Riemannian Interior Point Methods for Constrained Optimization on Manifolds

*Zhijian Lai Akiko Yoshise

University of Tsukuba

s2130117@s.tsukuba.ac.jp

International Workshop on Continuous Optimization
December 4, 2022

Overview

Riemannian Manifold

Riemannian manifold M is a locally linearizable set, equipped with a smoothly-varying inner product $\langle \cdot, \cdot \rangle_x$ on the tangent spaces T_xM.

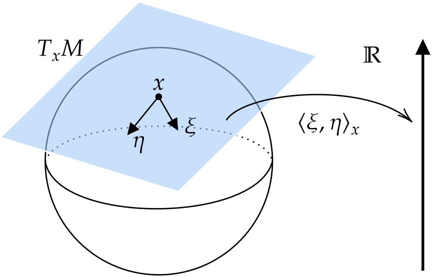
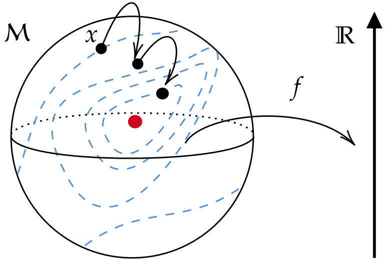


Figure: Manifold of unit sphere, $M = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$.

Given $f : M \to \mathbb{R}$, solve

$$\min_{x \in M} f(x) \tag{1}$$

where M is a Riemannian manifold.



Unconstrained problem on manifold.

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

Riemannian version of classical methods (2002-)

steepest decent, conjugate gradient, trust region, BFGS, proximal gradient, ADMM and more.

Applications

1. Stiefel manifold, $\mathrm{St}(n,k) = \{X \in \mathbb{R}^{n\times k} : X^\top X = I\}$.

$$\textbf{PCA:} \quad \min_{X\in \mathrm{St}(n,k)} -\mathrm{trace}(X^\top A^\top AX). \tag{2}$$

Applications

1. Stiefel manifold, $\mathrm{St}(n,k) = \{X \in \mathbb{R}^{n\times k} : X^\top X = I\}$.

$$\textbf{PCA:} \min_{X\in \mathrm{St}(n,k)} -\mathrm{trace}(X^\top A^\top A X). \tag{2}$$

2. Fixed rank manifold, $\mathbb{R}_r^{m\times n} = \{X \in \mathbb{R}^{m\times n} : \mathrm{rank}(X) = r\}$.

$$\textbf{Low-rank matrix completion:} \min_{X\in \mathbb{R}_r^{m\times n}} \sum_{(i,j)\in\Omega} (X_{ij} - M_{ij})^2. \tag{3}$$

More Requirements

1. Stiefel manifold, $\mathrm{St}(n,k) = \{X \in \mathbb{R}^{n\times k} : X^\top X = I\}$.

$$\textbf{Nonnegative PCA:} \min_{X\in \mathrm{St}(n,k)} -\mathrm{trace}(X^\top A^\top A X) \text{ s.t. } X \geq 0. \tag{4}$$

2. Fixed rank manifold, $\mathbb{R}_r^{m\times n} = \{X \in \mathbb{R}^{m\times n} : \mathrm{rank}(X) = r\}$.

$$\textbf{Nonnegative Low-rank matrix completion:} \min_{X\in \mathbb{R}_r^{m\times n}} \sum_{(i,j)\in\Omega} (X_{ij} - M_{ij})^2 \text{ s.t. } X \geq 0. \tag{5}$$

New Topic — Riemannian Constrained Optimization Problem

We consider

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

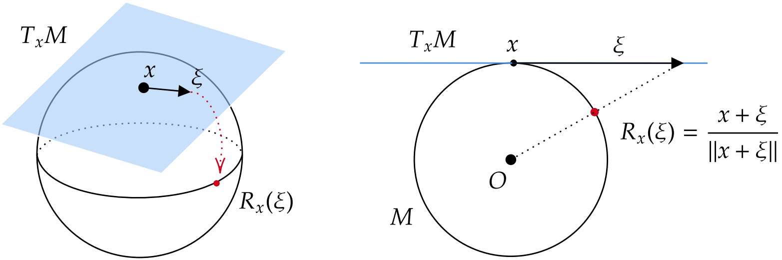
where $\mathbb{M}$ is a Riemannian manifold, $f : \mathbb{M} \to \mathbb{R}$, $h : \mathbb{M} \to \mathbb{R}^l$, and $g : \mathbb{M} \to \mathbb{R}^m$.

