Superlinear and Quadratic Convergence of Riemannian Interior Point Methods

*Zhijian Lai
Akiko Yoshise

University of Tsukuba

March 18, 2022

Riemannian Interior Point Methods

Riemannian manifold = manifold + Riemannian metric.

- A manifold $\mathcal{M}$ is a set that can be locally linearized.

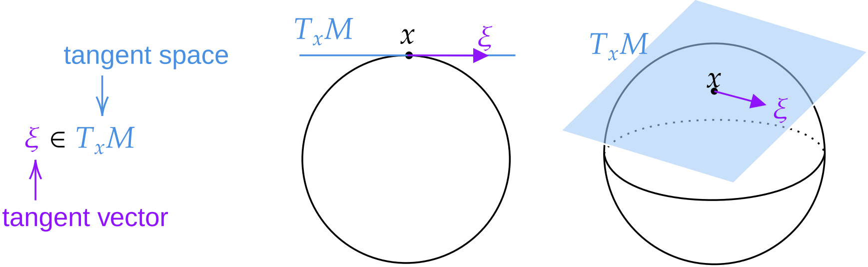


Figure: Manifold of unit sphere, $\mathcal{M} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$.

- Riemannian metric is the family of inner products on each tangent space.

Riemannian (Unconstrained) Optimization

Problem: Given $f(x) : \mathcal{M} \to \mathbb{R}$, solve

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

where $\mathcal{M}$ is a Riemannian manifold.

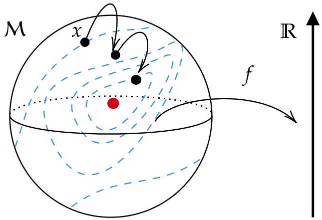


Figure: Iteration on manifold

Euclidean constrained problem $\Rightarrow$ Unconstrained problem on manifold $\mathcal{M}$.

1. Stiefel manifold, $\mathcal{M} = \{X \in \mathbb{R}^{n \times p} : X^\top X = I\}$.
2. Fixed rank manifold, $\mathcal{M} = \{X \in \mathbb{R}^{m \times n} : \mathrm{rank}(X) = r\}$.
3. And more.

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

Riemannian Constrained Optimization

We consider

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

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

Applications:

1. Nonnegative PCA:

$$\min_{X \in \mathbb{R}^{n \times p}} -\operatorname{tr}\left(X^\top A A^\top X\right) \text{ s.t. } X^\top X = I, X \geqslant 0. \tag{1}$$

2. Subproblem of K-indicators model for Data Clustering;
3. Minimum Balanced Cut for Graph Bisection.
4. And more.

Riemannian Constrained Optimization

We consider

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

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

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

Only 3 Riemannian algorithms exist! (2019~)
Augmented Lagrangian Method [LB19, YS22]; Exact Penalty Method [LB19]; Sequential Quadratic Method [SO20, OOT20].
Why not Interior Point Method?

Moving on a manifold — retractions

A retraction R yields a map $R_x : T_x\mathcal{M} \to \mathcal{M}$ for any x.

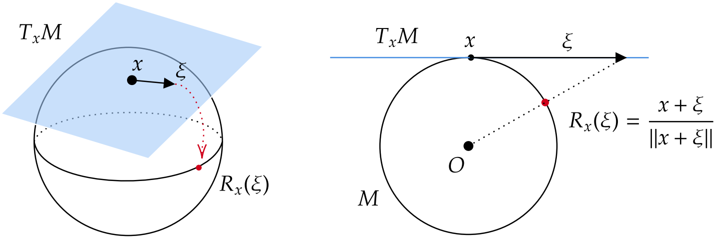


Figure: A retraction on $\mathcal{M} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$.

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

Riemannian gradient — a special vector field

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

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

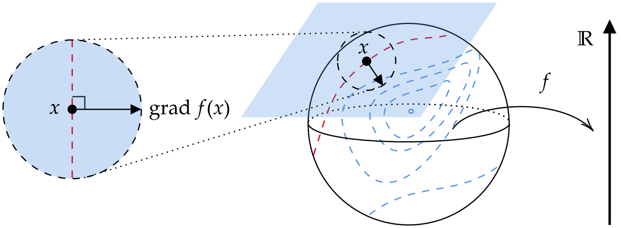


Figure: $\operatorname{grad} f(x)$ is perpendicular to the contour line of f on $\mathcal{M} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$.

