Riemannian Interior Point Methods

On the Global Convergence

*Zhijian Lai   Akiko Yoshise

University of Tsukuba

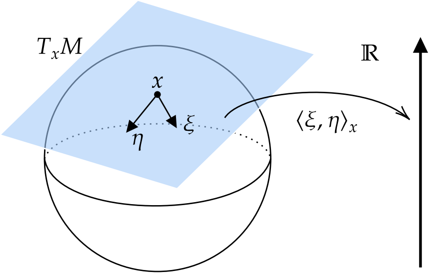
s2130117@s.tsukuba.ac.jp

JSIAM Annual Meeting 2022
September 8, 2022

Presentation Overview

Riemannian manifold

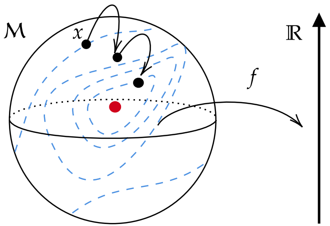
A **Riemannian manifold** M is a smooth (locally linearized) set equipped with a smoothly-varying inner product $\langle \cdot, \cdot \rangle_x$ on the tangent spaces.



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.

Examples of Applications

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

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

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

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

3. And many more.

More Requirements in Applications

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

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

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

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

3. And many more.

New Topic — Riemannian Constrained Optimization

We consider

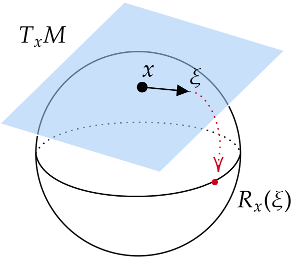
$$\min_{x \in \mathbb{M}} \quad f(x) \qquad \text{(RCOP)}$$
$$\text{s.t.} \quad h(x) = 0, \text{ and } g(x) \leq 0,$$

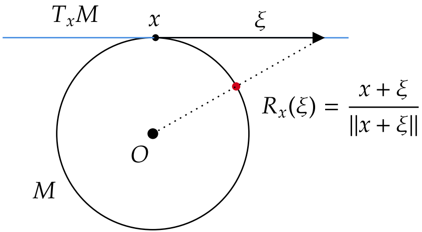
where $f : \mathbb{M} \to \mathbb{R}$, $h : \mathbb{M} \to \mathbb{R}^l$, and $g : \mathbb{M} \to \mathbb{R}^m$.

Riemannian version of optimality conditions:
KKT conditions; second-order necessary and sufficient conditions [YZS14]; More constraint qualifications [BH19]; Sequential optimality conditions [YS22].

Very little research (2019-)
Augmented Lagrangian Method [LB19, YS22]; Exact Penalty Method [LB19]; Sequential Quadratic Method [SO20, OOT20].
How about Interior Point Method?

A **retraction** R yields a map $R_x : T_xM \to M$ for any x.



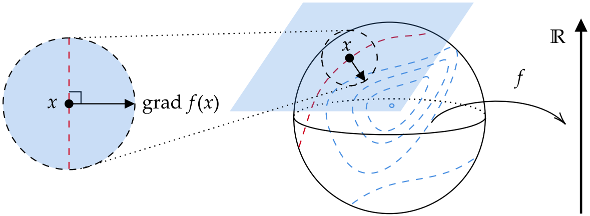


Euclidean	Riemannian
$x_{k+1} = x_k + \alpha_k \xi_k$	$x_{k+1} = R_{x_k}(\alpha_k \xi_k)$

Riemannian gradient — a special vector field

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

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



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

Riemannian Hessian — linear operators on tangent spaces

Covariant derivative of a vector field F:

Riemannian connection

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

general vector field

Specially, if $F = \operatorname{grad} f$, then $\operatorname{Hess} f(x) := \nabla \operatorname{grad} f(x)$ is called **Riemannian Hessian**.

Riemannian Hessian — linear operators on tangent spaces

Covariant derivative of a vector field F:

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

Riemannian connection (∇)

general vector field (F)

Specially, if $F = \operatorname{grad} f$, then $\operatorname{Hess} f(x) := \nabla \operatorname{grad} f(x)$ is called **Riemannian Hessian**.

