ G_x^* \Delta x + \Delta s \\ Z \Delta s + S \Delta z \end{pmatrix}$$

where $\Delta w = (\Delta x, \Delta y, \Delta s, \Delta z) \in T_x\mathcal{M} \times \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^m \equiv T_w\bar{\mathcal{M}}$.

Riemannian Interior Point Method (RIPM)

Algorithm 3 Prototype Algorithm of RIPM for (CRO)

Goal: To find singularity $w^* \in \bar{\mathcal{M}}$ such that $F(w^*) = 0, (z^*, s^*) \geq 0$.

Given initial w_0 with $(z_0, s_0) > 0$, barrier parameter $\mu_0 > 0$;

While stopping criterion not satisfied **do:**

Solve the perturbed Newton equation

$$\nabla F(w_k)\Delta w_k = -F(w_k) + \mu_k \hat{e},$$

where $\hat{e} \triangleq (0_x, 0, 0, e)$;

Compute the step sizes α_k such that $(z_{k+1}, s_{k+1}) > 0$;

Update $w_{k+1} = \bar{\mathrm{R}}_{w_k}(\alpha_k \Delta w_k)$;

Choose $0 < \mu_{k+1} < \mu_k$;

$k \leftarrow k + 1$;

End

Theorem (Local Convergence, L. 2022)

Under some standard assumptions.

(1) If $\mu_k = o(\|F(w_k)\|), \alpha_k \to 1$, then $\{w_k\}$ locally, superlinearly converges to w^*.

(2) If $\mu_k = O\left(\|F(w_k)\|^2\right), 1 - \alpha_k = O(\|F(w_k)\|)$, then $\{w_k\}$ locally, quadratically converges to w^*.

Next, we proposal a global convergent RIPM.

Global Algorithm for RIPM

① **Merit function**: Choose $\varphi(w) \triangleq \| F(w) \|^2$.

② **Backtracking for step size α_k** :

(1) Centrality conditions (See appendix in Page 74).

(2) Sufficient decreasing condition:

Let $\varphi(\alpha) \triangleq \varphi(\underbrace{\bar{R}_{w_k}(\alpha \Delta w_k)}_{\text{new iterate}})$ for fixed w_k and Δw_k, then $\varphi(0) = \varphi(w_k) =: \varphi_k$ and $\varphi'(0) = \langle \mathrm{grad}\varphi(w_k), \Delta w_k \rangle$. Sufficient decreasing asks $\varphi(\alpha_k) - \varphi(0) \leq \alpha_k \beta \varphi'(0)$.

How to ensure the descent direction?

Let Δw_k be the solution of $\nabla F(w_k) \Delta w_k = -F(w_k) + \rho_k \sigma_k \hat{e}$, then $\varphi'(0) < 0$ if we set $\rho_k := s_k^T z_k / m$, $\sigma_k \in (0,1)$. Then, $\{\varphi_k\}$ is monotonically decreasing.

Assumptions:

1. the functions $f(x), h(x), g(x)$ are smooth; the set $\{\mathrm{grad} h_i(x)\}_{i=1}^{l}$ is linearly independent in $T_x\mathcal{M}$ for all x; and $w \mapsto \nabla F(w)$ is Lipschitz continuous;
2. the sequences $\{x_k\}$ and $\{z_k\}$ are bounded;
3. the operator $\nabla F(w)$ is nonsingular.

In context of nonconvex!

Theorem (Global Convergence, L. 2022)
If Assumptions 1 ~ 3 hold, then $\{F(w_k)\}$ converges to zero; and for any limit point $w^* = (x^*, y^*, z^*, s^*)$ of $\{w_k\}$, x^* is a Riemannian KKT point of problem (CRO).

c.f. (El-Bakry et al., 1996)

Background

Preliminaries

Proposal - I

Proposal - II

Conclusions

Implementation: Condensed Form of Newton Equation

Dominant cost of RIPM is to solve **Newton equation:** $\nabla F(w)\Delta w = -F(w) + \mu\hat{e}$.
That is the following linear equation on $T_x\mathcal{M} \times \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^m$:

$$F(w) = \begin{pmatrix} F_x \\ F_y \\ F_z \\ F_s \end{pmatrix}, \quad \hat{e} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ e \end{pmatrix}.$$

$$\begin{pmatrix} \mathrm{Hess}_x \mathcal{L}(w)\Delta x + H_x \Delta y + G_x \Delta z \\ H_x^* \Delta x \\ G_x^* \Delta x + \Delta s \\ Z\Delta s + S\Delta z \end{pmatrix} = \begin{pmatrix} -F_x \\ -F_y \\ -F_z \\ -F_s + \mu e \end{pmatrix}.$$

Using two substitutions $\Delta s = Z^{-1}(\mu e - F_s - S\Delta z)$,
$\Delta z = S^{-1}[Z(G_x^*\Delta x + F_z) + \mu e - F_s]$ from 3rd and 4th rows.

It suffices to focus on **condensed form** on $T_x\mathcal{M} \times \mathbb{R}^l$:

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix},$$

where

$$\mathcal{A}_w := \mathrm{Hess}_x \mathcal{L}(w) + G_x S^{-1} Z G_x^*, \qquad c := -F_x - G_x S^{-1}(ZF_z + \mu e - F_s), \; q := -F_y.$$

$\mathcal{T}$ is self-adjoint (but may indefinite) operator on $T_x\mathcal{M} \times \mathbb{R}^l$.
The difficulty lies in...

- the Riemannian setting leaves us with no explicit matrix form available.
- a natural way is to find the representing matrix $\hat{\mathcal{T}}$ under some basis of tangent space. (Expensive!)

An ideal approach is to use iterative methods, e.g., Krylov subspace methods

Implementation: Krylov Subspace Methods on Tangent Space

For simplicity, we consider the case of **only inequality** constraints, where Δy vanishes and only a linear operator equation (Let $\mathcal{A} \equiv \mathcal{A}_w : T_x\mathcal{M} \to T_x\mathcal{M}$):

$$\mathcal{A}\Delta x = c. \quad \text{(OpEquation)}$$

- Krylov subspace method only needs to call $v \mapsto \mathcal{A}v$ once at each iteration.
- All the iterates v_k are in $T_x\mathcal{M}$.
- Since $\mathcal{A}$ is self-adjoint but indefinite, we use Conjugate Residual (CR) method to solve it.

Algorithm 4: Conjugate Residual (CR) Method on Tangent Spaces for (OpEquation)

Input: Sysmetric invertible linear operator $\mathcal{A} : T_x\mathcal{M} \to T_x\mathcal{M}$, nonzero $c \in T_x\mathcal{M}$ and an initial point $v_0 \in T_x\mathcal{M}$.

Output: Sequence $\{v_n\} \subset T_x\mathcal{M}$ such that $\{v_n\} \to v^*$ and $\mathcal{A}v^* = c$.

Set $n \leftarrow 0$, $r_0 := c - \mathcal{A}v_0, p_0 := r_0$ and compute $\mathcal{A}r_0, \mathcal{A}p_0$;

while *stopping criterion not satisfied* **do**

- Update number $\alpha_n := \langle r_n, \mathcal{A}r_n\rangle_x / \langle \mathcal{A}p_n, \mathcal{A}p_n\rangle_x$; `// step length`
- $v_{n+1} := v_n + \alpha_n p_n$; `// Iterate point`
- $r_{n+1} := r_n - \alpha_n \mathcal{A}p_n$; `// Residual`
- Compute $\mathcal{A}r_{n+1}$; `// This is the only call to` $\mathcal{A}$ `in while loop`
- Update number $\beta_n := \langle r_{n+1}, \mathcal{A}r_{n+1}\rangle_x / \langle r_n, \mathcal{A}r_n\rangle_x$;
- $p_{n+1} := r_{n+1} + \beta_n p_n$; `// Conjugate direction`
- $\mathcal{A}p_{n+1} := \mathcal{A}r_{n+1} + \beta_n \mathcal{A}p_n$; `// No need to call` $\mathcal{A}$ `here`
- $n \leftarrow n + 1$;

end

Comparison: Riemannian IPM (RIPM) is Better than Euclidean IPM (EIPM)

Constrained Riemannian Optimization (CRO)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, \ldots, m \\ & h_j(x) = 0, j = 1, \ldots, l \\ & x \in \mathcal{M}, \end{aligned}$$

where $f: \mathcal{M} \to \mathbb{R}, h: \mathcal{M} \to \mathbb{R}^l$, and $g: \mathcal{M} \to \mathbb{R}^m$.

← Generalize

Constrained Euclidean Optimization (CEO)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, \ldots, m \\ & h_j(x) = 0, j = 1, \ldots, l \\ & x \in \mathbb{R}^n, \end{aligned}$$

where $f: \mathbb{R}^n \to \mathbb{R}, h: \mathbb{R}^n \to \mathbb{R}^l$, and $g: \mathbb{R}^n \to \mathbb{R}^m$.

