Riemannian Interior Point Methods for Constrained Optimization on Manifolds

*Zhijian Lai Akiko Yoshise

University of Tsukuba

10th International Congress on Industrial and Applied Mathematics
Tokyo - Aug. 24, 2023

Interior Point Methods also Succeed in Nonconvex Case

Interior Point Methods is ***default algorithm*** *for constrained nonlinear optimization in MATLAB:*

```
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub)
```

Finds the minimum of a problem specified by

$$\min_x f(x) \text{ such that } \begin{cases} c(x) \leq 0 \\ ceq(x) = 0 \\ A \cdot x \leq b \\ Aeq \cdot x = beq \\ lb \leq x \leq ub, \end{cases}$$

- Byrd, R. H., J. C. Gilbert, and J. Nocedal. *"A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming."* Math. Program., 2000.
- Byrd, R.H., Mary E. Hribar, and Jorge Nocedal, *"An Interior Point Algorithm for Large-Scale Nonlinear Programming."* SIAM J. Optim., 1999.
- Waltz, R. A., J. L. Morales, J. Nocedal, and D. Orban. *"An interior algorithm for nonlinear optimization that combines line search and trust region steps."* Math. Program., 2006.

Overview

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

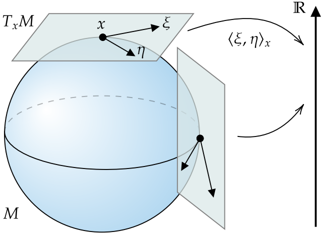


Figure: Example of sphere manifold $M = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ and $T_xM = \{v \in \mathbb{R}^n : \langle x, v \rangle = 0\}$.

Riemannian Optimization (RO)

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

$$\min_{x \in M} f(x) \qquad \text{(RO)}$$

where M is a Riemannian manifold.

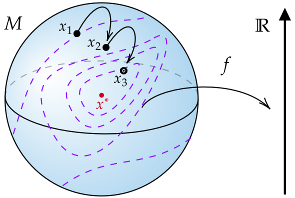


Figure: Iteration on unit sphere.

In a popular Riemannian optimization solver **"Manopt"**:

- **40+ available manifolds** M, e.g.,
  - Stiefel manifold, $\mathrm{St}(n,k) = \{X \in \mathbb{R}^{n \times k} : X^\top X = I\}$.
  - Fixed rank manifold, $\mathbb{R}_r^{m \times n} = \{X \in \mathbb{R}^{m \times n} : \mathrm{rank}(X) = r\}$.
- **9 available Riemannian algorithms**, e.g.,
  - steepest decent
  - conjugate gradient
  - trust region
  - BFGS and more.

Advantages of Riemannian Optimization (RO)

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

$$\min_{x \in M} f(x) \qquad \text{(RO)}$$

where M is a Riemannian manifold.

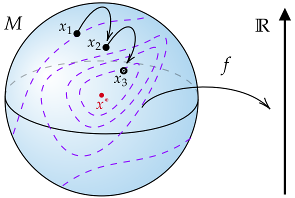


Figure: Iteration on unit sphere.

Advantages of (RO):

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

Applications of (RO)

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

$$\min_{X\in\mathrm{St}(n,k)} -\mathrm{trace}(X^\top A^\top A X)$$

- Matrix Completion on fixed rank manifold,
  $\mathbb{R}_r^{m\times n} = \{X \in \mathbb{R}^{m\times n} : \mathrm{rank}(X) = r\}$.

$$\min_{X\in\mathbb{R}_r^{m\times n}} \sum_{(i,j)\in\Omega} (X_{ij} - A_{ij})^2$$

$\rightsquigarrow$ **We can use the solver "Manopt" directly.**

More Requirements in Practical Applications

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

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

- Nonnegative Matrix Completion on fixed rank manifold,
  $\mathbb{R}_r^{m\times n} = \{X \in \mathbb{R}^{m\times n} : \mathrm{rank}(X) = r\}$.