$\operatorname{grad} f$ is a special vector field on $\mathcal{M}$.

Riemannian Newton method

Covariant derivative:

Riemannian connection

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

general vector field

When $F = \operatorname{grad} f$, $\operatorname{Hess} f(x) := \nabla \operatorname{grad} f(x)$ is called Riemannian Hessian at x.

Riemannian Newton method

Covariant derivative:

Riemannian connection

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

general vector field

When $F = \operatorname{grad} f$, $\operatorname{Hess} f(x) := \nabla \operatorname{grad} f(x)$ is called Riemannian Hessian at x.

Riemannian Newton method: Consider

$$F(x) = 0. \tag{2}$$

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:

(N1)There exists $x^* : F(x^*) = 0$.
(N2)$\nabla F(x^*)$ is nonsingular operator.
} $\Rightarrow$ superlinear[FFY17]
(N3)∇F is locally Lipschitz cont. at x^*.
} $\Rightarrow$ quadratic[FS12].

Our proposal: Riemannian Interior Point Method

We try to extend interior point methods to

$$\min_{x \in \mathcal{M}} \quad f(x) \qquad \text{s.t.} \quad h(x) = 0, \quad g(x) \geqslant 0, \tag{RCO}$$

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

Lagrangian function is

$$\mathcal{L}(x, y, z) = f(x) - y^\top h(x) - z^\top g(x). \tag{3}$$

$\mathcal{L}(\cdot, y, z)$ is a real function on $\mathcal{M}$, 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 [LB19, Definition 2.3] for problem (RCO) are

$$\begin{cases} \operatorname{grad}_x \mathcal{L}(x, y, z) = 0, \\ h(x) = 0, \\ g(x) \geqslant 0, \\ Zg(x) = 0, \\ z \geqslant 0. \end{cases} \tag{4}$$

KKT Vector Field

Riemannian KKT conditions [LB19, Definition 2.3] for problem (RCO) are

$$\begin{cases} \operatorname{grad}_x \mathcal{L}(x, y, z) = 0, \\ h(x) = 0, \\ g(x) \geqslant 0, \\ Zg(x) = 0, \\ z \geqslant 0. \end{cases} \tag{4}$$

With a slack variable $s = g(x)$, the above can be written

$$F(w) := \begin{pmatrix} \operatorname{grad}_x \mathcal{L}(x, y, z) \\ h(x) \\ g(x) - s \\ ZSe \end{pmatrix} = 0, \text{ and } (s, z) \geqslant 0, \tag{5}$$

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

Definition 2.1 (L. 2022)

F is a vector field on the product Riemannian manifold $\mathscr{M}$, named KKT vector field.

Formulation of $\nabla F(w)$

Lemma 2.2 (L. 2022)

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

$$\nabla F(w)\Delta w = \begin{pmatrix} \operatorname{Hess}_x \mathcal{L}(w)\Delta x - \sum_{i=1}^{l} \Delta y_i \operatorname{grad} h_i(x) - \sum_{i=1}^{m} \Delta z_i \operatorname{grad} g_i(x) \\ \langle \operatorname{grad} h_i(x), \Delta x \rangle, \text{ for } i = 1, \ldots, l \\ \langle \operatorname{grad} g_i(x), \Delta x \rangle - \Delta s_i, \text{ for } i = 1, \ldots, m \\ Z\Delta s + S\Delta z \end{pmatrix}, \tag{6}$$

where $\Delta w = (\Delta x, \Delta y, \Delta s, \Delta z) \in T_x\mathcal{M} \times \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^m \equiv T_w \mathscr{M}$, and $\operatorname{Hess}_x \mathcal{L}(w)$ denotes the Riemannian Hessian of Lagrangian $\mathcal{L}(\cdot, y, z)$.