Riemannian optimality conditions:
KKT conditions; Second-order conditions [Yang et al., 2014];
More constraint qualifications (CQ) [Bergmann and Herzog, 2019];
Sequential optimality conditions [Yamakawa and Sato, 2022].

Riemannian algorithms:
Augmented Lagrangian Method [Liu and Boumal, 2020, Yamakawa and Sato, 2022];
Exact Penalty Method [Liu and Boumal, 2020];
Sequential Quadratic Programming Method [Schiela and Ortiz, 2020, Obara et al., 2022].
In this talk, we consider Interior Point Method.

Retraction — moving on a manifold

A **retraction** R maps tangent vectors back to the manifold. $R_x : T_xM \to M$ for any x.

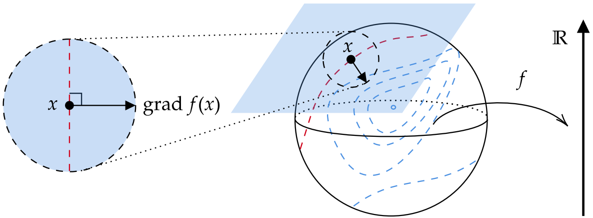


Euclidean	Riemannian
$x_{k+1} = x_k + \alpha_k \xi_k$	$x_{k+1} = R_{x_k}(\alpha_k \xi_k)$

Riemannian gradient — a vector field

Riemannian gradient, $\operatorname{grad} f(x)$, is the direction of steepest ascent in tangent space at x:

$$\frac{\operatorname{grad} f(x)}{\|\operatorname{grad} f(x)\|} = \underset{\xi \in T_xM : \|\xi\|=1}{\arg\max} \left(\lim_{\alpha \to 0} \frac{f(R_x(\alpha\xi)) - f(x)}{\alpha} \right).$$



Note that $x \mapsto \operatorname{grad} f(x)$ is a **vector field** on M.

Covariant derivative & Hessian & Newton method

Covariant derivative of a vector field F:

Riemannian connection

$$\nabla F(x) : T_x M \to T_x M, \ \textit{linear operator}.$$

general vector field

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

Covariant derivative & Hessian & Newton method

Covariant derivative of a vector field F:

Riemannian connection

$$\nabla F(x) : T_x M \to T_x M, \ \textit{linear operator}.$$

general vector field

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

Riemannian Newton method: To find **singularity** $x^* \in M$ such that $F(x^*) = 0_{x^*}$.
Solve a linear system on $T_{x_k}M \ni v_k$:

$$\nabla F(x_k) v_k = -F(x_k), \tag{6}$$

then $x_{k+1} = R_{x_k}(v_k)$.

Formulation of RIPM

We consider

$$\min_{x\in\mathbb{M}} \quad f(x) \qquad \text{s.t.} \quad h(x)=0, \text{ and } g(x)\leq 0, \tag{RCOP}$$

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

Lagrangian function is

$$\mathcal{L}(x,y,z) \triangleq f(x) + y^T h(x) + z^T g(x). \tag{7}$$

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

Formulation of RIPM

We consider

$$\min_{x \in \mathbb{M}} \quad f(x) \qquad \text{s.t.} \quad h(x) = 0, \ \text{and } g(x) \leq 0, \tag{RCOP}$$

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

Lagrangian function is

$$\mathcal{L}(x, y, z) \triangleq f(x) + y^T h(x) + z^T g(x). \tag{7}$$

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

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

KKT Vector Field

Riemannian KKT conditions [Liu and Boumal, 2020] are

$$\begin{cases} \operatorname{grad}_x \mathcal{L}(x, y, z) = 0_x, \\ h(x) = 0, \\ g(x) \leq 0, \\ Zg(x) = 0, (Z := \operatorname{diag}(z_1, \ldots, z_m)) \\ z \geq 0. \end{cases} \tag{8}$$