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

$\rightsquigarrow$ **Um..., can we use the solver "Manopt" directly? Answer is NO.**

Some Limitations of Riemannian Optimization (RO)

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

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

where M is a Riemannian manifold.

Some limitations of (RO):

1. (RO) requires that the *whole* feasible region to your problem forms a manifold. Unfortunately, the much common scenario is:

   "some of the constraints form a manifold that already exists in 'Manopt', but we have additional constraints."

2. If your feasible region *happens* to form a new manifold (not available in "Manopt"), you will need to find its geometric tools by yourself.

$\rightsquigarrow$ **To address these challenges, We aims to develop a new model instead of (RO).**

New Topic — Riemannian Constrained Optimization Problem

We consider

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

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

New Topic — Riemannian Constrained Optimization Problem

We consider

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

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, i.e., the advantages of (RO) are preserved.
2. Flexible! It only requires that a *portion* of the constraints in your problem form a manifold.
3. Easy programming! Because we can still use the existing geometric tools of M if it is available in “Manopt”.

New Topic — Riemannian Constrained Optimization Problem

We consider

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

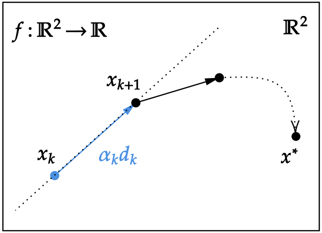
where $f : M \to \mathbb{R}$, $h : M \to \mathbb{R}^l$, and $g : M \to \mathbb{R}^m$.

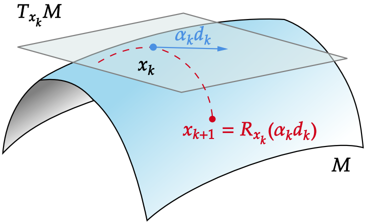
Riemannian version of classical algorithms:

- 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]
- $\rightsquigarrow$ In this talk, we propose *Riemannian version of Interior Point Method.*

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

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



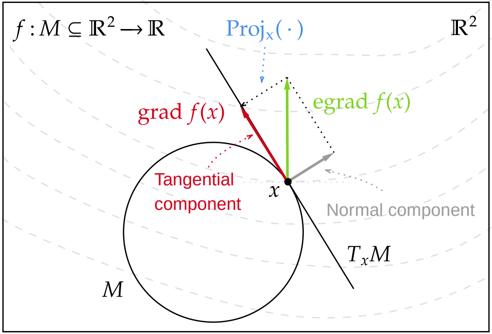


Euclidean	Riemannian
$x_{k+1} = x_k + \alpha_k d_k$	$x_{k+1} = R_{x_k}(\alpha_k d_k)$

Q2: In which direction do we move?? Riemannian Gradient!

For an embedded submanifold M, **Riemannian gradient of** $f : M \to \mathbb{R}$ is the orthogonal projection onto T_xM of the Euclidean gradient,

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



Supplementary: Vector fields on Manifolds

A **vector field** is a mapping F defined on M such that $F(x) \in T_xM$ for all $x \in M$.

Riemannian gradient,

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

is a vector field generated by scalar field $f : M \to \mathbb{R}$.

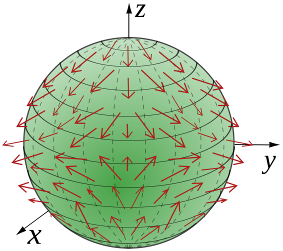


Figure: A vector field on a unit sphere. Source: Wikipedia.

Covariant derivative & Hessian & Riemannian 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 & Riemannian 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 $F(x^*) = 0_{x^*}$.