Remark

In Euclidean case, it reduces to the matrix multiplication:

$$F'(w)\Delta w = \begin{bmatrix} \nabla_x^2 L(w) & -\nabla h(x) & -\nabla g(x) & 0 \\ \nabla h(x)^T & 0 & 0 & 0 \\ \nabla g(x)^T & 0 & 0 & -I \\ 0 & 0 & Z & S \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \\ \Delta z \end{bmatrix}. \tag{7}$$

Implication of Riemannian assumptions for problem (RCO)

(R1) Existence. There exists w^* satisfying the KKT conditions.

(R2) Smoothness. The functions f, g, h are smooth on $\mathcal{M}$.

(R3) 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 $T_{x^*}\mathcal{M}$.

(R4) Strict Complementarity. $(z^*)_i > 0$ if $g_i(x^*) = 0$ for all $i = 1, \cdots, m$.

(R5) 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 2.3 (L. 2022)

If assumptions (R1)-(R5) hold, then standard Newton assumptions (N1)-(N3) hold for $F(w) = 0$.

Riemannian Interior Point Method (RIPM) (L. 2022)

Step 0. Choose an initial w_0 with $(s_0, z_0) > 0$.
Step 1. Solve the following system for $\Delta w_k = (\Delta x_k, \Delta y_k, \Delta s_k, \Delta z_k)$:

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

where $\hat{e} := \hat{e}(w) := (0_x, 0, 0, e)$.
Step 2. Compute the step sizes α_k such that $(s_{k+1}, z_{k+1}) > 0$.
Step 3. Update:

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

$(x_{k+1}, y_{k+1}, s_{k+1}, z_{k+1}) = (R_{x_k}(\alpha_k \Delta x_k), y_k + \alpha_k \Delta y_k, s_k + \alpha_k \Delta s_k, z_k + \alpha_k \Delta z_k)$.
Step 4. Shrink $\mu_k \to 0$. Return to step 1.

Theorem 2.4 Local Convergence of RIPM (L. 2022)

(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^*.

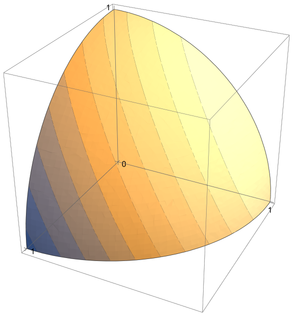
Riemannian Interior Point Methods

An example of Riemannian assumptions (R1)-(R5) (without $s = g(x)$)

$$\min_{x\in\mathcal{M}=\{x\in\mathbb{R}^3:\|x\|_2=1\}} \quad a^\top x \quad \text{s.t.} \quad x \geq 0. \tag{10}$$

$a = [-1, 2, 1]^\top$, $x^* = [1, 0, 0]^\top$. Check KKT condition:

$$\operatorname{grad} f(x^*) = (I - x^* x^{*\top})a = [0, 2, 1]^\top. \tag{11}$$



$x \geq 0$ implies $g_i(x) = e_i^\top x$ for $i = 1, 2, 3$; $\mathcal{A}(x^*) = \{2, 3\}$.

$$\operatorname{grad} g_1(x^*) = (I - x^* x^{*\top})e_1 = [0, 0, 0]^\top.$$
$$\operatorname{grad} g_2(x^*) = (I - x^* x^{*\top})e_2 = [0, 1, 0]^\top.$$
$$\operatorname{grad} g_3(x^*) = (I - x^* x^{*\top})e_3 = [0, 0, 1]^\top.$$

Regularity and strict complementarity hold at x^* with $z^* = [0, 2, 1]^\top$.

$$\operatorname{Hess}_x L(x^*, z^*)[u] = (z^* - a)^\top x^* \cdot u = u, \tag{12}$$

thus, second order sufficiency holds.