Our Proposal II:
Riemannian Interior Point Methods (**RIPM**)

Euclidean Interior Point Methods (**EIPM**)

- RIPM keeps the all advantages of Riemannian optimization.
- EIPM is a special case of RIPM when $\mathcal{M} \equiv \mathbb{R}^n$ or $\mathbb{R}^{n \times k}$.
- RIPM can solve some problems that EIPM cannot. → E.g., rank $(X) = r$ is not continuous, we can not apply EIPM.
- RIPM solves condensed Newton equation of smaller order on $T_x\mathcal{M} \times \mathbb{R}^l$ (See appendix in Page 67):

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}.$$

Numerical Experiments

Environment: MATLAB R2022a on a computer equipped with an Intel Core i7-10700 at 2.90G Hz with 16 GB of RAM.

We compare with the other Riemannian methods:

- RALM : Riemannian augmented Lagrangian method. (Liu & Boumal, 2020)
- REPM(LQH) : Riemannian exact penalty method with smoothing function LQH. (Liu & Boumal, 2020)
- REPM(LSE) : Riemannian exact penalty method with smoothing function LSE. (Liu & Boumal, 2020)
- RSQP : Riemannian sequential quadratic programming. (Obara et al., 2022)
- RIPM (Our method): Riemannian interior point method.

KKT residual is defined by

$$\sqrt{\|\mathrm{grad}_x \mathcal{L}(w)\|^2 + \sum_{i=1}^{m} \{\min(0, z_i)^2 + \max(0, g_i(x))^2 + |z_i g_i(x)|^2\} + \sum_{j=1}^{l} |h_j(x)|^2}.$$

Code: https://github.com/GALVINLAI/RIPM

Problem I — Nonnegative Low Rank Matrix Approximation (m varies under fixed n, r)

Problem I (Song & Ng, 2020)

$$\min_{X \in \mathbb{R}_r^{m \times n}} \| A - X \|_F^2 \quad \text{s.t. } X \geq 0,$$

where $\mathbb{R}_r^{m \times n} = \{X \in \mathbb{R}^{m \times n} : \text{rank}\,(X) = r\}$.

Experiments settings:
Fix $n = 20,\ r = 2$; we take $m \in \{8,16,24,32\}$. For each m, we generated 20 random instances* A to examine.
Each experiment stopped successfully if solution with KKT residual $< 10^{-8}$ was found before the maximum time 10 (s) was reached.

	Row Num. m	RALM	REPM (LQH)	REPM (LSE)	RSQP	RIPM
Success Rate	8	0.1	0.2	0	**1**	**1**
	16	0.05	0.25	0.1	0.85	**1**
	24	0	0.3	0.15	0.35	**1**
	32	0.05	0.3	0.25	0	**1**
Average Time (s)	8	0.68	**0.28**	-	2.57	**0.28**
	16	1.13	**0.48**	2.58	7.14	**0.68**
	24	-	**0.70**	3.89	10.12	**0.96**
	32	2.37	**0.95**	5.05	-	**1.63**

Success Rate = 1 in each row.

The first two fastest results in each row.

The results are similar when we vary the value of only one of m, n, r, so they are omitted here. (See appendix in Page 65-66 for more results)

(*): Let $\text{B} = \text{rand}(\text{m}, \text{r}); \text{C} = \text{rand}(\text{r}, \text{n}); \text{A} = \text{B} * \text{C} + 0.001 * \text{randn}(\text{m}, \text{n})$.

Background

Preliminaries

Proposal - I

Proposal - II

Conclusions

Problem I — Nonnegative Low Rank Matrix Approximation (Impacts of parameters m, n, r in RIPM)

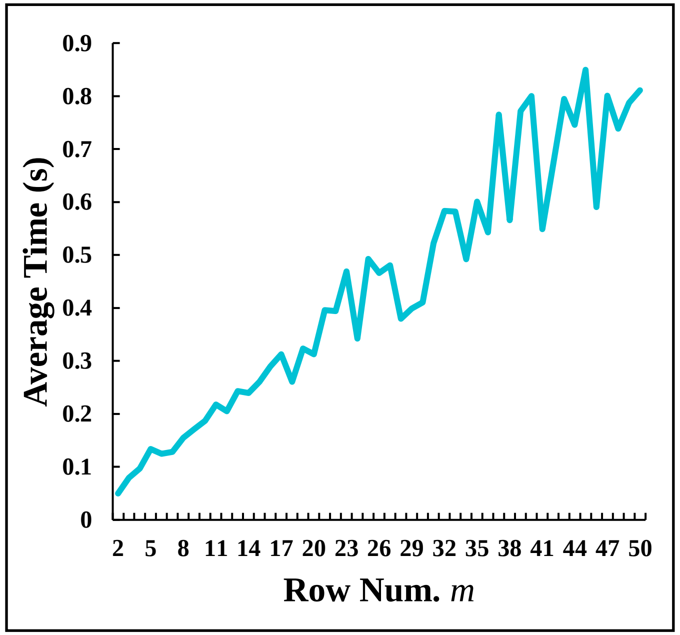
Problem I (Song & Ng, 2020)

$$\min_{X \in \mathbb{R}_r^{m \times n}} \| A - X \|_F^2 \quad \text{s.t. } X \geq 0,$$

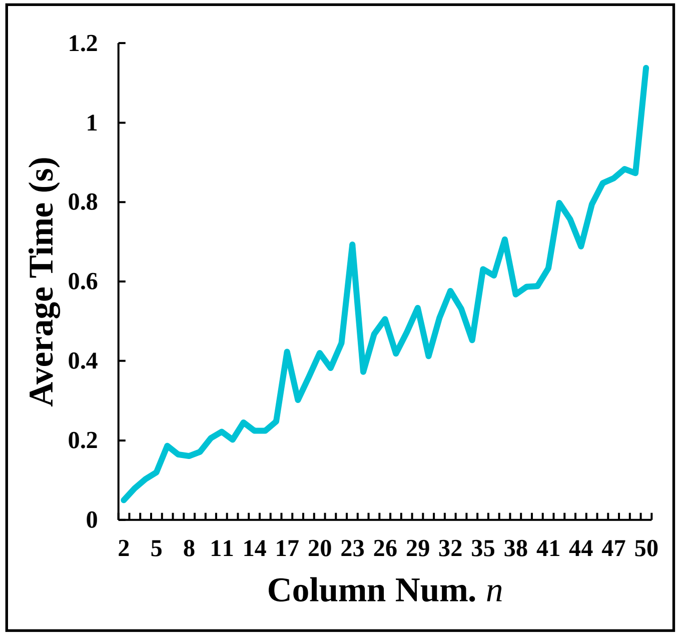
where $\mathbb{R}_r^{m \times n} = \{X \in \mathbb{R}^{m \times n} : \text{rank}(X) = r\}$.

Experiments settings:
For each case, we generated 10 random instances* to examine.
Each experiment stopped successfully if solution with KKT residual $< 10^{-6}$ was found before the maximum time 5 (s) was reached.

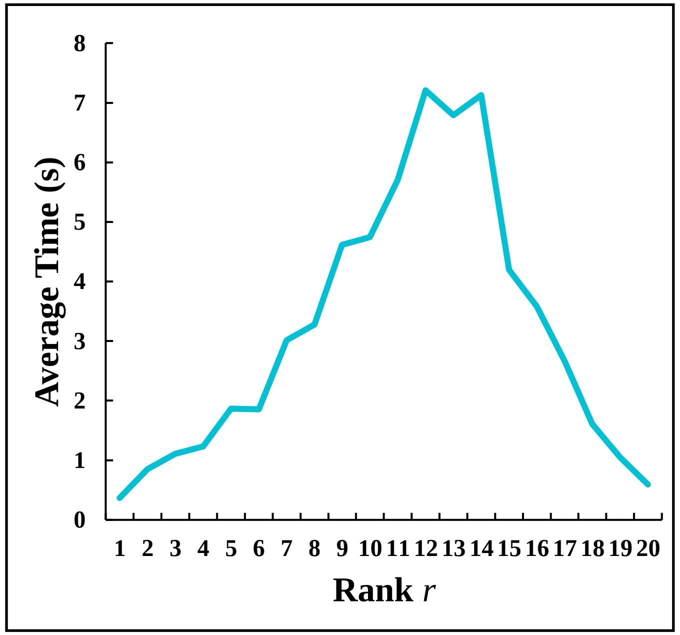
Fix $n = 10,\ r = 2.$



Fix $m = 10,\ r = 2.$



Fix $m = 20,\ n = 20.$



(*): Let $B = \text{rand}(m, r); C = \text{rand}(r, n); A = B * C + 0.001 * \text{randn}(m, n)$.

Problem II — Projection onto Nonnegative Stiefel Manifold

Problem II (Jiang et al., 2022) Given $C \in \mathbb{R}^{n\times k}$, we consider

$$\min_{X\in \mathrm{St}(n,k)} \| X - C \|_F^2, \text{ s.t. } X \geq 0, \qquad \textbf{(Model\_Stiefel)}$$

which can be equivalently reformulated into

$$\min_{X\in \mathrm{OB}(n,k)} \| X - C \|_F^2 \text{ s.t. } X \geq 0, \text{ and } \| XV \|_F = 1, \qquad \textbf{(Model\_Oblique)}$$

- Stiefel manifold, $\mathrm{St}(n,k) \triangleq \{X \in \mathbb{R}^{n\times k}: X^\top X = I\}$
- Oblique manifold, $\mathrm{OB}(n,k) \triangleq \{X \in \mathbb{R}^{n\times k}$: all columns have unit norm $\}$
- V is arbitrary satisfying $\| V \|_F = 1$ and $VV^\top > 0$.