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

$$\nabla F(x_k) v_k = -F(x_k), \tag{1}$$

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

Formulation of RIPM

We consider

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

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

Lagrangian function is

$$\mathcal{L}(x, y, z) \triangleq f(x) + y^T h(x) + z^T g(x). \tag{2}$$

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

$$\begin{cases} \operatorname{grad}_x \mathcal{L}(x,y,z) = 0_x, \\ h(x) = 0, \\ g(x) \le 0, \\ Zg(x) = 0, \ (Z := \operatorname{diag}(z_1,\ldots,z_m)) \\ z \ge 0. \end{cases} \tag{3}$$

Definition (KKT Vector Field, L. 2022)

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

$$F(w) \triangleq \begin{pmatrix} \operatorname{grad}_x \mathcal{L}(x,y,z) \\ h(x) \\ g(x) + s \\ ZSe \end{pmatrix} = 0_w := \begin{pmatrix} 0_x \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{ and } (z,s) \ge 0, \tag{4}$$

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

$$H_x : \mathbb{R}^l \to T_x M, \quad H_x v \triangleq \sum_{i=1}^{l} v_i \operatorname{grad} h_i(x). \tag{5}$$

Hence, the adjoint operator is

$$H_x^* : T_x M \to \mathbb{R}^l, \quad H_x^* \xi = \left[\langle \operatorname{grad} h_1(x), \xi \rangle_x, \cdots, \langle \operatorname{grad} h_l(x), \xi \rangle_x \right]^T. \tag{6}$$

Lemma (L. 2022)

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

$$\nabla F(w) \Delta w = \begin{pmatrix} \operatorname{Hess}_x \mathcal{L}(w) \Delta x + H_x \Delta y + G_x \Delta z \\ H_x^* \Delta x \\ G_x^* \Delta x + \Delta s \\ Z \Delta s + S \Delta z \end{pmatrix}, \tag{7}$$

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

Riemannian Interior Point Method (RIPM)

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

$$\nabla F(w_k)\Delta w_k = -F(w_k) + \mu_k \hat{e}, \tag{8}$$

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:

$$w_{k+1} = \bar{R}_{w_k}(\alpha_k \Delta w_k). \tag{9}$$

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

Merit function: Choose $\varphi(w) \triangleq \|F(w)\|^2$.
Backtracking for step size α_k:

1. Centrality conditions.
2. Sufficient decreasing condition.

With a slight abuse of notation, we also let

$$\varphi(\alpha) \triangleq \varphi(\underbrace{\bar{R}_{w_k}(\alpha \Delta w_k)}_{\text{new iterate}}) \quad \text{for fixed } w_k \text{ and } \Delta w_k, \tag{10}$$

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

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

Global Line Search RIPM Algorithm

Merit function: Choose $\varphi(w) \triangleq \|F(w)\|^2$.
Backtracking for step size α_k:

1. Centrality conditions.
2. Sufficient decreasing condition.

With a slight abuse of notation, we also let

$$\varphi(\alpha) \triangleq \varphi(\underbrace{\bar{R}_{w_k}(\alpha \Delta w_k)}_{\text{new iterate}}) \quad \text{for fixed } w_k \text{ and } \Delta w_k, \tag{10}$$

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 $\varphi'(0) < 0$ if we set $\rho_k := s_k^T z_k / m, \sigma_k \in (0, 1)$. Then, $\{\varphi_k\}$ is monotonically decreasing.

Global Convergence

C.f. 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*. J Optim Theory Appl, 1996.

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_xM 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).

Numerical Experiments

We compare with the other Riemannian methods:[1]

- 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

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

where Manvio measures the violation of manifold constraints.

[1]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

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

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

Data setting:

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

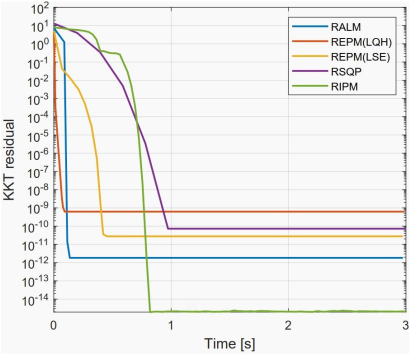


Figure: $m = 10, n = 8, r = 3$ and $\sigma = 0.01$.

