Interior-point methods on manifolds

Harold Nieuwboer

University of Amsterdam & Ruhr University Bochum

March 30th, 2023

Dutch Optimization Seminar - PhD talks

Joint work with Hiroshi Hirai & Michael Walter

arXiv:2212.10981 & **arXiv:2303.04771**

https://han28.github.io/

Convex optimization

$D \subseteq \mathbb{R}^n$ *convex*: $\forall p, q \in D$ and $t \in [0,1]$, $(1-t)p + tq \in D$.
$f : D \to \mathbb{R}$ convex if

$$f((1-t)p + tq) \leq (1-t)f(p) + tf(q)$$

Convex optimization:

$$\text{minimize } f(p)$$
$$\text{subject to } p \in D.$$

Algorithms for convex optimization

Euclidean setting:

- Specialized: simplex method for linear programming
- First-order: gradient descent, Frank–Wolfe, mirror descent, ...
- Second-order: Newton's method, trust region methods
- Ellipsoid and cutting-plane methods
- Interior-point methods

First- and second-order methods: generalized to manifolds.
Cutting plane methods could exist [Rus19], but no algorithms known.
This work: Interior-point methods!

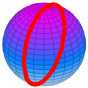
Previous proposals by Udriste [Udr97] and Jiang–Moore–Ji [JMJ07, Ji07], but unsatisfactory.

Convexity on (Riemannian) manifolds

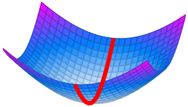
Manifold: space that locally looks like $\mathbb{R}^n$
Riemannian metric: **inner product** on each **tangent space**
Geodesic: $\gamma\colon \mathbb{R} \to M$ which is **locally length minimizing**.
Generalizes straight lines in Euclidean space.
Convexity of a function $f\colon M \to \mathbb{R}$:

$$f(\gamma((1-t)s_0 + ts_1)) \leq (1-t)\, f(\gamma(s_0)) + t\, f(\gamma(s_1))$$

for every geodesic γ, $t \in [0,1]$, $s_0, s_1 \in \mathbb{R}$.



Sphere S^2: $x^2 + y^2 + z^2 = 1$.



Hyperboloid $\mathbb{H}^2$: $x^2 + y^2 - z^2 = -1$.

Hyperboloid: interesting convex functions exist!
In contrast: on sphere, convex implies constant.

Why?

1. *Geometry:* Given points $p_1, \dots, p_m$ on a Riemannian manifold, what is the minimum radius ball that contains all these points? What is their geometric median, i.e., the point that minimizes the sum of distances to each p_i?

2. *Quantum marginals:* Given density matrices $\rho_1, \dots, \rho_k$, each describing the quantum state of one party, does there exist a k-party pure quantum state with marginals equal to the ρ_k?

3. *Tensor networks:* Given a $(2d+1)$-leg tensor, does it ever define a non-zero tensor network state of PEPS type? And how can one efficiently compute its canonical form?

4. *Brascamp–Lieb inequalities:* Given linear $L_k \colon \mathbb{R}^m \to \mathbb{R}^{m_k}$ and numbers $q_k > 0$ for $k \in [n]$, what is the optimal $C > 0$ (if it exists) s.t. for all integrable $f_k \colon \mathbb{R}^{m_k} \to \mathbb{R}_{\geq 0}$,

$$\int_{\mathbb{R}^m} \prod_{k=1}^{n} f_k(L_k x)\, dx \leq C \prod_{k=1}^{n} \|f_k\|_{1/q_k}?$$

Why?

These can all be rephrased as **geodesically convex** optimization problems over spaces of **nonpositive curvature**.
Important examples:

- Euclidean space $\mathbb{R}^n$: zero curvature
- Hyperbolic space $\mathbb{H}^n$: constant curvature -1
- Positive-definite matrices $\mathrm{PD}(n)$ with *affine-invariant* (Fisher–Rao) metric
- Determinant-1 matrices $\mathrm{SPD}(n) \subseteq \mathrm{PD}(n)$

Does **not** include e.g. *n*-sphere or embedded torus.
Problem 1 on minimum enclosing ball & geometric median: *distance to a point* is a convex function (in nonpos. curv.).
Problems 2–4 all related to **scaling problems**.

Why? Scaling problems!