Experiments settings:
For each model, we tested 20 random trials. It stopped successfully if solution with KKT residual $< 10^{-6}$ was found before the maximum time 600 (s) was reached.

	Size (n,k)	RALM	REPM (LQH)	REPM (LSE)	RSQP	RIPM
Success Rate	(60,12)	**1**	0	0	0.65	**1**
	(70,14)	**1**	0	0	0.85	**1**
Average Time (s)	(60,12)	**4.10**	-	-	78.02	**5.56**
	(70,14)	**6.23**	-	-	166.1	**7.57**

Results for (**Model_Stiefel**)

	Size (n,k)	RALM	REPM (LQH)	REPM (LSE)	RSQP	RIPM
Success Rate	(60,12)	0.6	0	0	0.7	**1**
	(70,14)	0.6	0	0	0.5	**1**
Average Time (s)	(60,12)	**5.73**	-	-	44.46	**7.13**
	(70,14)	**8.22**	-	-	91.38	**9.27**

Results for (**Model_Oblique**)

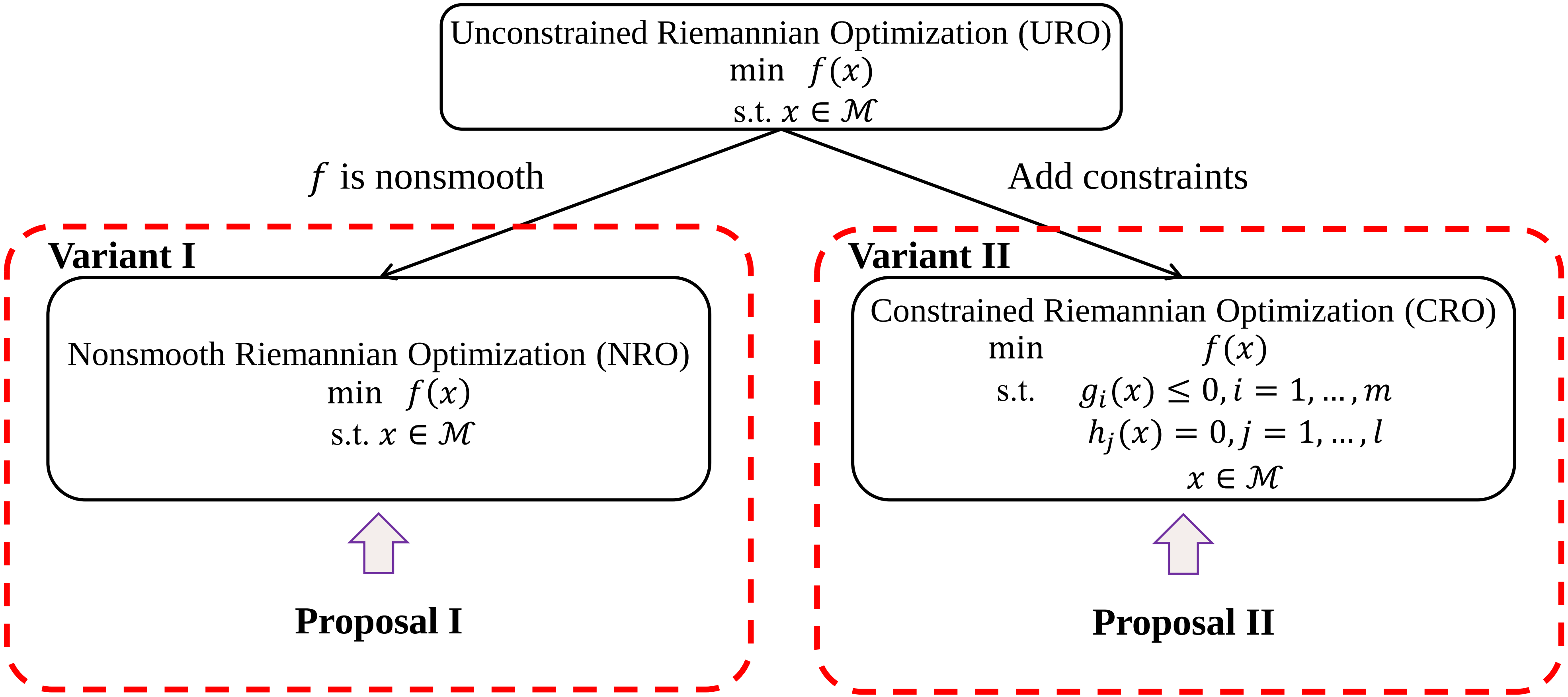
Success Rate = 1 in each row.

The first two fastest results in each row.

- Section 5

Conclusion

Summary of Our Works



Our contributions:

1. **Proposal I :** A general framework of Riemannian Smoothing Method
2. **Proposal II:** Riemannian version of the interior point method
   - ➢ We proved the local superlinear/quadratic and global convergence.
   - ➢ We established some foundational concepts, such as the KKT vector field.

Future Works

Background

Preliminaries

Proposal - I

Proposal - II

Conclusions

Finally, we discuss 3 promising future topics about Riemannian Interior Point Method (RIPM).

① Preconditioner for linear operator equation.

Due to complementary condition, as $k \to \infty$, the values of

$$S_k^{-1}Z_k = \begin{pmatrix} \frac{(z_k)_1}{(s_k)_1} \to 0 & & \\ & \ddots & \\ & & \frac{(z_k)_n}{(s_k)_n} \to \infty \end{pmatrix}$$

display a huge difference in magnitude.

Condensed form on $T_x\mathcal{M} \times \mathbb{R}^l$:

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix},$$

where

$$\mathcal{A}_w := Hess_x \mathcal{L}(w) + \Theta,$$

Hence, the operator $\Theta := G_x S^{-1} Z G_x^*$ in the condensed system (Above) makes it ill-conditioned, so the iterative method will likely fail unless it is carefully preconditioned.

No matrix form available!

② Treatment of more state-of-the-art interior point methods.

Our current global algorithm uses the simplest strategy. How about, e.g., the trust region?

③ Quasi-Newton RIPM

The quasi-Newton RIPM can approximate the Hessian of Lagrangian in $\nabla F(w_k)$ with gradient information while ensuring its local convergence. (See appendix in Page 68)

Thank you for your attention!
Questions?

Zhijian Lai

2024/01/22

PhD Thesis Final Defense

References

Berman, A., Dur, M., & Shaked-Monderer, N. (2015). Open problems in the theory of completely positive and copositive matrices. *The Electronic Journal of Linear Algebra*, *29*, 46–58. https://doi.org/10.13001/1081-3810.2943

Boţ, R. I., & Nguyen, D.-K. (2021). Factorization of completely positive matrices using iterative projected gradient steps. *Numerical Linear Algebra with Applications*, *28*(6), e2391. https://doi.org/10.1002/nla.2391

Boumal, N. (2023). *An Introduction to Optimization on Smooth Manifolds (2023)* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009166164

Boumal, N., Mishra, B., Absil, P.-A., & Sepulchre, R. (2014). Manopt, a Matlab Toolbox for Optimization on Manifolds. *Journal of Machine Learning Research*, *15*(42), 1455–1459.

Chen, C., Pong, T. K., Tan, L., & Zeng, L. (2020). A difference-of-convex approach for split feasibility with applications to matrix factorizations and outlier detection. *Journal of Global Optimization*, *78*(1), 107–136. https://doi.org/10.1007/s10898-020-00899-8

Chen, F., Yang, Y., Xu, L., Zhang, T., & Zhang, Y. (2019). *Big-Data Clustering: K-Means or K-Indicators?* (arXiv:1906.00938). arXiv. https://doi.org/10.48550/arXiv.1906.00938

El-Bakry, A. S., Tapia, R. A., Tsuchiya, T., & Zhang, Y. (1996). *On the formulation and theory of the Newton interior-point method for nonlinear programming*. *89*(3), 35.

Fernandes, T. A., Ferreira, O. P., & Yuan, J. (2017). On the Superlinear Convergence of Newton's Method on Riemannian Manifolds. *Journal of Optimization Theory and Applications*, *173*(3), 828–843. https://doi.org/10.1007/s10957-017-1107-2

Ferreira, O. P., & Silva, R. C. M. (2012). Local convergence of Newton's method under a majorant condition in Riemannian manifolds. *IMA Journal of Numerical Analysis*, *32*(4), 1696–1713. https://doi.org/10.1093/imanum/drr048

Groetzner, P., & Dür, M. (2020). A factorization method for completely positive matrices. *Linear Algebra and Its Applications*, *591*, 1–24. https://doi.org/10.1016/j.laa.2019.12.024

References

Hu, J., Liu, X., Wen, Z.-W., & Yuan, Y.-X. (2020). A Brief Introduction to Manifold Optimization. *Journal of the Operations Research Society of China*, *8*(2), 199–248. https://doi.org/10.1007/s40305-020-00295-9

Jiang, B., Meng, X., Wen, Z., & Chen, X. (2022). An exact penalty approach for optimization with nonnegative orthogonality constraints. *Mathematical Programming*. https://doi.org/10.1007/s10107-022-01794-8

Lai, Z., & Yoshise, A. (2022). Completely positive factorization by a Riemannian smoothing method. *Computational Optimization and Applications*. https://doi.org/10.1007/s10589-022-00417-4