KKT Vector Field

Riemannian KKT conditions [Liu and Boumal, 2020] are

$$\begin{cases} \operatorname{grad}_x \mathcal{L}(x,y,z) = 0_x, \\ h(x) = 0, \\ g(x) \leq 0, \\ Zg(x) = 0, (Z := \operatorname{diag}(z_1, \ldots, z_m)) \\ z \geq 0. \end{cases} \tag{8}$$

Definition (KKT Vector Field, L.2022)

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

$$F(w) \triangleq \begin{pmatrix} \operatorname{grad}_x \mathcal{L}(x,y,z) \\ h(x) \\ g(x) + s \\ ZSe \end{pmatrix} = 0_w := \begin{pmatrix} 0_x \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{ and } (z,s) \geq 0, \tag{9}$$

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

Covariant Derivative of KKT Vector Field

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

$$H_x : \mathbb{R}^l \to T_x\mathbb{M}, \qquad H_x v \triangleq \sum_{i=1}^{l} v_i \operatorname{grad} h_i(x). \tag{10}$$

Hence, the adjoint operator is

$$H_x^* : T_x\mathbb{M} \to \mathbb{R}^l, \qquad H_x^* \xi = \left[\langle \operatorname{grad} h_1(x), \xi \rangle_x, \cdots, \langle \operatorname{grad} h_l(x), \xi \rangle_x\right]^T. \tag{11}$$

Lemma (L. 2022)

The linear operator $\nabla F(w) : T_w\mathscr{M} \to T_w\mathscr{M}$ *is given by*

$$\nabla F(w)\Delta w = \begin{pmatrix} \operatorname{Hess}_x \mathcal{L}(w)\Delta x + H_x \Delta y + G_x \Delta z \\ H_x^* \Delta x \\ G_x^* \Delta x + \Delta s \\ Z\Delta s + S\Delta z \end{pmatrix}, \tag{12}$$

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

Riemannian Interior Point Method (RIPM)

Step 0. Initial w_0 with $(z_0, s_0) > 0$.
Step 1. Solve

$$\nabla F(w_k)\Delta w_k = -F(w_k) + \mu_k \hat{e}, \tag{13}$$

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:

$$w_{k+1} = \bar{R}_{w_k}(\alpha_k \Delta w_k). \tag{14}$$

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

Riemannian Interior Point Method (RIPM)

Step 0. Initial w_0 with $(z_0, s_0) > 0$.
Step 1. Solve

$$\nabla F(w_k)\Delta w_k = -F(w_k) + \mu_k \hat{e}, \tag{13}$$

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:

$$w_{k+1} = \bar{R}_{w_k}(\alpha_k \Delta w_k). \tag{14}$$

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

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

Global Line Search RIPM Algorithm[1]

- **Merit function:** Choose $\varphi(w) \triangleq \|F(w)\|^2$.
- **Backtracking for step size α_k:**
  1. Centrality conditions.
  2. With a slight abuse of notation, we also let

$$\varphi(\alpha) \triangleq \varphi(\underbrace{\bar{R}_{w_k}(\alpha \Delta w_k)}_{\text{new iterate}}) \text{ for fixed } w_k \text{ and } \Delta w_k, \tag{15}$$

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

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

[1]The most classic global algorithm for Euclidean IPM is [El-Bakry et al., 1996].

Global Line Search RIPM Algorithm[1]

- **Merit function:** Choose $\varphi(w) \triangleq \|F(w)\|^2$.
- **Backtracking for step size α_k:**
  1. Centrality conditions.
  2. With a slight abuse of notation, we also let

$$\varphi(\alpha) \triangleq \varphi(\underbrace{\bar{R}_{w_k}(\alpha \Delta w_k)}_{\text{new iterate}}) \text{ for fixed } w_k \text{ and } \Delta w_k, \tag{15}$$

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

$$\varphi'(0) < 0 \text{ when we set } \rho_k := s_k^T z_k / m, \sigma_k \in (0, 1).$$

The sequence $\{\varphi_k\}$ is monotonically decreasing.

[1] The most classic global algorithm for Euclidean IPM is [El-Bakry et al., 1996].

