Riemannian Interior Point Methods for Constrained Optimization on Manifolds

Zhijian Lai[1], Akiko Yoshise[1]

[1]University of Tsukuba

Introduction

A **Riemannian manifold** M is a set that can be locally linearizable, with a smooth mapping $x \mapsto \langle\cdot,\cdot\rangle_x$, which is an inner product on the tangent spaces T_xM.

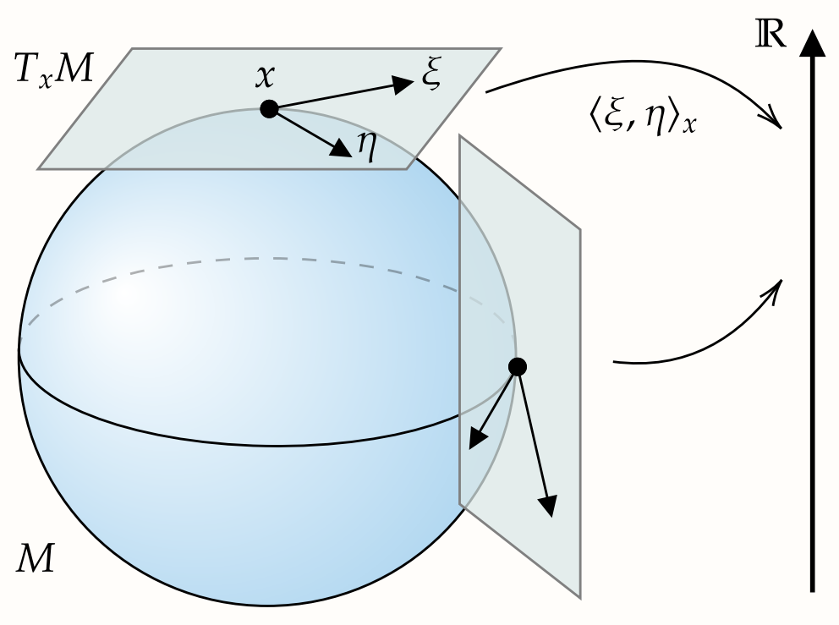


Figure 1: Unit sphere: $M = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ and $T_xM = \{v \in \mathbb{R}^n : \langle x, v\rangle = 0\}$.

Riemannian Optimization

Riemannian Optimization:

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

where $f : M \to \mathbb{R}$ and M is a Riemannian manifold.

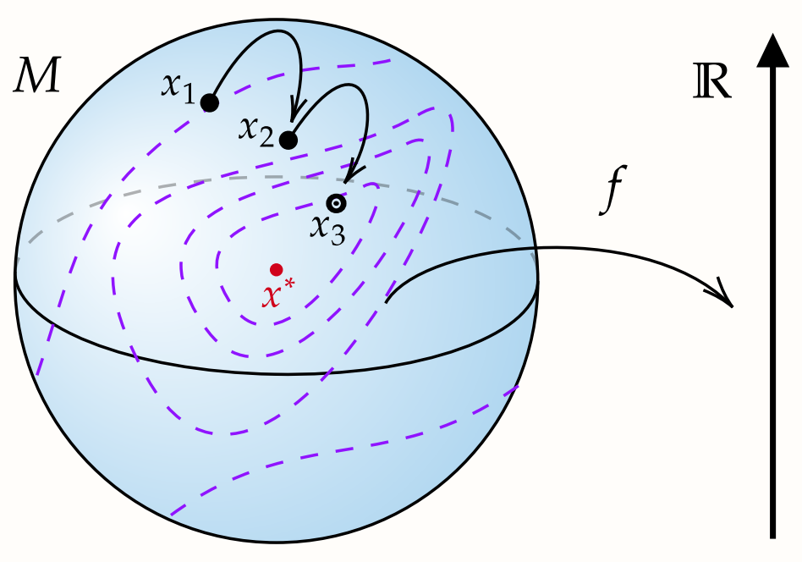


Figure 2: Iteration on unit sphere.