Riemannian Newton method: To find $x^* \in M$ 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 Riemannian Interior Point Method (RIPM)

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

Lagrangian function is

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

$x \mapsto \mathcal{L}(x, y, z)$ 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: F

Riemannian KKT conditions [LB19] are

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

Definition (KKT Vector Field, L.2022)

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

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

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

Formulation of $\nabla F(w)$

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

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

Hence, the adjoint operator of H 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 $\Delta w = (\Delta x, \Delta y, \Delta s, \Delta z) \in T_x\mathbb{M} \times \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^m \equiv T_w\mathscr{M}$.

Prototype — Riemannian Interior Point Method (RIPM)

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

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

where $\hat{e} := (0_x, 0, 0, e)$.
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 $\rho_k \to 0$. Return to 1.

Theorem (Local Convergence, L. 2022)

Under some standard assumptions.

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

Sketch — Global Line Search of RIPM

The most classic global algorithm for Euclidean IPM is [EBTTZ96]

- **Merit function:**

$$\varphi(w) := \|F(w)\|^2 \quad \Rightarrow \quad \operatorname{grad}\varphi(w) = 2\nabla F(w)^* F(w). \tag{15}$$

Goal:

$$\text{Keep } (z_k, s_k) > 0 \text{ and let } \|F(w_k)\|^2 \to 0. \tag{16}$$

- **Step size selection:** Given Δw, we obtain α_k by
  1. *Two Centrality conditions.*
  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

$$\varphi(\bar{R}_{w_k}(\alpha_k \Delta w_k)) - \varphi(w_k) \leq \alpha_k \beta \langle \operatorname{grad}\varphi_k, \Delta w_k \rangle. \tag{17}$$

Sufficient decreasing

We denote a real to real function $\alpha \mapsto \varphi(\alpha)$ by

$$\varphi(\alpha) := \varphi(\bar{R}_w(\alpha \Delta w)), \tag{18}$$

then

$$\varphi'(0) = \mathcal{D}\varphi(\bar{R}_w(0))\,[\mathcal{D}\bar{R}_w(0)[\Delta w]] = \mathcal{D}\varphi(w)[\Delta w] = \langle \operatorname{grad} \varphi(w), \Delta w \rangle_w. \tag{19}$$

Armijo condition

Hence, Armijo condition:

$$\varphi(\bar{R}_{w_k}(\alpha_k \Delta w_k)) - \varphi(w_k) \leq \alpha_k \beta \langle \operatorname{grad} \varphi_k, \Delta w_k \rangle \tag{20}$$

is to say

$$\varphi_k(\alpha_k) - \varphi_k(0) \leq \alpha_k \beta \varphi_k'(0). \tag{21}$$

Descent Direction

If direction Δw is given as the solution of

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

then

$$\varphi'(0) = \langle \operatorname{grad}\varphi(w), \Delta w\rangle = 2(-\|F(w)\|^2 + \sigma\mu z^T s). \tag{22}$$

Lemma (L. 2022)

1. Δw_k *is a descent direction, i.e.,* $\langle \operatorname{grad}\varphi(w_k), \Delta w_k\rangle < 0$*, for merit function* φ *at* w_k *if*

$$\mu_k := s_k^T z_k/m, \quad \sigma_k \in (0,1).$$

2. *if Armijo condition is satisfied, then the sequence* $\{\varphi_k\}$ *is monotonically decreasing.*

Global Convergence Theorem

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

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

Assumptions

1. in the set $\Omega(0)$, 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. in any compact subset of $\Omega(0)$ where s is bounded away from zero, 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

Dominant cost is to solve

$$\nabla F(w)\Delta w = -F(w) + \rho\hat{e}, \tag{23}$$

where

$$F(w) = \begin{pmatrix} F_x := \operatorname{grad}_x \mathcal{L}(x,y,z) \\ F_y := h(x) \\ F_z := g(x) + s \\ F_s := ZSe \end{pmatrix}, \quad \hat{e} := \begin{pmatrix} 0_x \\ 0 \\ 0 \\ e \end{pmatrix}. \tag{24}$$