Lee, J. M. (2012). *Introduction to Smooth Manifolds—2nd ed. 2012*. https://math.berkeley.edu/~jchaidez/materials/reu/lee_smooth_manifolds.pdf

Liu, C., & Boumal, N. (2020). Simple Algorithms for Optimization on Riemannian Manifolds with Constraints. *Applied Mathematics & Optimization*, *82*(3), 949–981. https://doi.org/10.1007/s00245-019-09564-3

Montanari, A., & Richard, E. (2016). Non-Negative Principal Component Analysis: Message Passing Algorithms and Sharp Asymptotics. *IEEE Transactions on Information Theory*, *62*(3), 1458–1484. https://doi.org/10.1109/TIT.2015.2457942

Nosaka, K., & Yoshise, A. (2023). Creating Collaborative Data Representations Using Matrix Manifold Optimal Computation and Automated Hyperparameter Tuning. *2023 IEEE 3rd International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB)*, 180–185. https://doi.org/10.1109/ICEIB57887.2023.10170466

Obara, M., Okuno, T., & Takeda, A. (2022). Sequential Quadratic Optimization for Nonlinear Optimization Problems on Riemannian Manifolds. *SIAM Journal on Optimization*, *32*(2), 822–853. https://doi.org/10.1137/20M1370173

Sato, H. (2021). *Riemannian Optimization and Its Applications*. Springer International Publishing. https://doi.org/10.1007/978-3-030-62391-3

Schiela, A., & Ortiz, J. (2021). An SQP Method for Equality Constrained Optimization on Hilbert Manifolds. *SIAM Journal on Optimization*, *31*(3), 2255–2284. https://doi.org/10.1137/20M1341325

References

Song, G.-J., & Ng, M. K. (2020). Nonnegative low rank matrix approximation for nonnegative matrices. *Applied Mathematics Letters*, *105*, 106300. https://doi.org/10.1016/j.aml.2020.106300

Yamashita, H., & Yabe, H. (1996). Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization. *Mathematical Programming*, *75*(3), 377–397. https://doi.org/10.1007/BF02592190

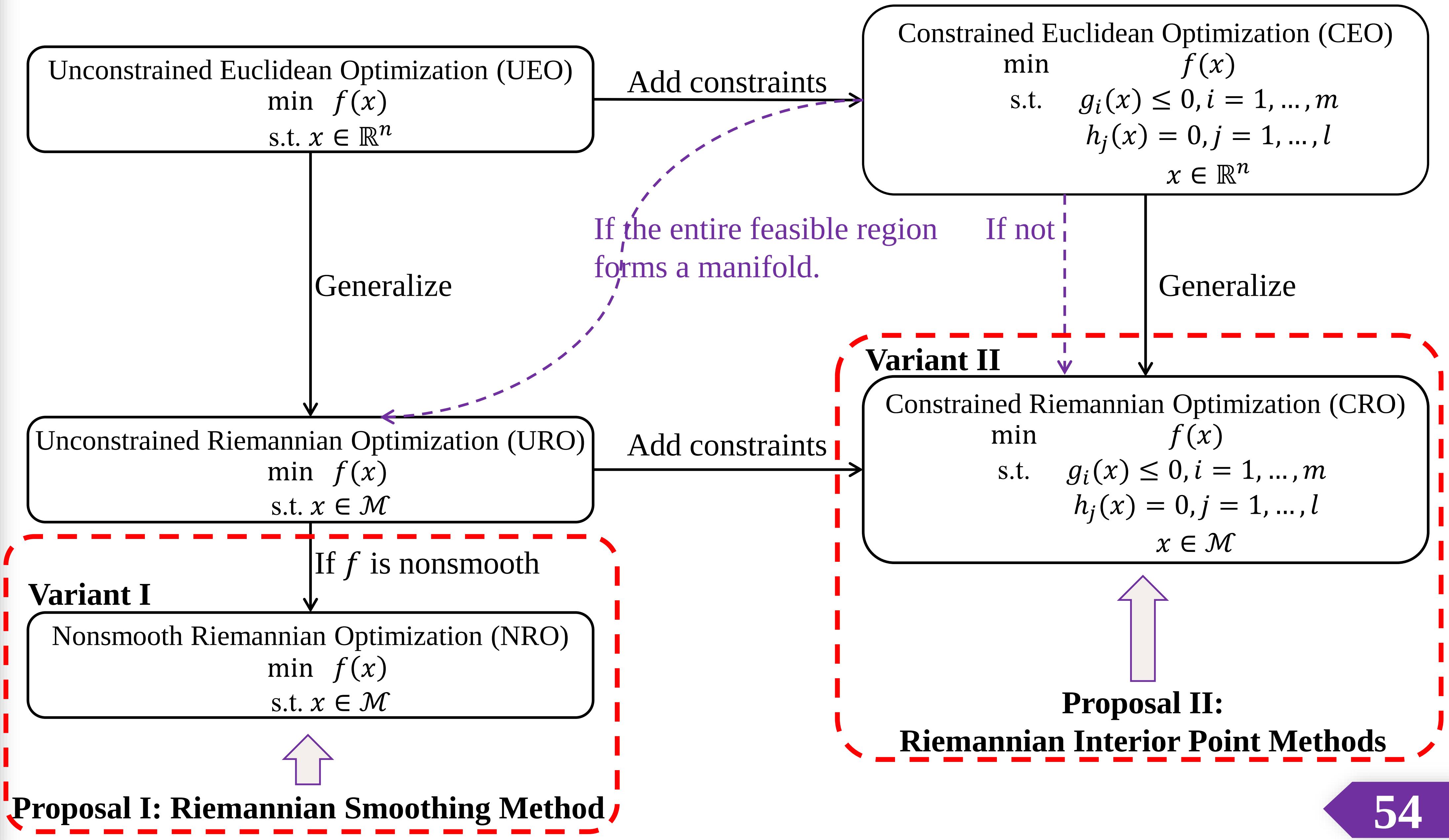
Yang, W. H., Zhang, L.-H., & Song, R. (2014). Optimality conditions for the nonlinear programming problems on Riemannian manifolds. *Pacific Journal of Optimization*, *10*(2), 415–434.

Zhang, C., Chen, X., & Ma, S. (2023). A Riemannian Smoothing Steepest Descent Method for Non-Lipschitz Optimization on Embedded Submanifolds of Rn. *Mathematics of Operations Research*, moor.2022.0286. https://doi.org/10.1287/moor.2022.0286

Appendix

Summary: Position of Our Works

Appendix



Grassmannian Manifold as a Quotient Manifold

Appendix

Grassmannian manifold is the set of linear subspaces of dimension p in $\mathbb{R}^n$.

$$\mathrm{Gr}(n,p) = \left\{\mathrm{span}(X): X \in \mathbb{R}^{n\times p}, X^\top X = I_p\right\}$$

We define an equivalence relation $\sim$ over $\mathrm{St}(n,p)$:

$$X \sim Y \Leftrightarrow \mathrm{span}(X) = \mathrm{span}(Y) \Leftrightarrow X = YQ \text{ for some } Q \in O(p),$$

where $\mathrm{O}(p)$is the orthogonal group. Formally, if $L = \mathrm{span}(X)$, we identify L with

$$[X] = \{Y \in \mathrm{St}(n,p): Y \sim X\}$$

This identification establishes a one-to-one correspondence between $\mathrm{Gr}(n,p)$ and the quotient set

$$\mathrm{St}(n,p)/{\sim} = \{[X]: X \in \mathrm{St}(n,p)\}.$$

Optimization over Grassmannian Manifold

Appendix

Principal Component Analysis (PCA)

Given k points $y_1, \dots, y_k \in \mathbb{R}^n$, the goal of PCA is to find a linear subspace $L \in \mathrm{Gr}(n,p)$ which fits the data $y_1, \dots, y_k$ as well as possible, in the sense that it solves

$$\min_{L \in \mathrm{Gr}(n,p)} \sum_{i=1}^{k} \mathrm{dist}(L, y_i)^2,$$

where $\mathrm{dist}(L, y)$ is the Euclidean distance between y and the point in L closest to y.

This objective function admits an explicit solution involving the SVD of the data matrix $M = [y_1, \dots, y_k]$. However, this is not the case for other objective functions.

For these, we may need more general optimization algorithms to address:

$$\min_{L \in \mathrm{Gr}(n,p)} f(L),$$

where objective function $f: \mathrm{Gr}(n,p) \to \mathbb{R}$.

Clearly, Euclidean optimization cannot solve these problems unless we convert the problem into some equivalent Euclidean problem.

Applications of Constrained Riemannian Optimization (CRO)

Nonnegative Principal Component Analysis (Montanari & Richard, 2016)
Find the nonnegative principal component vector of some data matrix $A \in \mathbb{R}^{n\times r}$:

$$\min_{X\in \mathrm{St}(n,k)} -\mathrm{tr}(X^\top A A^\top X) \ \text{ s. t. } X \geq 0.$$

Data Collaboration Analysis (Nosaka & Yoshise, 2023)
Nosaka and Yoshise created the collaborative data representations using fixed-rank manifolds.
To improve the performance, recently, Nosaka and me are trying to solve the new model:

$$\min_{G_i\in M_i} \frac{1}{2}\sum_{i,i'=1}^{N} \|\tilde{A}_i G_i - \tilde{A}_{i'} G_{i'}\|_F^2 \ \text{ s.t. } \|G_i\|_F^2 - \hat{m} = 0, \forall i \in N,$$

where $M_i = \{X \in \mathbb{R}^{\hat{m}_i\times\hat{m}} \mid \mathrm{Rank}(X) = \hat{m}\}$ and data matrices $\tilde{A}_i \in \mathbb{R} \times \tilde{m}_i$.