Related to *norm minimization problem*: for nice group G and representation V of G, given $v \in V$, goal is to minimize **norm** over the orbit $G \cdot v \subseteq V$.
Tensor scaling: $G = \mathrm{SL}(n, \mathbb{C}) \times \mathrm{SL}(n, \mathbb{C}) \times \mathrm{SL}(n, \mathbb{C})$ acts on $\mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}^n$ by

$$(g_1, g_2, g_3) \cdot v = (g_1 \otimes g_2 \otimes g_3) v$$

for $g_j \in \mathrm{SL}(n, \mathbb{C})$ and $v \in V$. Goal: find

$$\inf_{g_1, g_2, g_3 \in \mathrm{SL}(n, \mathbb{C})} \|(g_1 \otimes g_2 \otimes g_3) v\|_2^2.$$

Change of variables: $P_j = g_j^* g_j$, then

$$\inf_{P_1, P_2, P_3 \in \mathrm{SPD}(n)} \langle v | P_1 \otimes P_2 \otimes P_3 | v \rangle$$

Domain not convex! But with respect to a natural geometry on $\mathrm{SPD}(n)$, objective is *geodesically convex*. Geodesics on this space: $\sqrt{P} e^{tH} \sqrt{P}$ rather than $P + tH$.

Optimization in nonpositive curvature

The space $\mathrm{PD}(n)$ has nonpositive curvature, obstructs optimization:

- Volume of balls grows *exponentially* with the radius (even in constant dimension).
- Black-box lower bounds for convex optimization: #queries at least linear in distance to approximate minimizer.
- Scaling problems: approximate minimizers **far away**, and the search space is large! Current best algorithms have linear dependence on distance.

Implies that **new structured methods** necessary for efficient scaling algorithms.

Self-concordance and Newton's method

Self-concordance of convex $f \colon D \subseteq \mathbb{R}^n \to \mathbb{R}$:

$$|D^3 f(u,u,u)| \leq 2|D^2 f(u,u)|^{3/2}.$$

Strongly self-concordant = self-concordant and closed convex.
Examples:

- Linear and convex quadratics
- $x \mapsto -\ln(x)$ on $\mathbb{R}_{>0}$
- $-\ln\det(P)$ for $P \in \mathrm{PD}(n)$ (as Euclidean convex set)
- $(y,z) \mapsto -\ln(\ln(z) - y) - \ln(z)$ on $\{(y,z) : e^y < z\}$

Important properties:

- Sum of self-concordant functions is self-concordant
- Rescaling- and translation invariant

Self-concordance and Newton's method

Newton's iteration: update p to $p_+ =$ minimizer of quadratic approximation

$$f(q) \approx f(p) + df_p(q-p) + \frac{1}{2} D^2 f_p(q-p, q-p)$$

(squared) **Newton decrement** $\lambda_f(p)^2$ is twice the gap in function value.

Theorem

If f is self-concordant and $\lambda_f(p) < 1$, then $p_+ \in D$

$$\lambda_f(p_+) \leq \left(\frac{\lambda_f(p)}{1 - \lambda_f(p)} \right)^2$$

Implies **quadratic convergence** of Newton's method.

Self-concordance and Newton's method

M complete Riemannian manifold, $f\colon D \subseteq M \to \mathbb{R}$ smooth convex function on convex $D \subseteq M$.

Definition: Self-concordance on manifolds

Self-concordant if for all $p \in D$ and $u, v, w \in T_pM$,

$$|\nabla^3 f_p(u, v, w)| \leq 2\sqrt{\nabla^2 f_p(u, u)}\sqrt{\nabla^2 f_p(v, v)}\sqrt{\nabla^2 f_p(w, w)}.$$

where ∇ Levi–Civita connection.

Note: **stronger** than self-concordance along every geodesic (only bounds $u = v = w$)!
Key reason is that **non-zero curvature** implies $\nabla^3 f$ is an asymmetric 3-tensor, unlike in the Euclidean case.

Self-concordance and Newton's method

Result: Newton's method quadratic convergence

If f strongly self-concordant and $\lambda_f(p) < 1$, then $p_+ \in D$ and

$$\lambda_f(p_+) \leq \left(\frac{\lambda_f(p)}{1 - \lambda_f(p)} \right)^2 .$$

Other familiar results can also be obtained, e.g., Dikin ellipsoid of radius 1 contained in domain, existence of minimizers, damped Newton method guarantees, etc.

Path-following method

Goal: minimize $f : D \to M$ with D bounded.
Idea: F **self-concordant barrier** for domain D, with F diverging at the boundary.
Instead of minimizing f directly, follow **central path** $z(t)$ of minimizers of $F_t = t\,f + F$ by iterating:

- Increase t slightly.
- Take a Newton step with respect to F_t to move closer to $z(t)$.

Time-increase governed by barrier parameter:

$$\theta = \sup_{p \in D} \lambda_F(p)^2.$$

Path-following method

Result: Path-following method

Starting from approximate minimizer of θ-barrier F and **compatible** f, in

$$\widetilde{O}\Big(1 + \sqrt{\theta}\log(\|df_p\|^*_{F,p}/\varepsilon)\Big)$$

Newton iterations, can find $p_\varepsilon \in D$ such that

$$f(p_\varepsilon) - \inf_{p \in D} f(p) \leq \varepsilon.$$

Compatibility: encompasses linear- and convex quadratic objectives, but more generally is an estimate on third derivatives of f which implies $F_t = tf + F$ is self-concordant for all $t \geq 0$.