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

$$\begin{pmatrix} \operatorname{Hess}_x \mathcal{L}(w)\Delta x + H_x\Delta y + G_x\Delta z \\ H_x^*\Delta x \\ G_x^*\Delta x + \Delta s \\ Z\Delta s + S\Delta z \end{pmatrix} = \begin{pmatrix} -F_x \\ -F_y \\ -F_z \\ -F_s + \rho e \end{pmatrix}. \tag{25}$$

Condensed form of perturbed Newton Equation

By two substitutions:

$$\Delta s = Z^{-1}\left(\rho e - F_s - S\Delta z\right), \tag{26}$$

$$\Delta z = S^{-1}\left[Z\left(G_x^*\Delta x + F_z\right) + \rho e - F_s\right], \tag{27}$$

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

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}, \tag{28}$$

where

$$\begin{aligned} \mathcal{A}_w &:= \operatorname{Hess}_x \mathcal{L}(w) + G_x S^{-1} Z G_x^*, \\ c &:= -F_x - G_x S^{-1}\left(ZF_z + \rho e - F_s\right), \\ q &:= -F_y. \end{aligned} \tag{29}$$

Condensed form of perturbed Newton Equation

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + H_x \Delta y \\ H_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}. \tag{30}$$

Lemma (L. 2022)

Equivalence:

1. *If $(z, s) > 0$ holds, then $\nabla F(w)$ is nonsingular if and only if $\mathcal{T}$ is nonsingular.*

Symmetric linear system:

1. $\mathcal{A}_w := \mathrm{Hess}_x \mathcal{L}(w) + G_x S^{-1} Z G_x^*$ *is self-adjoint (i.e., $\mathcal{A}_w = \mathcal{A}_w^*$) on $T_x\mathbb{M}$.*
2. $\mathcal{T}$ *is self-adjoint (i.e., $\mathcal{T} = \mathcal{T}^*$) on product space $T_x\mathbb{M} \times \mathbb{R}^l$.*

If only the inequality constraint is present, then

$$\mathcal{T}(\Delta x) := \mathcal{A}_w \Delta x = c. \tag{31}$$

General solver — RIPM.m

Problem on Matrix Submanifold

We consider

$$\min_{X \in \mathbb{M}} \quad f(X)$$
$$\text{s.t.} \quad h(X) = O_{p\times q}, \text{ and } g(X) \leq O_{n\times k}, \tag{RCOP}$$

where $\mathbb{M} \subseteq \mathbb{R}^{r\times s}$ is a submanifold; $f : \mathbb{M} \to \mathbb{R}$, $h : \mathbb{M} \to \mathbb{R}^{p\times q}$, and $g : \mathbb{M} \to \mathbb{R}^{n\times k}$.

We established a general RIPM solver based on Manopt[1].

`RIPM.m` (L. 2022)

```
% function [x, cost, info, options] = RIPM(problem)
% function [x, cost, info, options] = RIPM(problem, x0)
% function [x, cost, info, options] = RIPM(problem, x0, options)
% function [x, cost, info, options] = RIPM(problem, [], options)
```

[1]Manopt, a matlab toolbox for optimization on manifolds.

Problem I — Computing projection onto $\mathrm{St}(n,k)_+$

Example (Problem I)

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

$$\min_{X\in \mathrm{St}(n,k)_+} \|X - C\|_F^2, \qquad \text{(Model\_St)}$$

which can be equivalently [JMWC22, Lemma 2.1] reformulated into

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

Here,

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

Problem I — Computing projection onto $\mathrm{St}(n,k)_+$

Example (Problem I)

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

$$\min_{X\in \mathrm{St}(n,k)_+} \|X - C\|_F^2, \qquad \text{(Model\_St)}$$

which can be equivalently [JMWC22, Lemma 2.1] reformulated into

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

Experiment setting:

- [JMWC22, Proposition 1] By choosing $X^* \in \mathrm{St}(n,k)_+$, we can construct a special C such that the solution is unique and equals to X^*.
- Define $gap := \frac{\|X^k - C\|_F}{\|X^* - C\|_F} - 1$. We test both (Model_St) and (Model_OB).