Riemannian Gradient (Strict definitions)

Appendix

Definition (differential D$F(x)$)
For manifolds $\mathcal{M}_1, \mathcal{M}_2$ and a mapping $F: \mathcal{M}_1 \to \mathcal{M}_2$, the differential of F at a point $x \in \mathcal{M}_1$, denoted as $\mathrm{D}F(x): T_x\mathcal{M}_1 \to T_{F(x)}\mathcal{M}_2$, is defined as

$$(\mathrm{D}F(x)[\xi])h = \xi(h \circ F), \ \forall h \in \mathfrak{F}_{F(x)}(\mathcal{M}_2).$$

Note that if $F: \mathcal{M}_1 \to \mathcal{M}_2$ and $h \in \mathfrak{F}_{F(x)}(\mathcal{M}_2)$, then $h \circ F \in \mathfrak{F}_x(\mathcal{M}_1)$.
We can show that $\mathrm{D}F(x): T_x\mathcal{M}_1 \to T_{F(x)}\mathcal{M}_2$ is a linear map.

Special case. Consider the differential of $f: \mathcal{M} \to \mathbb{R}$ at a point $x \in \mathcal{M}$. If we identify $T_{f(x)}\mathbb{R} \cong \mathbb{R}$, for $\xi \in T_x\mathcal{M}$, we have

$$\mathrm{D}f(x)[\xi] = \xi f.$$

$\rightsquigarrow$ $\mathrm{D}f(x): T_x\mathcal{M} \to \mathbb{R}$ is linear functional.

Definition (Riemannian gradient)
The Riemannian gradient of a function $f: \mathcal{M} \to \mathbb{R}$ at a point $x \in \mathcal{M}$, denoted as grad $f(x)$, is a unique element* in $T_x\mathcal{M}$ that satisfies

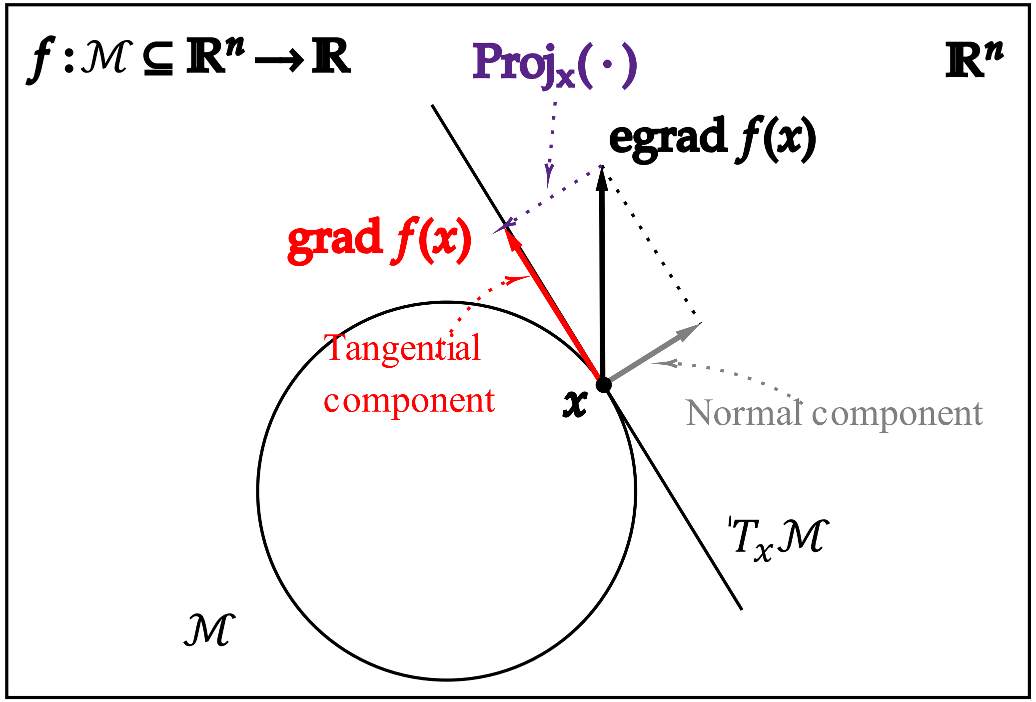
$$\langle \mathrm{grad}\, f(x), \xi \rangle_x = \mathrm{D}f(x)[\xi], \ \forall \xi \in T_x\mathcal{M}.$$

(*) Riesz Theorem: For an inner product space $(V, \langle\cdot,\cdot\rangle)$, if $T: V \to \mathbb{R}$ is a linear, then there is a unique $y \in V$ such that $T(x) = \langle v, x \rangle$ for all $x \in V$.

Calculation of Gradient on Embedded Submanifold

Proposition: For any embedded submanifold $\mathcal{M}$, Riemannian gradient $f: \mathcal{M} \to \mathbb{R}$ is the orthogonal projection onto $T_x\mathcal{M}$ of the Euclidean gradient:

$$\operatorname{grad} f(x) = \operatorname{Proj}_x(\operatorname{egrad} f(x))$$



Example: For $f(x) = x^\top A x$ on $\mathbb{S}^{n-1}$, we have egrad $f(x) = 2Ax$, and

$$\operatorname{Proj}_x(u) = (I_n - xx^\top)u.$$

It follows that grad $f(x) = \operatorname{Proj}_x(\operatorname{egrad} f(x)) = 2(I_n - xx^\top)Ax$.

Riemannian Metric Induces the Distance Space

The norm of a tangent vector ξ at any point x on $\mathcal{M}$ can be defined as

$$\| \xi \|_x := \sqrt{\langle \xi, \xi \rangle_x}$$

Furthermore, the length $L(c)$ of a curve $c: [a, b] \to \mathcal{M}$ on $\mathcal{M}$ can be defined as

$$L(c) := \int_a^b \|c'(t)\|_{c(t)} dt.$$

A natural distance on $\mathcal{M}$, called the Riemannian distance,

$$\operatorname{dist}(x, y) := \inf_c L(c)$$

where the infimum is taken over all curve segments which connect x to y, and thus $\mathcal{M}$ becomes a distance space.

Equivalent Definitions of Retraction

Definition (retraction). A smooth mapping $\mathrm{R}: T\mathcal{M} \to \mathcal{M}$ is called a retraction if for all $x \in \mathcal{M}$, the restriction $\mathrm{R}_x: T_x\mathcal{M} \to \mathcal{M}$ satisfies

$$\mathrm{R}_x(0) = x,\ \mathrm{DR}_x(0) = \mathrm{id}_{T_x\mathcal{M}}.$$

Here, $T\mathcal{M} = \{(x, v): x \in \mathcal{M} \text{ and } v \in T_x\mathcal{M}\}$ is called the tangent bundle of $\mathcal{M}$, and $\mathrm{id}_{T_x\mathcal{M}}$ represents the identity mapping in $T_x\mathcal{M}$.

Equivalent.

Definition (Retraction). A retraction on a manifold $\mathcal{M}$ is a smooth map

$$\mathrm{R}: \mathrm{T}\mathcal{M} \to \mathcal{M}: (x, \xi) \mapsto \mathrm{R}_x(\xi)$$

such that for each $(x, \xi) \in \mathrm{T}\mathcal{M}$ the curve $\gamma(t) := \mathrm{R}_x(t\xi)$ satisfies $\dot{\gamma}(0) = \xi$.

Appendix

Numerical Experiments of CP factorization

Appendix

- SpFeasDC_ls (Chen et al., 2020): **A difference-of-convex functions** approach for solving the split feasibility problem.
- RIPG_mod (Boţ & Nguyen, 2021): **This is a projected gradient method** with relaxation and inertia parameters for solving (4.4).
- APM_mod (Groetzner & Dür, 2020): **A modified alternating projection method for CP** factorization.

2nd Order Geometry: Calculation of Hessian on Embedded Submanifold

Appendix

Hess $f(x)$ is self-adjoint (i.e., symmetric) operator from and to $T_x\mathcal{M}$.

Proposition: For any embedded submanifold $\mathcal{M}$,

$$\text{Hess}\, f(x)[u] = \text{Proj}_x\,(\text{D}\,\text{grad} f(x)[u]).$$

Example: For $f(x) = x^\top Ax$ on $\mathbb{S}^{n-1}$, we have egrad $f(x) = 2Ax$, and

$$\text{Proj}_x\,(u) = (I_n - xx^\top)u.$$

It follows that grad $f(x) = \text{Proj}_x(\text{egrad}\, f(x)) = 2(I_n - xx^\top)Ax$.