40+ available manifolds in solver "Manopt"[1]

- Stiefel manifold, $\mathrm{St}(n,k) = \{X \in \mathbb{R}^{n\times k} : X^\top X = I\}$.
- Fixed rank manifold, $\mathbb{R}_r^{m\times n} = \{X \in \mathbb{R}^{m\times n} : \mathrm{rk}(X) = r\}$.

Riemannian version of classical methods: steepest decent, conjugate gradient, Newton, trust region, proximal, ADMM.

Advantages of Riemannian Optimization

1. Exploit the geometric structure of the constrained set.
2. Convergence properties of like optimization on Euclidean.
3. Transfer the constrained problem to the unconstrained one.

More Requirements in Applications

- Nonnegative PCA on Stiefel manifold

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

- Nonnegative matrix completion on fixed rank manifold

$$\min_{X\in\mathbb{R}_r^{m\times n}} \sum_{(i,j)\in\Omega} (X_{ij} - A_{ij})^2 \quad \text{s.t. } X \geq 0$$

$\rightsquigarrow$ Can we use the solver "Manopt" directly?

Some limitations of Riemannian optimization:

1. Existing manifold solvers lack flexibility, and adding even one more constraint can make it impossible to use them directly. E.g., $x \in M, x \geq 0$.
2. Adding new constraints does not necessarily guarantee that the feasible set is still a manifold.
3. Even if a feasible set is proven to be a manifold, there are no available software packages to support it.

$\rightsquigarrow$ We aim to develop a new model to address these issues.

Riemannian Constrained Optimization Problem

Riemannian Constrained Optimization Problem:

$$\begin{aligned} \min_{x\in M}\ & f(x) \\ \text{s.t.}\ & h(x) = 0, \text{ and } g(x) \leq 0, \end{aligned} \tag{RCOP}$$

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

Advantages of (RCOP):

1. Still using the geometric structure of M. The advantages of Riemannian optimization are maintained.
2. Very flexible, even if h, g cannot form a new manifold.
3. Based directly on an existing solver "Manopt".

Riemannian version of classical algorithms:

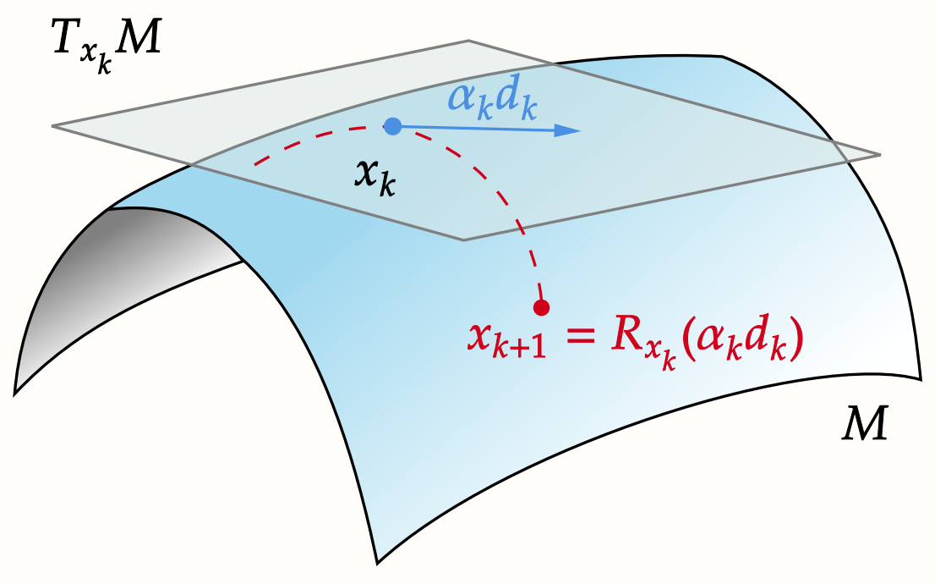
- Augmented Lagrangian Method [2];
- Exact Penalty Method [2];
- Sequential Quadratic Programming Method [3].
- $\rightsquigarrow$ In this talk, we consider Riemannian version of Interior Point Method.