Problem I — Computing projection onto $\mathrm{St}(n,k)_+$

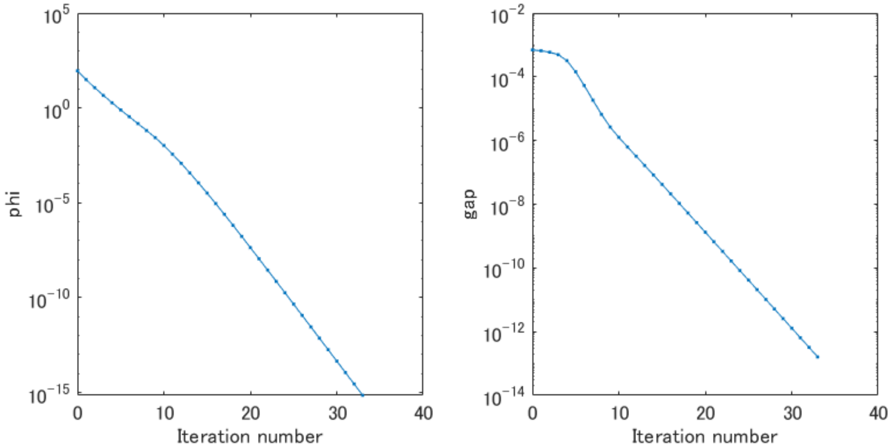


Figure: For $n = 10, k = 5$.

Problem I — Computing projection onto $\mathrm{St}(n,k)_+$

Table: For each (n,k), we test 5 instances. "gap" and "iter" are average values.

		Model_St		Model_OB	
n	$k = 0.1n$	gap	iter	gap	iter
30	3	5.03E-09	19	7.28E-09	23
50	5	3.38E-09	19	7.30E-09	22
70	7	4.19E-09	24	6.04E-09	21
90	9	6.98E-09	27	3.09E-09	20
110	11	8.71E-09	27	2.32E-09	22
130	13	8.18E-09	25	6.78E-09	21

Problem II — nonnegative low rank matrix approximation

Example (Problem II)

[SN20] proposed the nonnegative low-rank matrix approximation:

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

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

Experiment setting:

- `B = rand(m, r); C = rand(r, n); A = B*C; % original data`
  `Gaussian_Noise = sigma*randn(m,n); % zero mean and standard deviation` σ
  `A_test = A+Gaussian_Noise;`
- Define *relative_residual* $\stackrel{\text{def}}{=} \frac{\|A-X^k\|_F}{\|A\|_F}$.

Problem II — nonnegative low rank matrix approximation

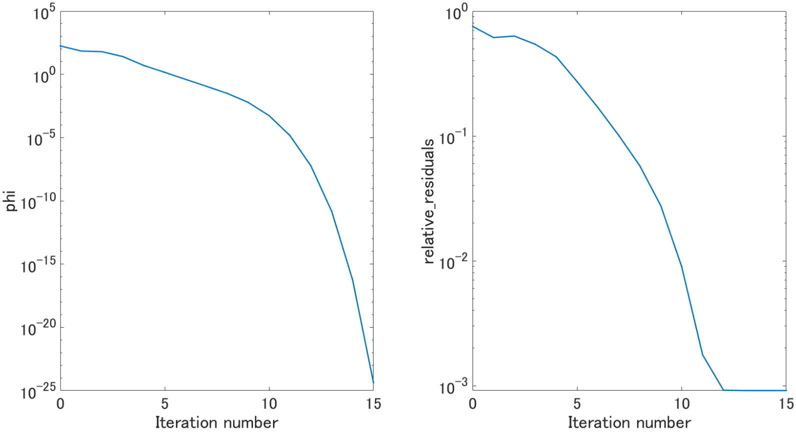


Figure: For $m = 10, n = 8, r = 3$ and $\sigma = 0.001$. It is as good as the results in [SN20].

Future Work — Practical Algorithm.

If no equality $h(x) = 0$, we only need to solve

$$\mathcal{T}\Delta x = c, \tag{32}$$

where $\mathcal{T}$ is a self-adjoint operator on $T_x\mathbb{M}$ (say, $\dim T_x\mathbb{M} =: d$).