Global Convergence Theorem

Assumptions:

1. the functions $f(x), h(x), g(x)$ are smooth; the set $\{\operatorname{grad} h_i(x)\}_{i=1}^{l}$ is linearly independent in $T_x\mathbb{M}$ 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.

Theorem (Global Convergence, L. 2022)

Let $\{\sigma_k\} \subset (0,1)$ bounded away from zero and one. 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 (RCOP).

Dominant cost — solving Newton equation

Dominant cost is to solve

$$\nabla F(w)\Delta w = -F(w) + \mu\hat{e}, \tag{16}$$

where

$$F(w) = \begin{pmatrix} F_x \triangleq \operatorname{grad}_x \mathcal{L}(x,y,z) \\ F_y \triangleq h(x) \\ F_z \triangleq g(x) + s \\ F_s \triangleq ZSe \end{pmatrix}, \quad \hat{e} \triangleq \begin{pmatrix} 0_x \\ 0 \\ 0 \\ e \end{pmatrix}. \tag{17}$$

Thus, we need to solve the following linear system on $T_x\mathbb{M} \times \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^m$:

$$\begin{pmatrix} \operatorname{Hess}_x \mathcal{L}(w)\Delta x + H_x\Delta y + G_x\Delta z \\ H_x^*\Delta x \\ G_x^*\Delta x + \Delta s \\ Z\Delta s + S\Delta z \end{pmatrix} = \begin{pmatrix} -F_x \\ -F_y \\ -F_z \\ -F_s + \mu e \end{pmatrix}. \tag{18}$$

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

Condensed form of Newton equation

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

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}, \tag{19}$$

where

$$\begin{aligned} \mathcal{A}_w &\triangleq \operatorname{Hess}_x \mathcal{L}(w) + G_x S^{-1} Z G_x^*, \\ c &:= -F_x - G_x S^{-1} \left(Z F_z + \mu e - F_s \right), \quad q := -F_y. \end{aligned} \tag{20}$$

- $\mathcal{A}_w$ is self-adjoint (but may indefinite) on $T_x\mathbb{M}$.
- $\mathcal{T}$ is self-adjoint (but may indefinite) on $T_x\mathbb{M} \times \mathbb{R}^l$. This is a saddle point problems on Hilbert space.
- The Riemannian situation leaves us with no explicit matrix form available.
- A simple approach is to first find the representing matrix $\hat{\mathcal{T}}$. (Expensive !)

Krylov subspace methods on Tangent space

An ideal approach is to use iterative methods, such as **Krylov subspace methods** (e.g., Conjugate Gradients method [Boumal, 2022, Chapter 6.3]), on $T_x\mathbb{M} \times \mathbb{R}^l$ directly.

For simplicity, we consider the case of only inequality constraints, where Δy vanishes and only

$$\mathcal{A}_w \Delta x = c \text{ on } T_x\mathbb{M} \tag{21}$$

needs to be solved.

Krylov subspace methods on Tangent space

An ideal approach is to use iterative methods, such as **Krylov subspace methods** (e.g., Conjugate Gradients method [Boumal, 2022, Chapter 6.3]), on $T_x\mathbb{M} \times \mathbb{R}^l$ directly.

For simplicity, we consider the case of only inequality constraints, where Δy vanishes and only

$$\mathcal{A}_w \Delta x = c \text{ on } T_x\mathbb{M} \tag{21}$$

needs to be solved.

- It only needs to call an abstract linear operator $v \mapsto \mathcal{A}_w v$. (matrix-vector product)
- All the iterates v_k are in $T_x\mathbb{M}$.
- Since operator $\mathcal{A}_w$ is self-adjoint but indefinite, we use Conjugate Residual (CR) method to solve it.

The discussion of above can be naturally extended to the general case.

Numerical Experiments

We compare with the other Riemannian methods:[2]

- RALM : Riemannian augmented Lagrangian method.
- REPM(LQH) : Riemannian exact penalty method with smoothing function LQH.
- REPM(LSE) : Riemannian exact penalty method with smoothing function LSE.
- RSQP : Riemannian sequential quadratic programming.
- RIPM (Our method): Riemannian interior point method.