Preliminaries

Q1: How to move on manifolds? Retraction!

A **retraction** R maps tangent vectors back to the manifold.

$$R_x : T_xM \to M \text{ for any } x.$$



Euclidean	Riemannian
$x_{k+1} = x_k + \alpha_k d_k$	$x_{k+1} = R_{x_k}(\alpha_k d_k)$

Q2: Where to move towards on manifolds? Riemannian Gradient!

For embedded submanifold M, **Riemannian gradient** of $f : M \to \mathbb{R}$ is the orthogonal projection onto T_xM of the Euclidean gradient, $\mathrm{grad}\, f(x) = \mathrm{Proj}_x(\mathrm{egrad}\, f(x))$.

Supplementary:

A **vector field** is a mapping F defined on M such that $F(x) \in T_xM$ for all $x \in M$. Riemannian gradient, $x \mapsto \mathrm{grad}\, f(x)$, is a vector field generated by scalar field $f : M \to \mathbb{R}$.

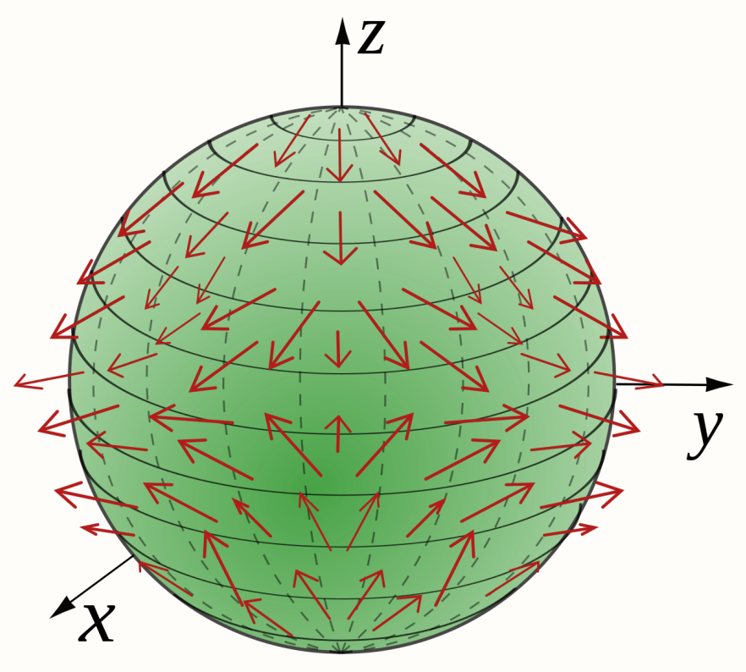


Figure 3: A vector field on a unit sphere. Source: Wikipedia.

Covariant derivative of a vector field F:

$$\overset{\substack{\text{Riemannian connection}\\ \downarrow}}{\nabla}\underset{\substack{\uparrow\\ \text{general vector field}}}{F}(x) : T_xM \to T_xM,\ \textit{linear operator}.$$

$\mathrm{Hess}\, f(x) \triangleq \nabla\, \mathrm{grad}\, f(x)$ is called **Riemannian Hessian**.

Riemannian Newton method: To find singularity $x^* \in M$ such that $F(x^*) = 0_{x^*}$.

(Step 1.) Solve a linear system on $T_{x_k}M \ni v_k$:

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

(Step 2.) $x_{k+1} = R_{x_k}(v_k)$. Return to Step 1.

Our proposal: Riemannian Interior Point Methods

KKT Vector Field

Lagrangian function of (RCOP) is

$$\mathcal{L}(x,y,z) \triangleq f(x) + y^T h(x) + z^T g(x).$$

$x \mapsto \mathcal{L}(x,y,z)$ is a real-valued function on M, then we have

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

Riemannian KKT conditions [2] are

$$\begin{cases} \mathrm{grad}_x\, \mathcal{L}(x,y,z) = 0_x, \\ h(x) = 0, \\ g(x) \leq 0, \\ Zg(x) = 0, (Z := \mathrm{diag}\,(z_1,\ldots,z_m)) \\ z \geq 0. \end{cases}$$

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

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

where $w := (x,y,z,s) \in \mathscr{M} \triangleq M \times \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^m$. Note that $T_w\mathscr{M} \equiv T_xM \times \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^m$.