Simple Implement of RIPM

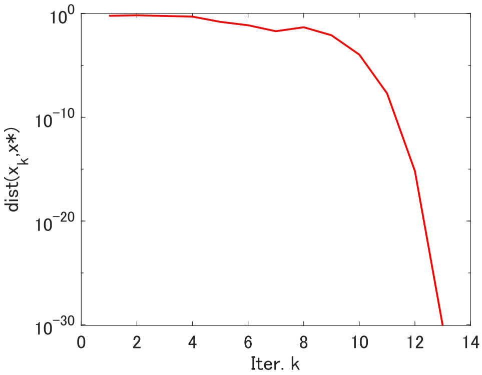
$$\min_{x\in\mathcal{M}=\{x\in\mathbb{R}^3:\|x\|_2=1\}} \quad a^\top x \quad \text{s.t.} \quad x \geq 0. \qquad (13)$$

$a = [-1, 2, 1]^\top, x^* = [1, 0, 0]^\top, n = 3.$

$x_0 = M.rand()$; Random point on manifold.
$s_0 = ones(n, 1) * (0.5); z_0 = ones(n, 1) * (0.5);$

$F_k := \|F(w_k)\|;$
$\mu_k = min(\mu_k/1.5, 0.5 * F_k^2);$
$\gamma_k = 0.5 * (1 + max(0, 1 - F_k));$

A result as shown on the right.



Simple Implement of RIPM

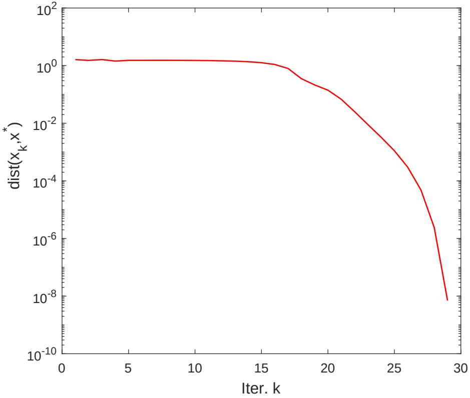
$$\min_{x\in\mathcal{M}=\{x\in\mathbb{R}^n:\|x\|_2=1\}} a^\top x \quad \text{s.t.} \quad x \geq 0. \qquad (14)$$

$a = [-1; abs(rand(n-1,1))]^\top,$
$x^* = [1; zeros(n-1,1)]^\top, n = 1000.$

$x_0 = M.rand()$; Random point on manifold.
$s_0 = ones(n,1) * (0.5); z_0 = ones(n,1) * (0.5);$

$F_k := \|F(w_k)\|;$
$\mu_k = min(\mu_k/1.5, 0.5 * F_k^2);$
$\gamma_k = 0.5 * (1 + max(0, 1 - F_k));$

A result as shown on the right.



Future Work — Global algorithm of RIPM

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

- **Merit function:** A simple one is

$$\varphi(w) = \|F(w)\|^2.$$

- **Step size selection:**
  1. Centrality conditions: $\bar{\alpha}_k = \min\left(\alpha^I, \alpha^{II}\right)$.
  2. Sufficient decreasing: Let $\alpha_k = \theta^t \bar{\alpha}_k$, where t is the smallest nonnegative integer such that α_k satisfies

$$\varphi\left(\bar{R}_{w_k}\left(\alpha_k \Delta w_k\right)\right) - \varphi\left(w_k\right) \leqslant \alpha_k \beta \left\langle \operatorname{grad} \varphi_k, \Delta w_k \right\rangle . \tag{15}$$

Future works:

1. Global Convergence to be proved.
2. Other merit functions; linear search $\longrightarrow$ trust region.

END.

Reference I

[AMS09] P-A Absil, Robert Mahony, and Rodolphe Sepulchre.
Optimization algorithms on matrix manifolds.
Princeton University Press, 2009.

[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.

[Bou20] Nicolas Boumal.
An introduction to optimization on smooth manifolds.
Available online, Aug, 2020.

[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.

[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.

Reference II

[HLWY20] Jiang Hu, Xin Liu, Zai-Wen Wen, and Ya-Xiang Yuan.
A brief introduction to manifold optimization.
Journal of the Operations Research Society of China, 8(2):199–248, 2020.

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

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

[Sat21] Hiroyuki Sato.
Riemannian Optimization and Its Applications.
Springer Nature, 2021.

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

Reference III

[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.

[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.