Examples of self-concordant functions & barriers

Result: squared distance on $\mathrm{PD}(n)$ self-concordant

For $P_0 \in \mathrm{PD}(n)$, half the **squared distance**

$$\frac{1}{2} d(P, P_0)^2 = \frac{1}{2} \|\log(P_0^{-1/2} P P_0^{-1/2})\|_{\mathrm{HS}}^2$$

is self-concordant.

Implies similar statement for Hadamard symmetric spaces!
Warning: this result is **highly** non-trivial: the third derivative of the squared distance is **not** zero, unlike on Euclidean space!

Examples of self-concordant functions & barriers

Barrier for a ball on $\mathrm{PD}(n)$

$D = \{(P, S) \in \mathrm{PD}(n) \times \mathbb{R} : \frac{1}{2}d(P, P_0)^2 < S\}$. Then

$$F(P, S) = -\log(S - d(P, P_0)^2) + \frac{1}{2}d(P, P_0)^2$$

is strongly self-concordant and $\lambda_F(P, S) \leq 1 + d(P, P_0)^2$.

Yields a self-concordant barrier for ball of radius R with barrier parameter $O(R^2)$.

Applications: geometry problems

Corollary: algorithm for minimum enclosing ball

Given points $P_1, \ldots, P_m \in \mathrm{PD}(n)$, $R_0 = \max_{i,j} d(P_i, P_j)$, $\varepsilon > 0$, can find P_ε such that

$$\max_i d(P_\varepsilon, P_i) \leq \min_P \max_i d(P, P_i) + \varepsilon$$

in $\widetilde{O}\left(mR_0^2 + \sqrt{mR_0^2}\log(1/\varepsilon)\right)$ Newton steps.

Best previous result (only on $\mathbb{H}^n$): multiplicative error δ in $O(1/\delta^2)$ iterations [NH15].

Get similar result for approximate geometric median on $\mathbb{H}^n$: requires a non-trivial strengthening of self-concordance estimate to construct a barrier for the "second order cone".

Applications: norm minimization

Theorem: algorithm for norm minimization

Let $G \subseteq \mathrm{GL}(N)$ complex reductive, $\pi \colon G \to \mathrm{GL}(V)$ algebraic representation, $v \in V$, and set

$$\phi_v(g) = \log\|g \cdot v\|_2^2.$$

Then for $R_0 > 0$ and $\varepsilon > 0$, can find $g_\varepsilon \in G$ such that

$$\phi_v(g_\varepsilon) - \inf_{\|\log(g^* g)\|_{\mathrm{HS}} < R_0} \phi_v(g) \leq \varepsilon$$

within $\widetilde{O}(R_0 N(\pi) \log(1/\varepsilon))$ Newton steps.

$N(\pi) = \|\Pi\|$ weight norm of π, $\Pi = d\pi_I$.

This essentially matches the state-of-the-art for non-commutative optimization- and scaling algorithms.

Summary

- Extend self-concordance to Riemannian manifolds, and analyze Newton's method.
- Implement a path-following method with same guarantees as in Euclidean setting.
- Examples: squared distance on $\mathbb{H}^n$, $\mathrm{PD}(n)$ and general Hadamard symmetric spaces.
- First algorithms for efficiently finding high-precision solutions for minimum enclosing ball & geometric median.
- State-of-the-art complexity guarantees for non-commutative scaling problems.

Outlook

Open questions:

- Better barriers and/or lower bounds?
- Universal/entropic barrier?
- Preliminary stage?
- Primal-dual algorithms?

Thank you!

References

Huibo Ji.
Optimization approaches on smooth manifolds.
PhD thesis, Australian National University, 2007.

Danchi Jiang, John B Moore, and Huibo Ji.
Self-concordant functions for optimization on smooth manifolds.
Journal of Global Optimization, 38(3):437–457, 2007.
doi:10.1007/s10898-006-9095-z.

Frank Nielsen and Gaëtan Hadjeres.
Approximating Covering and Minimum Enclosing Balls in Hyperbolic Geometry.
In *Geometric Science of Information*, Lecture Notes in Computer Science, pages 586–594, Cham, 2015. Springer International Publishing.
doi:10.1007/978-3-319-25040-3_63.

Alexander Rusciano.
A Riemannian Corollary of Helly's Theorem, 2019.
arXiv:1804.10738.

Constantin Udriște.
Optimization Methods on Riemannian Manifolds.
Algebras, Groups and Geometries, 14:339–359, 1997.