1. The basic approach to obtain $d \times d$ symmetric matrix $\mathcal{T}_{mat}$ (Tools in Manopt).
2. The features of the manifold itself should be utilized. [AS17]

$$\begin{array}{ccccc}
T_xSt(n,k) & \xrightarrow{\mathcal{S}_1} & Skew(k) \times Mat(n-k,k) & \xrightarrow{\mathcal{S}_2} & \mathbb{R}^N \\
\uparrow \mathcal{T} & & \uparrow \mathcal{T}' & & \uparrow \mathcal{T}'' \\
T_xSt(n,k) & \xleftarrow{\mathcal{S}_1^{-1}} & Skew(k) \times Mat(n-k,k) & \xleftarrow{\mathcal{S}_2^{-1}} & \mathbb{R}^N
\end{array}$$

3. Krylov Subspace Methods (Iterative Solver) for symmetric system $\mathcal{T}''\Delta x'' = c''$.

Reference I

[AS17] Kensuke Aihara and Hiroyuki Sato.
A matrix-free implementation of riemannian newton's method on the stiefel manifold.
Optimization Letters, 11(8):1729–1741, 2017.

[BH19] Ronny Bergmann and Roland Herzog.
Intrinsic formulation of kkt conditions and constraint qualifications on smooth manifolds.
SIAM Journal on Optimization, 29(4):2423–2444, 2019.

[DCFF92] Manfredo Perdigao Do Carmo and J Flaherty Francis.
Riemannian geometry, volume 6.
Springer, 1992.

[EBTTZ96] AS El-Bakry, Richard A Tapia, T Tsuchiya, and Yin Zhang.
On the formulation and theory of the newton interior-point method for nonlinear programming.
Journal of Optimization theory and Applications, 89(3):507–541, 1996.

Reference II

[FFY17] Teles A Fernandes, Orizon P Ferreira, and Jinyun Yuan.
On the superlinear convergence of newton's method on riemannian manifolds.
Journal of Optimization Theory and Applications, 173(3):828–843, 2017.

[FS12] Orizon P Ferreira and Roberto CM Silva.
Local convergence of newton's method under a majorant condition in riemannian manifolds.
IMA Journal of Numerical Analysis, 32(4):1696–1713, 2012.

[JMWC22] Bo Jiang, Xiang Meng, Zaiwen Wen, and Xiaojun Chen.
An exact penalty approach for optimization with nonnegative orthogonality constraints.
Mathematical Programming, pages 1–43, 2022.

[LB19] Changshuo Liu and Nicolas Boumal.
Simple algorithms for optimization on riemannian manifolds with constraints.
arXiv preprint arXiv:1901.10000, 2019.

Reference III

[OOT20] Mitsuaki Obara, Takayuki Okuno, and Akiko Takeda.
Sequential quadratic optimization for nonlinear optimization problems on riemannian manifolds.
arXiv preprint arXiv:2009.07153, 2020.

[SN20] Guang-Jing Song and Michael K Ng.
Nonnegative low rank matrix approximation for nonnegative matrices.
Applied Mathematics Letters, 105:106300, 2020.

[SO20] Anton Schiela and Julian Ortiz.
An sqp method for equality constrained optimization on manifolds.
arXiv preprint arXiv:2005.06844, 2020.

[YS22] Yuya Yamakawa and Hiroyuki Sato.
Sequential optimality conditions for nonlinear optimization on riemannian manifolds and a globally convergent augmented lagrangian method.
Computational Optimization and Applications, pages 1–25, 2022.

Reference IV

[YY96] Hiroshi Yamashita and Hiroshi Yabe.
Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization.
Mathematical Programming, 75(3):377–397, 1996.

[YZS14] Wei Hong Yang, Lei-Hong Zhang, and Ruyi Song.
Optimality conditions for the nonlinear programming problems on riemannian manifolds.
Pacific Journal of Optimization, 10(2):415–434, 2014.

The End

Questions? Comments?

Appendix.