Perturbed KKT Conditions and Damped Step Size

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

- Introducing the **perturbed** complementary equation,

$$Z\Delta s + S\Delta z = -ZSe + \mu e, \tag{16}$$

  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

$$\alpha_k := \min\left\{1, \gamma_k \min_i \left\{-\frac{(s_k)_i}{(\Delta s_k)_i} \mid (\Delta s_k)_i < 0\right\}, \gamma_k \min_i \left\{-\frac{(z_k)_i}{(\Delta z_k)_i} \mid (\Delta z_k)_i < 0\right\}\right\}, \tag{17}$$

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

The relation of α_k and γ_k: [YY96]

1. If $\gamma_k \longrightarrow 1$, then $\alpha_k \longrightarrow 1$.
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-)

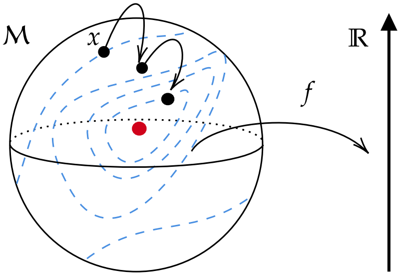
Early phase (1990-1995)

- 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.

Comparison with Constrained Optimization



1. All iterates on the manifold
2. Convergence properties of unconstrained optimization algorithms
3. No need to consider Lagrange multipliers or penalty functions
4. Exploit the structure of the constrained set

from `https://www.math.fsu.edu/~whuang2/pdf/NanjingUniversity_2019-10-23.pdf`

Riemannian submanifold

The letter $\mathcal{E}$ always denotes a linear space.

Embedded submanifold = manifold + subset of $\mathcal{E}$;

- Sphere $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$.

Riemannian submanifold = Embedded submanifold + inherited metric;

- Let
$$\langle u, v\rangle_x := u^\top v \tag{18}$$
for all $u, v \in T_x\mathbb{S}^{n-1} = \{y \in \mathbb{R}^n : x^\top y = 0\}$.
- We have
$$\mathrm{T}_x\mathcal{M} = \{\gamma'(0) \in \mathcal{E} \mid \gamma : I \to \mathcal{M} \text{ is smooth curve around } 0, \gamma(0) = x\} \tag{19}$$

Riemannian gradient for Riemannian submanifold

Proposition

With

$$\mathrm{Proj}_x : \mathcal{E} \to \mathrm{T}_x\mathcal{M} \subseteq \mathcal{E} \tag{20}$$

we denote the orthogonal projector from $\mathcal{E}$ *to* $\mathrm{T}_x\mathcal{M}$*, then*

$$\mathrm{grad}\, f(x) = \mathrm{Proj}_x(\nabla f(x)). \tag{21}$$

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

$$\mathrm{Proj}_x(u) = (I_n - xx^\top)u. \tag{22}$$

Then, Riemannian gradient of f on $\mathbb{S}^{n-1}$ is

$$\mathrm{grad}\, f(x) = 2(I_n - xx^\top)Ax. \tag{23}$$

Distance, metric space

Given a smooth curve segment $c : [a, b] \to \mathcal{M}$, we define the length of c as

$$L(c) := \int_a^b \|c'(t)\|_{c(t)} \, \mathrm{d}t. \tag{24}$$

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

$$\mathrm{dist}(x, y) := \inf_c L(c) \tag{10.2}$$

where the infimum is taken over all curve segments which connect x to y.

Riemannian Optimization Libraries I

Other manifolds:

- Oblique manifold,

$$\left\{X \in \mathbb{R}^{m\times n} : \mathrm{diag}(X^\top X) = I_n\right\}. \tag{25}$$

- Generalized Stiefel manifold,

$$\left\{X \in \mathbb{R}^{n\times p} : X^\top BX = I_p\right\} \text{ for some } B \succ 0. \tag{26}$$

- Manifold of symmetric positive semidefinite, fixed-rank with unit diagonal,

$$\left\{X \in \mathbb{R}^{n\times n} : X = X^\top \succeq 0, \mathrm{rank}(X) = k, \mathrm{diag}(X) = 1\right\}.$$

- And many more.