Problem II — Projection onto Nonnegative Stiefel Manifold

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

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

which can be equivalently reformulated into

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

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 solution with KKT residual $< 10^{-6}$ was found.
- It failed if the maximum iteration 10,000 or maximum time 600 [s] was reached.[2]

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

[2] 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 IPM (RIPM) vs. Euclidean IPM (EIPM)

Riemannian Constrained Optimization Problem

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

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 IPM (RIPM) vs. Euclidean IPM (EIPM)

1. RIPM inherits the advantages of Riemannian optimization.
2. EIPM is a special case of RIPM when $\mathbb{M} = \mathbb{R}^n$ or $\mathbb{R}^{n \times k}$.
3. RIPM solves Newton equation (12) of smaller order on $\mathrm{T}_x\mathbb{M} \times \mathbb{R}^l$:

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}. \qquad (11)$$

4. RIPM can solve some problems that EIPM cannot. For example, $\operatorname{rank}(X) = r$ is not continuous, we can not apply EIPM.

Concluding Remarks

Riemannian Constrained Optimization Problem

We consider

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

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

Our contributions:

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

Future work:

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

Reference I

[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.

[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.

Reference II

[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.

[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.

Reference III

[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.

The End

Questions? Comments?

Appendix.

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

RIPM can solve a condensed equation (12) of smaller order.

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}, \qquad (12)$$

For example, the Stiefel manifold can be used as the equality constraints; i.e., we set $h : M \equiv \mathbb{R}^{n\times k} \to \mathrm{Sym}(k)$, where $h(X) = X^\top X - I_k$. Here, EIPM requires us to solve (12) 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 $\mathrm{St}(n, k)$.

Riemannian Newton method: Consider

$$F(x) = 0. \tag{13}$$

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:

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

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.

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

- Introducing the **perturbed** complementary equation,

$$Z\Delta s + S\Delta z = -ZSe + \mu e, \tag{14}$$

  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

$$\alpha_k := \min\left\{1, \gamma_k \min_i \left\{-\frac{(s_k)_i}{(\Delta s_k)_i} \mid (\Delta s_k)_i < 0\right\}, \gamma_k \min_i \left\{-\frac{(z_k)_i}{(\Delta z_k)_i} \mid (\Delta z_k)_i < 0\right\}\right\}, \tag{15}$$

  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.,

$$w(\alpha) := \bar{R}_w(\alpha \Delta w), \tag{16}$$

for some step length $\alpha > 0$.

By introducing

$$w(\alpha) = (x(\alpha), y(\alpha), z(\alpha), s(\alpha)), \tag{17}$$

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:

$$f^I(\alpha) := \min(Z(\alpha)S(\alpha)e) - \gamma\tau_1 \frac{z(\alpha)^T s(\alpha)}{m}, \qquad (18)$$

$$f^{II}(\alpha) := z(\alpha)^T s(\alpha) - \gamma\tau_2 \|F(w(\alpha))\|. \qquad (19)$$

For $i = I, II$, let

$$\alpha^i := \max_{\alpha \in (0,1]} \left\{ \alpha : f^i(t) \geq 0, \text{ for all } t \in (0, \alpha] \right\}. \qquad (20)$$

Global RIP Algorithm

1. Choose $\sigma_k \in (0,1)$; for w_k, compute the perturbed Newton direction Δw_k with
$$\mu_k = z_k^T s_k / m \tag{21}$$
and by
$$\nabla F(w) \Delta w = -F(w) + \sigma_k \mu_k \hat{e}. \tag{22}$$
2. Step length selection.
    1. Centrality conditions: Choose $1/2 < \gamma_k < \gamma_{k-1} < 1$; compute $\alpha^i, i = I, II$, from (20); and let
    $$\bar{\alpha}_k = \min(\alpha^I, \alpha^{II}). \tag{23}$$
    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
    $$\varphi(\bar{R}_{w_k}(\alpha_k \Delta w_k)) - \varphi(w_k) \leq \alpha_k \beta \langle \operatorname{grad} \varphi_k, \Delta w_k \rangle. \tag{24}$$
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,

$$\|A_x\| = \|\hat{A}_x\|_2 \leq \|\hat{A}_x\|_F. \tag{25}$$

Lemma (L. 2022)

$$x \mapsto \|\widehat{\operatorname{Hess} f(x)}\| \tag{26}$$

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

$$x \mapsto \|H_x\| \text{ and } x \mapsto \|G_x\| \tag{27}$$

are continuous scalar field on 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 $p \in M$ *and let* $v \in T_pM$ *such that* $\exp_p(v)$ *is well defined. Let* $w \in T_pM \approx T_v(T_pM)$. *Then*

$$\left\langle \mathcal{D}\exp_p(v)[v], \mathcal{D}\exp_p(v)[w] \right\rangle = \langle v, w \rangle. \tag{28}$$