KKT residual is defined by

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

where Manvio measures the violation of manifold constraints.

[2]The numerical experiments were performed in Matlab R2022a on a computer equipped with an Intel Core i7-10700 at 2.90GHz with 16GB of RAM.

Problem I — Nonnegative Low Rank Matrix Approximation (NLRM)

Problem I

[Song and Ng, 2020] proposed

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

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

Data setting:

```
B = rand(m, r);
C = rand(r, n);
A = B*C+sigma*randn(m,n);
```

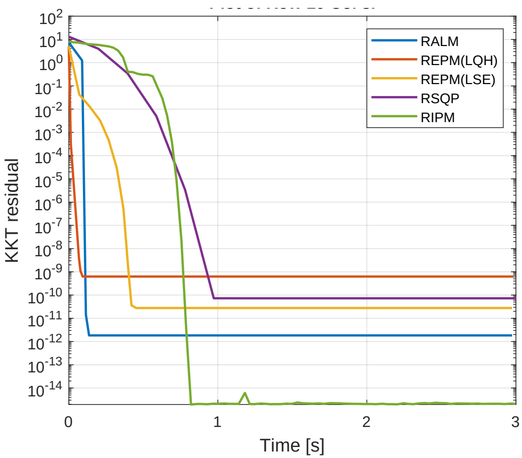


Figure: $m = 10, n = 8, r = 3$ and $\sigma = 0.01$.

Problem II — Projection onto nonnegative Stiefel manifold

Problem II

Given $C \in \mathbb{R}^{n\times k}$, we consider

$$\min_{X\in \mathrm{St}(n,k)} \|X - C\|_F^2, \quad \text{s.t. } X \geq 0, \tag{Model_St}$$

which can be equivalently reformulated [Jiang et al., 2022, Lemma 2.1] into

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

Here,

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

Problem II — Projection onto nonnegative Stiefel manifold

- For each Model, we conducted 20 random trials.
- Each experiment terminated successfully if a solution with KKT residual $\epsilon_{kkt} = 10^{-6}$ was found.
- It failed if the maximum iteration 10,000 or maximum time 600 [s] was reached.[3]

Table: Model_St

(n,k)	(60,12)			(70,14)		
	Rate	Time [s]	Iter.	Rate	Time [s]	Iter.
RALM	**1**	**4.097**	34	**1**	**6.234**	37
REPM(LQH)	0	-	-	0	-	-
REPM(LSE)	0	-	-	0	-	-
RSQP	0.65	78.02	7	0.85	166.1	7
RIPM	**1**	**5.555**	32	**1**	**7.574**	33

Table: Model_Ob

(n,k)	(60,12)			(70,14)		
	Rate	Time [s]	Iter.	Rate	Time [s]	Iter.
RALM	0.6	5.725	49	0.6	8.223	52
REPM(LQH)	0	-	-	0	-	-
REPM(LSE)	0	-	-	0	-	-
RSQP	0.7	44.46	5	0.5	91.38	5
RIPM	**1**	**7.134**	23	**1**	**9.268**	24

[3] The success rate (Rate) over the total number of trials, the average time in seconds (Time [s]) and the average iteration number (Iter.) among the successful trials.

Riemannian Interior Point Methods (RIPM)

Concluding remarks

We consider

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

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

Contributions:

1. We proposed a Riemannian version of the interior point method.
2. We proved the local superlinear/quadratic and global convergence.

Future Work

Future Work

1. **Preconditioner for linear operator equation.** Recall that we use Krylov subspace methods to solve

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}, \tag{22}$$

where $\mathcal{A}_w \triangleq \mathrm{Hess}_x\, \mathcal{L}(w) + G_x S^{-1} Z G_x^*$. Due to the strictly complementary condition, as $k \to \infty$, the values of $S_k^{-1} Z_k$ display a huge difference of magnitude. Hence, $\Theta := G_x S^{-1} Z G_x^*$ makes $\mathcal{T}$ very ill-conditioned.

One possible way is to find another nonsingular operator $\mathcal{P}$ such that the condition number of new operator $\mathcal{P}^{-1}\mathcal{T}$ becomes smaller.