List of Riemannian methods (2002∼):

- Steepest decent
- Newton
- trust region
- adaptive cubic overestimation
- conjugate gradient
- Quasi-Newton (BFGS)
- ADMM
- proximal gradient
- stochastic algorithms
- and many more.

Riemannian Optimization Libraries II

Monographs:

- Optimization algorithms on matrix manifolds [AMS09]
- An introduction to optimization on smooth manifolds [Bou20]
- Riemannian Optimization and Its Applications [Sat21]

Survey:

- A brief introduction to manifold optimization [HLWY20]
- History of Riemannian Optimization
  `https://www.math.fsu.edu/~whuang2/pdf/NanjingUniversity_2019-10-23.pdf`

Available solvers:

- Manopt (for Matlab, Python, Julia)
- McTorch (Riemannian optimization for deep learning)

Euclidean KKT Conditions

$$\begin{array}{ll} \text{Minimize} & f(x), \quad x \in \mathbb{R}^n \\ \text{subject to} & h(x) = 0, \quad g(x) \geqslant 0. \end{array} \tag{27}$$

The Lagrangian function is

$$L(x, y, z) = f(x) - y^\top h(x) - z^\top g(x). \tag{28}$$

The KKT conditions in slack variable form is

$$F(x, y, s, z) \equiv \begin{bmatrix} \nabla_x L(x, y, z) \\ h(x) \\ g(x) - s \\ ZSe \end{bmatrix} = 0, \quad (s, z) \geq 0. \tag{29}$$

Let $w = (x, y, s, z)$, then our goal just is

$$F(w) = 0, \quad (s, z) \geq 0. \tag{30}$$

Euclidean Interior Point Method (EIP)

To solve $F(w) = 0$, but $(s, z) \geq 0$. Note that $w = (x, y, s, z)$.

Standard assumptions of (27):

(C1) Existence. There exists (x^*, y^*, z^*) satisfying the KKT conditions.

(C2) Smoothness. The functions f, g, h are smooth.

(C3) Regularity. Linear independence constraint qualification at x^*.

(C4) Strict Complementarity. $z_i^* > 0$ if $g_i(x^*) = 0$.

(C5) Second-Order Sufficiency.

Standard assumptions of (27) imply Newton assumptions of F:
If conditions (C1)-(C5) hold, then the standard assumptions (A1)-(A3) hold for $F(w) = 0$. [EBTTZ96]

Euclidean Interior Point Method (EIP)

To solve $F(w) = 0$, but $(s, z) \geq 0$. Note that $w = (x, y, s, z)$.

Algorithm (Euclidean Interior Point Method)

Step 0. Choose an initial w_0 with $(s_0, z_0) > 0$.
Step 1. Solve the following system for Δw_k:

$$\nabla F(w_k)\,\Delta w_k = -F(w_k) + \mu_k \hat{e}, \tag{31}$$

where $\hat{e} = (0, 0, 0, e^\top)^\top$.
Step 2. Compute the step sizes α_k such that $(s_{k+1}, z_{k+1}) > 0$.
Step 3. Update:

$$w_{k+1} = w_k + \alpha_k \Delta w_k. \tag{32}$$

Step 4. Shrink the parameter $\mu_k > 0$. Return to step 1.

Euclidean Interior Point Method (EIP)

Interior Point Method as Perturbed damped Newton iterates:

$$w_{k+1} = w_k - \alpha_k \nabla F(w_k)^{-1}(F(w_k) - \mu_k \hat{e}), \qquad k = 0, 1, \cdots . \tag{33}$$

Theorem (Local Convergence Theory of EIP [EBTTZ96])

1. *If* $\alpha_k \to 1$ *and* $\mu_k = o(\|F(w_k)\|)$*, then local superlinear convergence holds.*
2. *If* $1 - \alpha_k = O(\|F(w_k)\|)$ *and* $\mu_k = O(\|F(w_k)\|^2)$*, then local quadratic convergence holds.*

Examples of Manifolds and Applications

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

$$\text{Smallest eigenvalue of symmetric matrix } A = \min_{x \in \mathbb{S}^{n-1}} x^\top A x. \tag{34}$$