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)
    $\alpha_n = \frac{\|r_{n-1}\|_x^2}{\langle p_{n-1}, Hp_{n-1}\rangle_x}$
    $v_n = v_{n-1} + \alpha_n p_{n-1}$
    $r_n = r_{n-1} - \alpha_n Hp_{n-1}$
    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 $V \equiv T_x M$.
3. Converges very fast if H is PD with small condition number.

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

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

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

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

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

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

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

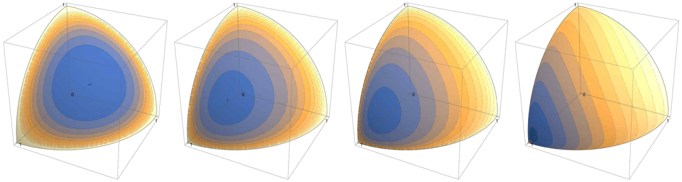
where $a = [-1, 2, 1]^T$. Its solution is $x^* = [1, 0, 0]^T$.

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

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

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

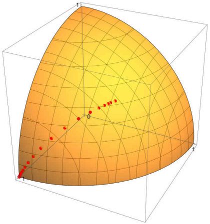


Figure: Iterates x_k of barrier method for (SP).

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

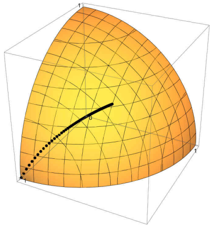


Figure: Existence of a central path for (SP).

Dominant cost — Newton equation

Dominant cost is to solve

$$\nabla F(w)\Delta w = -F(w) + \mu\hat{e}, \tag{29}$$

where

$$F(w) = \begin{pmatrix} F_x \triangleq \operatorname{grad}_x \mathcal{L}(x, y, z) \\ F_y \triangleq h(x) \\ F_z \triangleq g(x) + s \\ F_s \triangleq ZSe \end{pmatrix}, \quad \hat{e} \triangleq \begin{pmatrix} 0_x \\ 0 \\ 0 \\ e \end{pmatrix}. \tag{30}$$

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

$$\begin{pmatrix} \operatorname{Hess}_x \mathcal{L}(w)\Delta x + H_x\Delta y + G_x\Delta z \\ H_x^*\Delta x \\ G_x^*\Delta x + \Delta s \\ Z\Delta s + S\Delta z \end{pmatrix} = \begin{pmatrix} -F_x \\ -F_y \\ -F_z \\ -F_s + \mu e \end{pmatrix}. \tag{31}$$

Condensed form of Newton equation

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

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}, \tag{32}$$

where

$$\begin{aligned} \mathcal{A}_w &:= \operatorname{Hess}_x \mathcal{L}(w) + G_x S^{-1} Z G_x^*, \\ c &:= -F_x - G_x S^{-1} \left(Z F_z + \mu e - F_s \right), \quad q := -F_y. \end{aligned} \tag{33}$$

- $\mathcal{A}_w$ is self-adjoint (but may indefinite) on T_xM.
- $\mathcal{T}$ is self-adjoint (but may indefinite) on $T_xM \times \mathbb{R}^l$. 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 $\hat{\mathcal{T}}$ under some basis. (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), on $T_x M \times \mathbb{R}^l$ directly.

For simplicity, we consider the case of only inequality constraints, where Δy vanishes, thus we only needs to

$$\text{solve } \mathcal{A}_w \Delta x = c \text{ for } \Delta x \in T_x M. \tag{34}$$

- 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 $T_x M$.
- 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.