Riemannian Newton method

Riemannian Newton method: Consider

$$F(x) = 0. \tag{33}$$

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:

$$\left.\begin{array}{l}\left.\begin{array}{l}\text{(N1)There exists } x^* : F(x^*) = 0. \\ \text{(N2)}\nabla F(x^*) \text{ is nonsingular operator.}\end{array}\right\} \Rightarrow \text{superlinear[FFY17]} \\ \text{(N3)}\nabla F \text{ is locally Lipschitz cont. at } x^*.\end{array}\right\} \Rightarrow \text{quadratic[FS12].}$$

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,

$$Z\Delta s + S\Delta z = -ZSe + \mu e, \tag{34}$$

  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

$$\alpha_k := \min\left\{1, \gamma_k \min_i \left\{-\frac{(s_k)_i}{(\Delta s_k)_i} \mid (\Delta s_k)_i < 0\right\}, \gamma_k \min_i \left\{-\frac{(z_k)_i}{(\Delta z_k)_i} \mid (\Delta z_k)_i < 0\right\}\right\}, \tag{35}$$

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

The relation of α_k and γ_k: [YY96]

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
  by El-Bakry, Tapia, Tsuchiya, and Zhang[EBTTZ96], Yamashita and Yabe [YY96].
- Global algorithms
  by El-Bakry, Tapia, Tsuchiya, and Zhang[EBTTZ96]

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.,

$$w(\alpha) := \bar{R}_w(\alpha \Delta w), \tag{36}$$

for some step length $\alpha > 0$.

By introducing

$$w(\alpha) = (x(\alpha), y(\alpha), z(\alpha), s(\alpha)), \tag{37}$$

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:

$$f^I(\alpha) := \min(Z(\alpha)S(\alpha)e) - \gamma\tau_1 \frac{z(\alpha)^T s(\alpha)}{m}, \tag{38}$$

$$f^{II}(\alpha) := z(\alpha)^T s(\alpha) - \gamma\tau_2 \|F(w(\alpha))\|. \tag{39}$$

For $i = I, II$, let

$$\alpha^i := \max_{\alpha \in (0,1]} \left\{ \alpha : f^i(t) \geq 0, \text{ for all } t \in (0, \alpha] \right\}. \tag{40}$$

- Widely used?

Global RIP Algorithm

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

$$\mu_k = z_k^T s_k / m \tag{41}$$

and by

$$\nabla F(w)\Delta w = -F(w) + \sigma_k \mu_k \hat{e}. \tag{42}$$

2. Step length selection.
   1. Centrality conditions: Choose $1/2 < \gamma_k < \gamma_{k-1} < 1$; compute $\alpha^i, i = I, II$, from (40); and let

$$\bar{\alpha}_k = \min(\alpha^I, \alpha^{II}). \tag{43}$$

   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

$$\varphi(\bar{R}_{w_k}(\alpha_k \Delta w_k)) - \varphi(w_k) \leq \alpha_k \beta \langle \operatorname{grad} \varphi_k, \Delta w_k \rangle. \tag{44}$$

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

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

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

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,

$$\|A_x\| = \|\hat{A}_x\|_2 \leq \|\hat{A}_x\|_F. \tag{45}$$

Lemma (L. 2022)

$$x \mapsto \|\widehat{\mathrm{Hess} f(x)}\| \tag{46}$$

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

$$x \mapsto \|H_x\| \text{ and } x \mapsto \|G_x\| \tag{47}$$

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

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

Lemma (Gauss [DCFF92, 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* $w \in T_p\mathcal{M} \approx T_v\left(T_p\mathcal{M}\right)$*. Then*

$$\left\langle \mathcal{D}\exp_p(v)[v], \mathcal{D}\exp_p(v)[w]\right\rangle = \langle v, w\rangle. \tag{48}$$

Manopt, a matlab toolbox for optimization on manifolds

- Manifolds in Manopt are represented as structures and are obtained by calling a factory.

```
M = euclideanfactory(m,n);
M = symmetricfactory(n);
M = skewsymmetricfactory(n);

M = spherefactory(n);
M = obliquefactory(n,m);
M = stiefelfactory(n,k);
M = fixedrankembeddedfactory(m,n);
⋮
```

- M = Ma x Mb x Mc...

```
M = productmanifold(Ma,Mb,Mc...);
```

Manopt — Manifold structure

- A manifold structure has a number of fields, most of which contain function handles.