Covariant Derivative of KKT Vector Field

For each $x \in M$, we define

$$H_x : \mathbb{R}^l \to T_xM, \quad H_x v \triangleq \sum_{i=1}^{l} v_i\, \mathrm{grad}\, h_i(x).$$

Hence, the adjoint operator is

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

The linear operator $\nabla F(w) : T_w\mathscr{M} \to T_w\mathscr{M}$ 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 $\Delta w = (\Delta x, \Delta y, \Delta s, \Delta z) \in T_xM \times \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^m \equiv T_w\mathscr{M}$.

Riemannian Interior Point Method (RIPM)

Basic RIPM Algorithm:

Step 0. Initial w_0 with $(z_0, s_0) > 0$.

Step 1. Solve

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

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

Step 2. Compute the step sizes α_k such that $(z_{k+1}, s_{k+1}) > 0$.

Step 3. Update:

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

Step 4. Let $\mu_k \to 0$. Return to Step 1.

Local Convergence: Under some standard assumptions.

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

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

Global Line Search RIPM Algorithm:

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

Backtracking for step size α_k: With a slight abuse of notation, we also let

$$\varphi(\alpha) \triangleq \varphi(\underbrace{\bar{R}_{w_k}(\alpha\Delta w_k)}_{\text{new iterate}}) \text{ for fixed } w_k \text{ and } \Delta w_k,$$

then $\varphi(0) = \varphi(w_k) =: \varphi_k$ and $\varphi'(0) = \langle \mathrm{grad}\, \varphi(w_k), \Delta w_k\rangle$. Sufficient decreasing condition asks

$$\varphi(\alpha_k) - \varphi(0) \leq \alpha_k\beta\varphi'(0).$$

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.

Global Convergence: Under some standard assumptions.

For any limit point $w^* = (x^*, y^*, z^*, s^*)$ of $\{w_k\}, x^*$ is a Riemannian KKT point of problem (RCOP).

Conclusion

Riemannian IPM (RIPM) vs. Euclidean IPM (EIPM)

1. RIPM inherits the advantages of Riemannian optimization and can exploit the geometric structure of the constraints.
2. EIPM is a special case of RIPM when $\mathbb{M} = \mathbb{R}^n$.
3. RIPM solves Newton equation of smaller order on $T_x\mathbb{M} \times \mathbb{R}^l$:
$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w\Delta x + H_x\Delta y \\ H_x^*\Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}.$$
4. RIPM can solve some problems that EIPM cannot. E.g., $\mathrm{rk}(X) = r$ is not continuous, we can not apply EIPM.

Our contributions:

1. Propose a Riemannian version of the interior point method.
2. Prove the local superlinear/quadratic and global convergence.
3. Establish some foundational concepts, such as the KKT vector field and its covariant derivative.

Future work: The more sophisticated and robust global strategies are often based on the trust region or filter line-search method.

References

[1] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre.
Manopt, a Matlab toolbox for optimization on manifolds.
Journal of Machine Learning Research, 15(42):1455–1459.

[2] Changshuo Liu and Nicolas Boumal.
Simple algorithms for optimization on Riemannian manifolds with constraints.
Applied Mathematics & Optimization, 82(3):949–981, 2020.

[3] Mitsuaki Obara, Takayuki Okuno, and Akiko Takeda.
Sequential quadratic optimization for nonlinear optimization problems on riemannian manifolds.
SIAM Journal on Optimization, 32(2):822–853, 2022.

Summer School on Continuous Optimization and Related Fields. Institute of Statistical Mathematics, Tokyo, 9 August, 2023.