Future Work

1. **Preconditioner for linear operator equation.** Recall that we use Krylov subspace methods to solve

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}, \tag{22}$$

where $\mathcal{A}_w \triangleq \mathrm{Hess}_x\, \mathcal{L}(w) + G_x S^{-1} Z G_x^*$. Due to the strictly complementary condition, as $k \to \infty$, the values of $S_k^{-1} Z_k$ display a huge difference of magnitude. Hence, $\Theta := G_x S^{-1} Z G_x^*$ makes $\mathcal{T}$ very ill-conditioned.

One possible way is to find another nonsingular operator $\mathcal{P}$ such that the condition number of new operator $\mathcal{P}^{-1}\mathcal{T}$ becomes smaller.

2. **Sophisticated global strategies.** Recall that now we use
   - Merit function $\varphi(w) = \|F(w)\|^2$. (too simple)
   - Backtracking for line-search.

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

Reference I

[Bergmann and Herzog, 2019] Bergmann, R. and Herzog, R. (2019).
Intrinsic formulation of KKT conditions and constraint qualifications on smooth manifolds.
SIAM Journal on Optimization, 29(4):2423–2444.

[Boumal, 2022] Boumal, N. (2022).
An introduction to optimization on smooth manifolds.
To appear with Cambridge University Press.

[Do Carmo and Flaherty Francis, 1992] Do Carmo, M. P. and Flaherty Francis, J. (1992).
Riemannian geometry, volume 6.
Springer.

[El-Bakry et al., 1996] El-Bakry, A., Tapia, R. A., Tsuchiya, T., and Zhang, Y. (1996).
On the formulation and theory of the Newton interior-point method for nonlinear programming.
Journal of Optimization Theory and Applications, 89(3):507–541.

[Fernandes et al., 2017] Fernandes, T. A., Ferreira, O. P., and Yuan, J. (2017).
On the superlinear convergence of Newton's method on Riemannian manifolds.
Journal of Optimization Theory and Applications, 173(3):828–843.

Reference II

[Ferreira and Silva, 2012] Ferreira, O. P. and Silva, R. C. (2012).
Local convergence of Newton's method under a majorant condition in Riemannian manifolds.
IMA Journal of Numerical Analysis, 32(4):1696–1713.

[Forsgren et al., 2002] Forsgren, A., Gill, P. E., and Wright, M. H. (2002).
Interior methods for nonlinear optimization.
SIAM Review, 44(4):525–597.

[Jiang et al., 2022] Jiang, B., Meng, X., Wen, Z., and Chen, X. (2022).
An exact penalty approach for optimization with nonnegative orthogonality constraints.
Mathematical Programming, pages 1–43.

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

[Obara et al., 2022] Obara, M., Okuno, T., and Takeda, A. (2022).
Sequential quadratic optimization for nonlinear optimization problems on riemannian manifolds.
SIAM Journal on Optimization, 32(2):822–853.

Reference III

[Schiela and Ortiz, 2020] Schiela, A. and Ortiz, J. (2020).
An SQP method for equality constrained optimization on manifolds.
arXiv preprint arXiv:2005.06844.

[Song and Ng, 2020] Song, G.-J. and Ng, M. K. (2020).
Nonnegative low rank matrix approximation for nonnegative matrices.
Applied Mathematics Letters, 105:106300.

[Yamakawa and Sato, 2022] Yamakawa, Y. and Sato, H. (2022).
Sequential optimality conditions for nonlinear optimization on Riemannian manifolds and a globally convergent augmented lagrangian method.
Computational Optimization and Applications, pages 1–25.

[Yamashita and Yabe, 1996] Yamashita, H. and Yabe, H. (1996).
Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization.
Mathematical Programming, 75(3):377–397.

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

The End

Questions? Comments?

Appendix.