Example: For $f(x) = \frac{1}{2}x^\top Ax$ on $\mathbb{S}^{n-1}$, we have grad $f(x) = (I_n - xx^\top)Ax$. Its differential is

$$\text{D}\,\text{grad}\, f(x)[u] = Au - (u^\top Ax + x^\top Au)x - (x^\top Ax)u$$

project to the tangent space at x to reveal

$$\text{Hess}\, f\,(x)[u] = Au - (x^\top Au)x - (x^\top Ax)u.$$

Note that Hess $f(x)\colon T_x\mathbb{S}^{n-1} \to T_x\mathbb{S}^{n-1}$ is true for $x \in T_x\mathbb{S}^{n-1} = \{u \in \mathbb{R}^n\colon x^T u = 0\}$.

Local Convergence — Riemannian Newton Method

Appendix

Definition (singularity)
Let $F: \mathcal{M} \to T\mathcal{M}$ be a smooth vector field. A point $p \in \mathcal{M}$ is call singularity of F if

$$F(p) = 0_p \in T_p\mathcal{M}$$

where 0_p is the zero element of $T_p\mathcal{M}$.

Algorithm 2 Riemannian Newton method

Goal: To find the singularity of the given vector field F.

Take $x_0 \in \mathcal{M}$ and set $k = 0$.

While stopping criterion not satisfied **do:**

> Solve the Newton equation
> $$\nabla F(x_k) v_k = -F(x_k),$$
> Update $x_{k+1} := \mathrm{R}_{x_k}(v_k)$;
> $k \leftarrow k + 1$;

End

Standard Newton assumptions:
(N1) There exists $x^*: F(x^*) = 0$.
(N2) $\nabla F(x^*)$ is nonsingular operator.
(N3) ∇F is locally Lipschitz cont. at x^*.

Local Convergence Rate:
(N1)-(N2) ⇒ superlinear (Fernandes et al., 2017)
(N1)-(N3) ⇒ quadratic (Ferreira & Silva, 2012)

Appendix

Problem I — Nonnegative Low Rank Matrix Approximation (Appendix: n varies under fixed m, r)

Problem I (Song & Ng, 2020)

$$\min_{X \in \mathbb{R}_r^{m\times n}} \| A - X \|_F^2 \quad \text{s.t. } X \geq 0,$$

where $\mathbb{R}_r^{m\times n} = \{X \in \mathbb{R}^{m\times n} : \text{rank}\,(X) = r\}$.

Experiments settings:

Fix $m = 20,\ r = 2$; we take $n \in \{8,16,24,32\}$. For each n, we generated 20 random instances* to examine.

Each experiment stopped successfully if solution with KKT residual $< 10^{-8}$ was found before the maximum time 10 (s) was reached.

	Column Num. n	RALM	REPM (LQH)	REPM (LSE)	RSQP	RIPM
Success Rate	8	0.05	0.25	0	0.9	**1**
	16	0.2	0.2	0.15	0.85	**1**
	24	0.05	0.45	0.15	0.15	**1**
	32	0.1	0.2	0.2	0	**1**
Average Time (s)	8	0.57	**0.29**	-	2.54	**0.30**
	16	1.27	**0.54**	2.91	6.83	**0.47**
	24	1.59	**0.71**	3.93	10.44	**0.90**
	32	2.35	**0.92**	5.13	-	**1.85**

Success Rate = 1 in each row.

The first two fastest results in each row.

(*): Let $B = \text{rand}(m, r)$; $C = \text{rand}(r, n)$; $A = B * C + 0.001 * \text{randn}(m, n)$.

Appendix

Problem I — Nonnegative Low Rank Matrix Approximation (Appendix: r varies under fixed m, n)

Problem I (Song & Ng, 2020)

$$\min_{X \in \mathbb{R}_r^{m\times n}} \| A - X \|_F^2 \quad \text{s.t. } X \geq 0,$$

where $\mathbb{R}_r^{m\times n} = \{X \in \mathbb{R}^{m\times n}: \text{rank}(X) = r\}$.

Experiments settings:
Fix $m = 20,\ n = 20$; we take $r \in \{2,4,8,16\}$. For each r, we generated 20 random instances* to examine.
Each experiment stopped successfully if solution with KKT residual $< 10^{-8}$ was found before the maximum time 10 (s) was reached.

	Rank r	RALM	REPM (LQH)	REPM (LSE)	RSQP	RIPM
Success Rate	2	0.15	0.2	0.25	0.75	**1**
	4	0.15	0.15	0	0	**1**
	8	0.15	0.05	0.1	0	**1**
	16	0.05	0.05	0.05	0	**1**
Average Time (s)	2	1.60	**0.60**	3.57	9.26	**0.64**
	4	**0.92**	**0.57**	-	-	1.28
	8	**0.70**	**0.56**	1.86	-	3.79
	16	**0.71**	**0.62**	1.93	-	3.41

Success Rate = 1 in each row.

The first two fastest results in each row.

(*): Let $B = \text{rand}(m, r); C = \text{rand}(r, n); A = B * C + 0.001 * \text{randn}(m, n)$.

RIPM vs. EIPM: RIPM Solves Newton Equation of Smaller Order

Appendix

Constrained Riemannian Optimization (CRO)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, \dots, m \\ & h_j(x) = 0, j = 1, \dots, l \\ & x \in \mathcal{M}, \end{aligned}$$

where $f: \mathcal{M} \to \mathbb{R}, h: \mathcal{M} \to \mathbb{R}^l$, and $g: \mathcal{M} \to \mathbb{R}^m$.

← Generalize

Constrained Euclidean Optimization (CEO)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, \dots, m \\ & h_j(x) = 0, j = 1, \dots, l \\ & x \in \mathbb{R}^n, \end{aligned}$$

where $f: \mathbb{R}^n \to \mathbb{R}, h: \mathbb{R}^n \to \mathbb{R}^l$, and $g: \mathbb{R}^n \to \mathbb{R}^m$.

- RIPM inherits the all advantages of Riemannian optimization.
- EIPM is a special case of RIPM when $\mathcal{M} \equiv \mathbb{R}^n$ or $\mathbb{R}^{n \times k}$.
- RIPM can solve some problems that EIPM cannot.
- RIPM solves condensed Newton equation of smaller order on $T_x\mathcal{M} \times \mathbb{R}^l$:

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}.$$

E.g., the Stiefel manifold can be used as the equality constraints; i.e., we set $h: \mathcal{M} \equiv \mathbb{R}^{n \times k} \to \mathrm{Sym}(k)$, where $h(X) = X^\top X - I_k$.
Here, EIPM requires us to solve an equation of order $nk + k(k+1)/2$.
But RIPM only requires us to solve an equation of order $nk - k(k+1)/2$, i.e., the dimension of $\mathrm{St}(n, k)$.

Future Works: Quasi-Newton RIPM

◆**Quasi-Newton RIPM**

The quasi-Newton RIPM can approximate the Hessian of Lagrangian in $\nabla F(w_k)$ with gradient information while ensuring its local convergence.

Algorithm of RIPM for (CRO)

Solve the perturbed Newton equation

$$\nabla F(w_k)\Delta w_k = -F(w_k) + \mu_k \hat{e},$$

Compute the step α_k such that $(z_{k+1}, s_{k+1}) > 0$;

Update $w_{k+1} = \bar{R}_{w_k}(\alpha_k \Delta w_k)$;

Choose $0 < \mu_{k+1} < \mu_k$;

$$\nabla F(w)\Delta w = \begin{pmatrix} \mathrm{Hess}_x \mathcal{L}(w)\,\Delta x + H_x \Delta y + G_x \Delta z \\ H_x^* \Delta x \\ G_x^* \Delta x + \Delta s \\ Z\Delta s + S\Delta z \end{pmatrix}$$

Quasi-Newton method, BFGS formulate, etc.

$$B(w)\Delta w = \begin{pmatrix} G(w)\Delta x + H_x \Delta y + G_x \Delta z \\ H_x^* \Delta x \\ G_x^* \Delta x + \Delta s \\ Z\Delta s + S\Delta z \end{pmatrix}$$