Table: Part I — Basic

Field usage	Functionality
`M.name()`	Returns the name of M.
`M.rand()`	Computes a random point on M.
`M.dist(x,y)`	Computes the Riemannian distance.
`M.proj(x,u)`	Computes $\mathrm{Proj}_x\, u$.
`M.exp(x,u,t)`	Computes exponential map, $\mathrm{Exp}_x(tu)$.
`M.retr(x,u,t)`	Computes retraction, $\mathrm{Retr}_x(tu)$.
`M.egrad2rgrad(x,egrad)`	Euclidean to Riemannian gradient.
`M.ehess2rhess(x,egrad,ehess,u)`	Euclidean to Riemannian Hessian.

Manopt — Manifold structure

- A manifold structure has a number of fields, most of which contain function handles.

Table: Part II — Tangent space

Field usage	Functionality
`M.dim()`	Returns the dimension of M.
`M.zerovec(x)`	Returns the zero tangent vector at x.
`M.randvec(x)`	Computes a random tangent vector at x.
`M.lincomb(x,a1,u1,a2,u2)`	Computes the linear combination $a_1u_1 + a_2u_2$, where a_1, a_2 scalars and u_1, u_2 tangent vectors at x.
`M.inner(x,u,v)`	Computes the Riemannian metric $\langle u, v\rangle_x$.
`M.norm(x,u)`	Computes the Riemannian norm $\|u\|_x = \sqrt{\langle u, u\rangle_x}$.

Manopt — Other tools

- A number of generically useful tools in Manopt.

Table: Linear operator

Function usage	Functionality
`Bx = tangentorthobasis(M, x)`	Returns an orthonormal basis of tangent space at x.
`matT = operator2matrix(M,x,x,T,Bx,Bx)`	Forms a matrix representing a linear operator between two tangent spaces.
`c_vec = tangent2vec(M,x,Bx,c)`	Expands tangent vector c by an orthonormal basis Bx.
`vec = lincomb(M,x,vecs,coeffs)`	Computes a linear combination of vec = coeffs(1)*vecs1 + ... + coeffs(n)*vecsn

```
RIPM.m
% function [x, cost, info, options] = RIPM(problem)
% function [x, cost, info, options] = RIPM(problem, x0)
% function [x, cost, info, options] = RIPM(problem, x0, options)
% function [x, cost, info, options] = RIPM(problem, [], options)
```

This function calls:

- `RIPM_getNTdirection.m % Solve NT equation.`
- `RIPM_linesearch.m`
- `RIPM_stoppingcriterion.m % Allow the user defined stop criterion.`
- `RIPM_applyStatsfun.m % Allow the user defined stats function.`

1. Euclidean IPM is a special case when $\mathbb{M}$ is Euclidean space.
2. If the equality constraints are considered as $\mathbb{M}$, $\dim \mathcal{T}$ can become smaller.

Manifold $\mathbb{M}$	$h(X)$	codomain of h	$\dim \mathcal{T}$
$\mathbb{R}^{n\times n}$	$X^T - X = O$	$\mathrm{Skew}(n)$	$n^2 + n(n-1)/2$
$\mathrm{Sym}(n)$	-	-	$n(n+1)/2$
$\mathbb{R}^n$	$\|\|x\|\|^2 - 1 = 0$	$\mathbb{R}$	$n+1$
sphere(n)	-	-	$n-1$
$\mathbb{R}^{n\times k}$	$X^T X - I_k = O$	$\mathrm{Sym}(k)$	$nk + k(k+1)/2$
stiefel(n,k)	-	-	$nk - k(k+1)/2$
$\mathbb{R}^{m\times n}$	$\mathrm{rank}(X) = r$ is not continuous	-	-
fixedrank(m,n,r)	-	-	$r(m+n-r)$

3. Not all manifolds are equivalent to the smooth equality constraints.