2. Stiefel manifold, $\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n\times p} \mid X^\top X = I_p\}$.

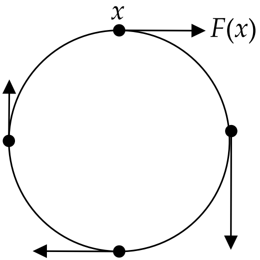
$$\text{Spare PCA:} \min_{X \in \mathrm{St}(p,n)} -\mathrm{tr}\left(X^\top A^\top A X\right) + \rho \|X\|_1. \tag{35}$$

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

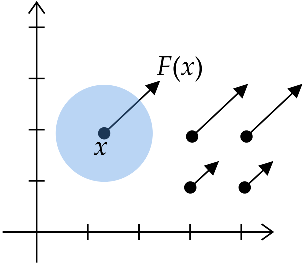
$$\text{Low-rank matrix completion:} \min_{X \in \mathrm{Fr}(m,n,r)} \sum_{(i,j)\in\Omega} (X_{ij} - M_{ij})^2. \tag{36}$$

Riemannian gradient — a special vector field

A vector field F on $\mathcal{M}$ is an assignment of a tangent vector to each point in M.



$F : M \to TM \equiv$ union of all tangent spaces



$F : \mathbb{R}^2 \to \mathbb{R}^2$

Riemannian gradient, grad f — a special vector field on $\mathcal{M}$.

An illustrative example by barrier approach

Consider only inequalities:

$$\min_{x \in \mathcal{M}} \quad f(x) \quad \text{s.t.} \quad g(x) \geq 0. \tag{RCO-I}$$

The logarithmic barrier function of (RCO-I) is

$$B(x;\mu) := f(x) - \mu \sum_{i=1}^{m} \ln g_i(x), \text{ and } \mu > 0. \tag{37}$$

$B(\cdot,\mu)$ is defined on $\{x \in \mathcal{M} : g(x) > 0\}$ — an open subset of $\mathcal{M}$.

Algorithm (Basic Barrier Method (L. 2022))

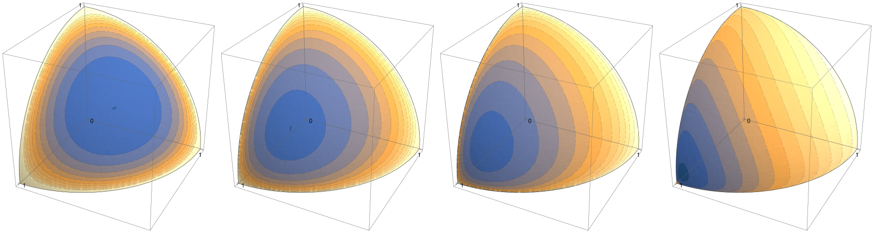
Step 1. Compute an unconstrained minimizer $x(\mu_k)$ of $B(x, \mu_k)$.
Step 2. $x_{k+1} \leftarrow x(\mu_k)$; choose $\mu_{k+1} < \mu_k$; $k \leftarrow k+1$; return to the Step 1.

An illustrative example by barrier approach

Consider a simple problem (SP):

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

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



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

Figure: For (SP), the contour plots of logarithmic barrier function $B(\cdot, \mu)$ with (a) $\mu = 10$; (b) $\mu = 1$; (c) $\mu = 0.5$; (d) $\mu = 0.1$.

Riemannian gradient of $B(x;\mu)$ is

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

An unconstrained minimizer of $B(x,\mu)$ will be denoted by either x_μ or $x(\mu)$, then

1. $\operatorname{grad} B(x_\mu,\mu) = 0$.
2. $x(\mu)$ is a smooth curve on $\mathcal{M}$, and $\lim_{\mu\to 0_+} x(\mu) = x^*$.

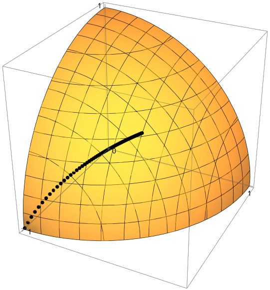


Figure: For (SP), we plot the positive solutions $(x_1(\mu), x_2(\mu), x_3(\mu))$ for different $\mu \to 0$.