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

1. **EIPM is a special case of RIPM** when $\mathbb{M} \equiv \mathbb{R}^n$ or $\mathbb{R}^{n\times k}$.
2. **RIPM can solve a condensed equation (23) of smaller order.**

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}, \tag{23}$$

   For example, the Stiefel manifold can be used as the equality constraints; i.e., we set $h : \mathbb{M} \equiv \mathbb{R}^{n\times k} \to \mathrm{Sym}(k)$, where $h(X) = X^\top X - I_k$. Here, EIPM requires us to solve (23) of order $nk + k(k+1)/2$.
   But RIPM only requires us to solve a problem of order $nk - k(k+1)/2$, i.e., the dimension of $\mathrm{St}(n, k)$.
3. **Not all manifolds are equivalent to the smooth equality constraints.**
   For example, $\mathrm{rank}(X) = r$ is not continuous, we can not apply EIPM.

Riemannian Newton method

Riemannian Newton method: Consider

$$F(x) = 0. \tag{24}$$

Solve a linear system on $T_{x_k}M \ni v_k$:

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

then $x_{k+1} = R_{x_k}(v_k)$.

Standard Newton assumptions & Local Convergence Results:

$$\left.\begin{array}{l}\left.\begin{array}{l}\text{(N1)There exists } x^* : F(x^*) = 0. \\ \text{(N2)}\nabla F(x^*) \text{ is nonsingular operator.}\end{array}\right\} \Rightarrow \text{superlinear[Fernandes et al., 2017]} \\ \text{(N3)}\nabla F \text{ is locally Lipschitz cont. at } x^*.\end{array}\right\} \Rightarrow \text{quadratic[Ferreira}$$

1. Existence. There exists w^* satisfying the KKT conditions.
2. Smoothness. The functions f, g, h are smooth on $\mathcal{M}$.
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 $T_{x^*}\mathcal{M}$.
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 $\xi \in T_{x^*}\mathbb{M}$ 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.

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

- Introducing the **perturbed** complementary equation,

$$Z\Delta s + S\Delta z = -ZSe + \mu e, \tag{25}$$

  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

$$\alpha_k := \min\left\{1, \gamma_k \min_i \left\{-\frac{(s_k)_i}{(\Delta s_k)_i} \mid (\Delta s_k)_i < 0\right\}, \gamma_k \min_i \left\{-\frac{(z_k)_i}{(\Delta z_k)_i} \mid (\Delta z_k)_i < 0\right\}\right\}, \tag{26}$$

  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.,

$$w(\alpha) := \bar{R}_w(\alpha \Delta w), \tag{27}$$

for some step length $\alpha > 0$.

By introducing

$$w(\alpha) = (x(\alpha), y(\alpha), z(\alpha), s(\alpha)), \tag{28}$$

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:

$$f^I(\alpha) := \min(Z(\alpha)S(\alpha)e) - \gamma\tau_1 \frac{z(\alpha)^T s(\alpha)}{m}, \tag{29}$$

$$f^{II}(\alpha) := z(\alpha)^T s(\alpha) - \gamma\tau_2 \|F(w(\alpha))\|. \tag{30}$$

For $i = I, II$, let

$$\alpha^i := \max_{\alpha \in (0,1]} \left\{ \alpha : f^i(t) \geq 0, \text{ for all } t \in (0, \alpha] \right\}. \tag{31}$$

Global RIP Algorithm

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

$$\mu_k = z_k^T s_k / m \tag{32}$$

and by

$$\nabla F(w)\Delta w = -F(w) + \sigma_k \mu_k \hat{e}. \tag{33}$$

2. Step length selection.
   1. Centrality conditions: Choose $1/2 < \gamma_k < \gamma_{k-1} < 1$; compute $\alpha^i, i = I, II$, from (31); and let

   $$\bar{\alpha}_k = \min(\alpha^I, \alpha^{II}). \tag{34}$$

   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

   $$\varphi(\bar{R}_{w_k}(\alpha_k \Delta w_k)) - \varphi(w_k) \le \alpha_k \beta \langle \operatorname{grad} \varphi_k, \Delta w_k \rangle. \tag{35}$$

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

$$\|A_x\| = \|\hat{A}_x\|_2 \leq \|\hat{A}_x\|_F. \tag{36}$$

Lemma (L. 2022)

$$x \mapsto \|\widehat{\mathrm{Hess}\, f(x)}\| \tag{37}$$

is a continuous scalar field on $\mathbb{M}$. It is true for all h_i, g_i.

$$x \mapsto \|H_x\| \text{ and } x \mapsto \|G_x\| \tag{38}$$

are continuous scalar field on $\mathbb{M}$.