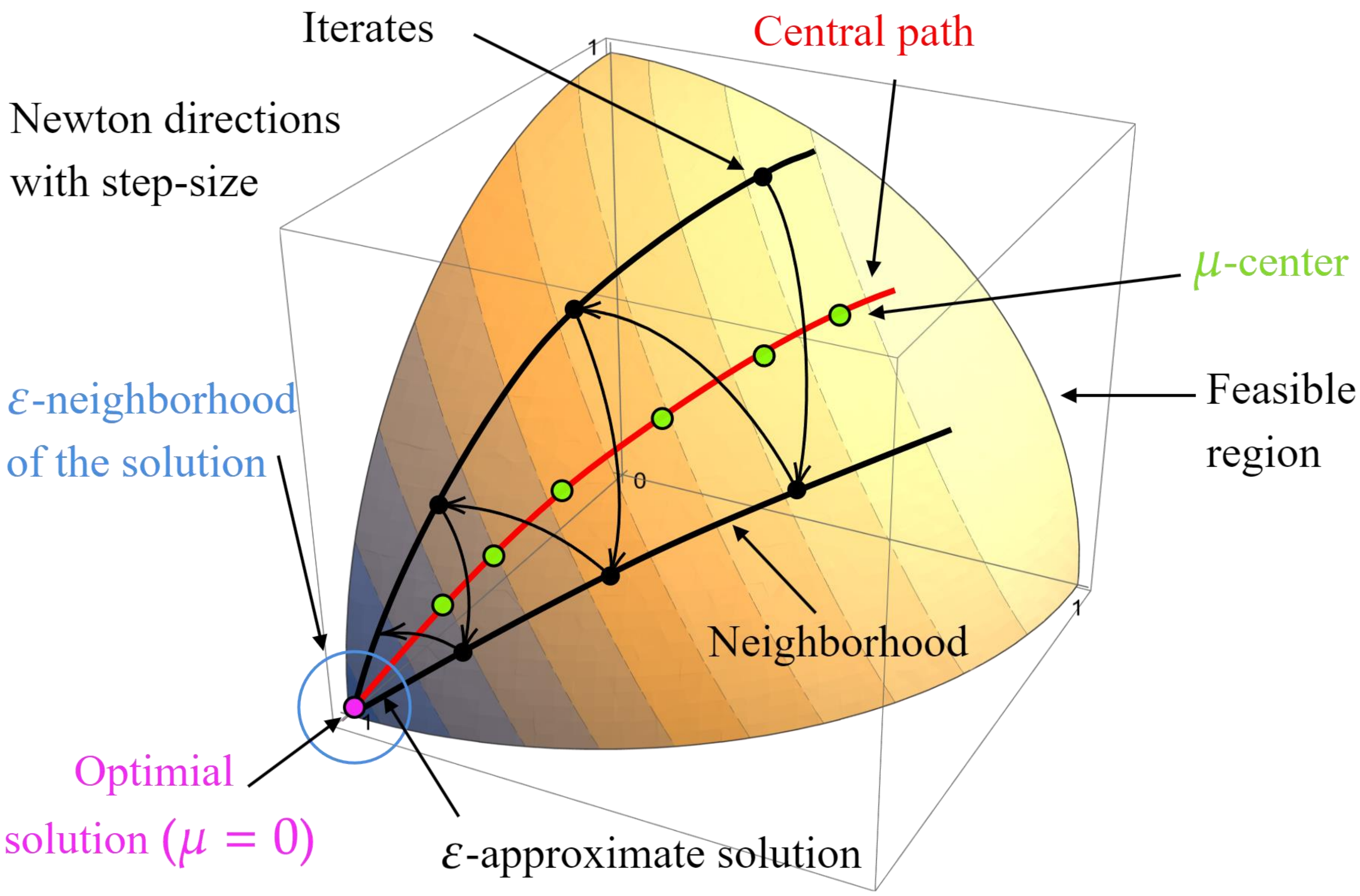
$$B(w_k)\Delta w_k = -F(w_k) + \mu_k \hat{e},$$

As the last chapter of my thesis, we only give some theoretical results.

There is still a great deal of work to be done to refine the quasi-Newton RIPM.

Primal Interior Point Method on Manifold

Appendix



Riemannian Interior Point Methods

Superlinear and Quadratic Convergence

1. Existence. There exists w^* satisfying the KKT conditions.
2. Smoothness. The functions f, g, h are smooth on M.
3. Regularity. The set $\{\operatorname{grad} h_i(x^*) : i = 1, \cdots, l\} \cup \{\operatorname{grad} g_i(x^*) : i \in \mathcal{A}(x)\}$ is linearly independent in $T_{x^*}M$.
4. Strict Complementarity. $(z^*)_i > 0$ if $g_i(x^*) = 0$ for all $i = 1, \cdots, m$.
5. Second-Order Sufficiency. $\langle \operatorname{Hess}_x \mathcal{L}(w^*)\xi, \xi \rangle > 0$ for all nonzero $\xi \in T_{x^*}M$ satisfying $\langle \xi, \operatorname{grad} h_i(x^*) \rangle = 0$ for $i = 1, \cdots, l$, and $\langle \xi, \operatorname{grad} g_i(x^*) \rangle = 0$ for $i \in \mathcal{A}(x^*)$.

Proposition (L. 2022)

If assumptions (1)-(5) hold, then standard Newton assumptions (N1)-(N3) hold for KKT vector field F.

Riemannian Interior Point Methods

Superlinear and Quadratic Convergence

On the other hand, to keep $(s_k, z_k) \geq 0$:

- Introducing the **perturbed** complementary equation,

$$Z\Delta s + S\Delta z = -ZSe + \mu e, \tag{15}$$

  so that we are able to keep the iterates far from the boundary.

- Compute the **damped** step sizes α_k, e.g., choose $\gamma_k \in (0, 1)$ and compute

$$\alpha_k := \min\left\{1, \gamma_k \min_i \left\{-\frac{(s_k)_i}{(\Delta s_k)_i} \mid (\Delta s_k)_i < 0\right\}, \gamma_k \min_i \left\{-\frac{(z_k)_i}{(\Delta z_k)_i} \mid (\Delta z_k)_i < 0\right\}\right\}, \tag{16}$$

  such that $(s_{k+1}, z_{k+1}) > 0$.

The relation of α_k and γ_k: [Yamashita and Yabe, 1996]

1. If $\gamma_k \to 1$, then $\alpha_k \to 1$.
2. If $1 - \gamma_k = O(\|F(w_k)\|)$, then $1 - \alpha_k = O(\|F(w_k)\|)$.

History of Euclidean Interior Point Method

Interior Point (IP) Method for NONLINEAR, NONCONVEX (1990-)

Early phase (1990-1995)

- Local algorithms with superlinear/ quadratic convergence [El-Bakry et al., 1996, Yamashita and Yabe, 1996].
- Global algorithms [El-Bakry et al., 1996]

Variations (1995-2010)

- Inexact Newton/ Quasi Newton IP Method
- Global strategy: *many* merit functions; linear search, or trust region, etc.

Update by Retraction

At a current point $w = (x, y, z, s)$ and direction $\Delta w = (\Delta x, \Delta y, \Delta z, \Delta s)$, the next iterate is calculated along a curve on $\mathscr{M}$, i.e.,

$$w(\alpha) := \bar{R}_w(\alpha \Delta w), \tag{17}$$

for some step length $\alpha > 0$.

By introducing

$$w(\alpha) = (x(\alpha), y(\alpha), z(\alpha), s(\alpha)), \tag{18}$$

we have

$$x(\alpha) = R_x(\alpha \Delta x),$$

and $y(\alpha) = y + \alpha \Delta y, z(\alpha) = z + \alpha \Delta z, s(\alpha) = s + \alpha \Delta s.$

Centrality conditions

Given $w_0 = (x_0, y_0, z_0, s_0)$ with $(z_0, s_0) > 0$, let $\tau_1 := \frac{\min(Z_0 S_0 e)}{z_0^T s_0 / m}$, $\quad \tau_2 := \frac{z_0^T s_0}{\|F(w_0)\|}$.
Let $\gamma \in (0, 1)$ be a constant. Define centrality functions:

$$f^I(\alpha) := \min(Z(\alpha)S(\alpha)e) - \gamma\tau_1 \frac{z(\alpha)^T s(\alpha)}{m}, \tag{19}$$

$$f^{II}(\alpha) := z(\alpha)^T s(\alpha) - \gamma\tau_2 \|F(w(\alpha))\|. \tag{20}$$

For $i = I, II$, let

$$\alpha^i := \max_{\alpha \in (0,1]} \left\{ \alpha : f^i(t) \geq 0, \ \text{for all } t \in (0, \alpha] \right\}. \tag{21}$$

Global RIP Algorithm

1. Choose $\sigma_k \in (0,1)$; for w_k, compute the perturbed Newton direction Δw_k with

$$\mu_k = z_k^T s_k / m \tag{22}$$

and by

$$\nabla F(w) \Delta w = -F(w) + \sigma_k \mu_k \hat{e}. \tag{23}$$

2. Step length selection.
   1. Centrality conditions: Choose $1/2 < \gamma_k < \gamma_{k-1} < 1$; compute $\alpha^i, i = I, II$, from (21); and let

$$\bar{\alpha}_k = \min(\alpha^I, \alpha^{II}). \tag{24}$$

   2. Sufficient decreasing: Choose $\theta \in (0,1)$, and $\beta \in (0, 1/2]$. Let $\alpha_k = \theta^t \bar{\alpha}_k$, where t is the smallest nonnegative integer such that α_k satisfies

$$\varphi(\bar{R}_{w_k}(\alpha_k \Delta w_k)) - \varphi(w_k) \leq \alpha_k \beta \langle \operatorname{grad} \varphi_k, \Delta w_k \rangle. \tag{25}$$

3. Let $w_{k+1} = \bar{R}_{w_k}(\alpha_k \Delta w_k)$ and $k \leftarrow k+1$.

Auxiliary Results I: Boundedness of the sequences

Given $\epsilon \geq 0$, let us define the set

$$\Omega(\epsilon) := \left\{ w \in \mathscr{M} : \epsilon \leq \varphi(w) \leq \varphi_0, \min(ZSe)/(z^T s/m) \geq \tau_1/2, z^T s/\|F(w)\| \geq \tau_2/2 \right\}.$$

Lemma (Boundedness of the sequences I, L. 2022)

If $\epsilon > 0$ and $w_k \in \Omega(\epsilon)$ for all k, then

1. *the sequence $\{z_k^T s_k\}$ and $\{(z_k)_i(s_k)_i\}$, $i = 1, 2, \ldots, m$, are all bounded above and below away from zero.*
2. *the sequence $\{z_k\}$ and $\{s_k\}$ are bounded above and component-wise bounded away from zero;*
3. *the sequence $\{w_k\}$ is bounded;*
4. *the sequence $\{\|\nabla F(w_k)^{-1}\|\}$ is bounded;*
5. *the sequence $\{\Delta w_k\}$ is bounded.*

Lemma (Boundedness of the sequences II, L. 2022)

If $\{\sigma_k\}$ is bounded away from zero. Then, $\{\bar{\alpha}_k\}$ is bounded away from zero.