Global Convergence Theorem

This theorem, now, is only proved under exponential map exp.

Lemma (Gauss [Do Carmo and Flaherty Francis, 1992, Lemma 3.5])

Let $p \in \mathcal{M}$ *and let* $v \in T_p\mathcal{M}$ *such that* $\exp_p(v)$ *is well defined. Let* $w \in T_p\mathcal{M} \approx T_v(T_p\mathcal{M})$*. Then*

$$\left\langle \mathcal{D}\exp_p(v)[v], \mathcal{D}\exp_p(v)[w] \right\rangle = \langle v, w \rangle. \tag{39}$$

Conjugate Gradients (CG) on a tangent space

Input: positive definite map H on $\mathrm{T}_x\mathcal{M}$ and $b \in \mathrm{T}_x\mathcal{M}$, $b \neq 0$

Set $v_0 = 0, r_0 = b, p_0 = r_0$

For $n = 1, 2, \ldots$

Compute Hp_{n-1} (this is the only call to H)

$$\alpha_n = \frac{\|r_{n-1}\|_x^2}{\langle p_{n-1}, Hp_{n-1}\rangle_x}$$

$$v_n = v_{n-1} + \alpha_n p_{n-1}$$

$$r_n = r_{n-1} - \alpha_n H p_{n-1}$$

If $r_n = 0$, **output** $s = v_n$: the solution of $Hs = b$

$$\beta_n = \frac{\|r_n\|_x^2}{\|r_{n-1}\|_x^2}$$

$$p_n = r_n + \beta_n p_{n-1}$$

1. Exactly the same in form of usual CG.
2. Every vectors v_n, r_n, p_n belong to tangent space $V \equiv \mathrm{T}_x\mathcal{M}$.
3. Converges very fast if H is PD with small condition number.

An Intuitive Barrier Method on Manifolds

Consider

$$\min_{x \in \mathbb{M}} \quad f(x) \quad \text{s.t.} \quad c(x) \geq 0. \tag{RCOP_Ineq}$$

Its logarithmic barrier function is

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

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

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

Barrier Method on Manifolds

1. Set $x_0 \in \mathbb{M}$ 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.

C

onsider 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):

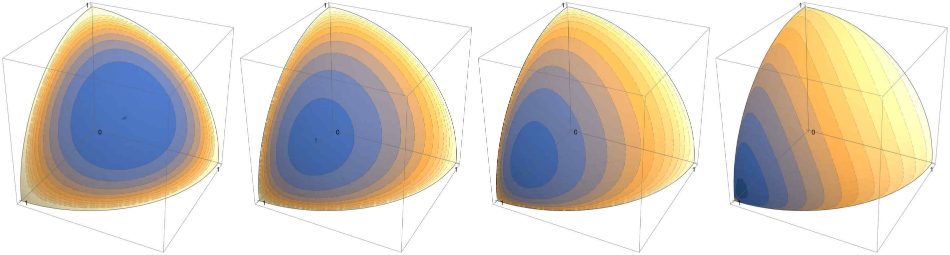
$$\operatorname{grad} f(x) = (I_n - xx^T)a = [0, 2, 1]^T.$$

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.

An Intuitive 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.

An Intuitive 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].

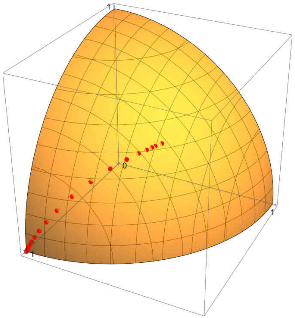


Figure: Iterates x_k of barrier method for (SP).

An Intuitive 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 $\operatorname{grad} B(x_\mu;\mu) = 0$.

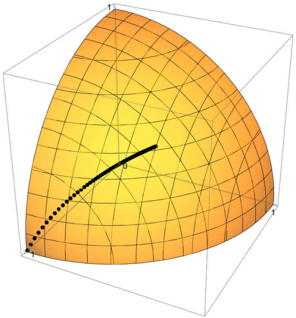


Figure: Existence of a central path for (SP).