Auxiliary Results II: Continuity of Some Special Scalar Fields

Lemma (L. 2022)

Let $x \in M$ and A_x be a linear operator on T_xM. Then, the values $\|\widehat{A}_x\|_2$ and $\|\hat{A}_x\|_F$ are invariant under a change of orthonormal basis; moreover,

$$\|A_x\| = \|\hat{A}_x\|_2 \leq \|\hat{A}_x\|_F. \tag{26}$$

Lemma (L. 2022)

$$x \mapsto \|\widehat{\mathrm{Hess} f(x)}\| \tag{27}$$

is a continuous scalar field on M. It is true for all h_i, g_i.

$$x \mapsto \|H_x\| \text{ and } x \mapsto \|G_x\| \tag{28}$$

are continuous scalar field on M.

An Intuitive Barrier Method on Manifolds

Consider

$$\min_{x \in M} \quad f(x) \quad \text{s.t.} \quad c(x) \geq 0. \qquad \text{(RCOP\_Ineq)}$$

Its logarithmic barrier function is

$$B(x;\mu) := f(x) - \mu \textstyle\sum_{i=1}^{m} \log c_i(x),$$

where $\mu > 0$. Note that the function $x \mapsto B(x;\mu)$ is differentiable on, strict $\mathcal{F} := \{x \in M : c(x) > 0\}$. Its Riemannian gradient is

$$\operatorname{grad} B(x;\mu) = \operatorname{grad} f(x) - \textstyle\sum_{i=1}^{m} \frac{\mu}{c_i(x)} \operatorname{grad} c_i(x).$$

Barrier Method on Manifolds

1. Set $x_0 \in M$ to a strictly feasible point, i.e., $c(x_0) > 0$, and set $\mu_0 > 0$ and $k \leftarrow 0$.
2. Check whether x_k satisfies a stopping test for (RCOP_Ineq).
3. Compute an unconstrained minimizer $x(\mu_k)$ of $B(x;\mu_k)$ with a warm starting point x_k.
4. $x_{k+1} \leftarrow x(\mu_k)$; choose $\mu_{k+1} < \mu_k$; $k \leftarrow k+1$. Return to Step 1.

Barrier Method on Manifolds

Barrier Method

Consider the following simple problem on a sphere manifold, $\mathbb{S}^2 := \{x \in \mathbb{R}^3 : \|x\|_2 = 1\}$,

$$\min_{x \in \mathbb{S}^2} \quad a^T x \quad \text{s.t.} \quad x \geq 0, \tag{SP}$$

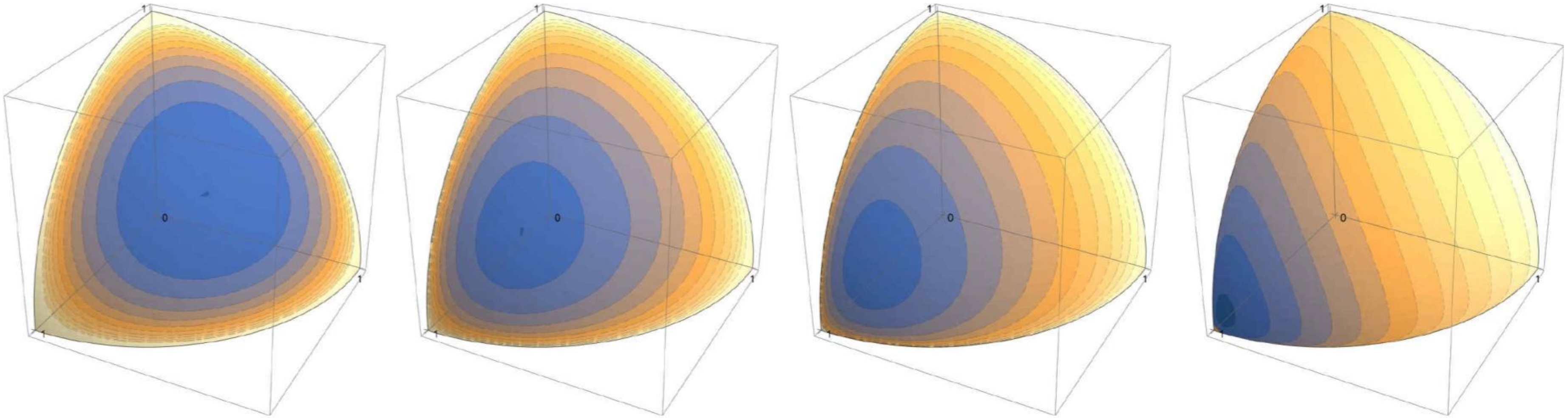
where $a = [-1, 2, 1]^T$. Its solution is $x^* = [1, 0, 0]^T$.

Now, check the KKT conditions at x (asterisks omitted below): $\operatorname{grad} f(x) = (I_n - xx^T)a = [0, 2, 1]^T$. The constraint $x \geq 0$ implies $c_i(x) = e_i^T x$ for $i = 1, 2, 3$;

$$\operatorname{grad} c_1(x) = (I_n - xx^T)e_1 = [0, 0, 0]^T;$$
$$\operatorname{grad} c_2(x) = (I_n - xx^T)e_2 = [0, 1, 0]^T;$$
$$\operatorname{grad} c_3(x) = (I_n - xx^T)e_3 = [0, 0, 1]^T.$$

Clearly, the multipliers $z^* = [0, 2, 1]^T$, and LICQ and strict complementarity hold.

Barrier Method on Manifolds



(a) (b) (c) (d)

Figure: Contour plots of logarithmic barrier function $B(x; \mu)$ of (SP) for (a) $\mu = 10$ (b) $\mu = 1$ (c) $\mu = 0.5$ (d) $\mu = 0.1$. The blue area indicates low values.

Barrier Method on Manifolds

Finally, we find that $\lim_{k\to\infty} x_k = x^*$ and that

$$\lim_{k\to\infty} \mu_k/c_1(x_k) = 0 = z^*_{(1)}, \lim_{k\to\infty} \mu_k/c_2(x_k) = 2 = z^*_{(2)}, \lim_{k\to\infty} \mu_k/c_3(x_k) = 1 = z^*_{(3)},$$

which are the notable features of the classical barrier method; see [Forsgren et al., 2002, Theorem 3.10 & 3.12].

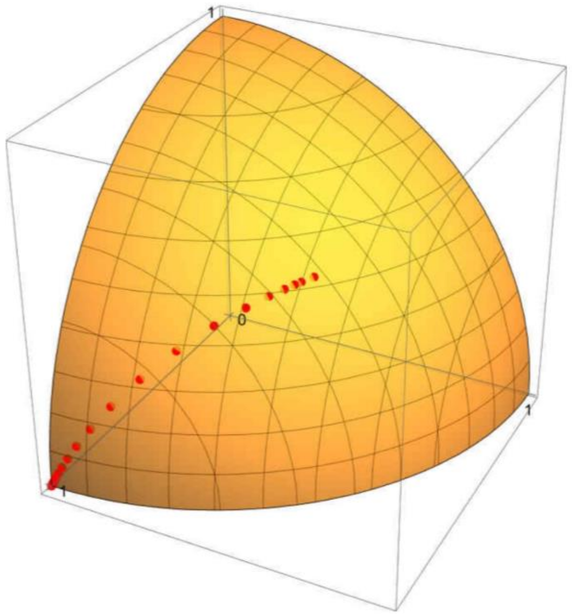


Figure: Iterates x_k of barrier method for (SP).

Barrier Method on Manifolds

Furthermore, if we denote the minimizer of $B(x;\mu)$ by either x_μ or $x(\mu)$, it must be that $\operatorname{grad} B(x_\mu;\mu) = 0$.

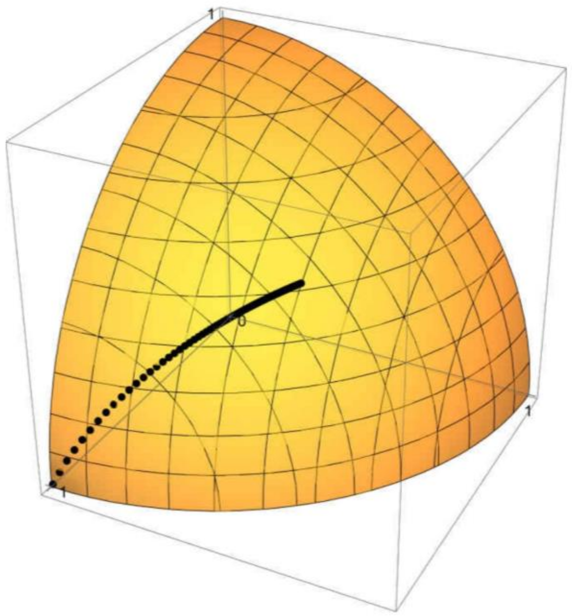


Figure: Existence of a central path for (SP).