Curvature and Complexity: Better lower bounds for geodesically convex optimization

Christopher Criscitiello, Nicolas Boumal (EPFL, Institute of Mathematics)

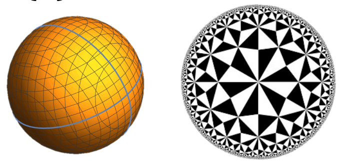
EPFL

OPTIM@EPFL

1 Setting: g-convex optimization

$$\min_{x \in \mathcal{M}} f(x) \quad \textbf{(P)}$$

$\mathcal{M}$ is a Riemannian manifold



f is *geodesically* convex (g-convex)
Def: For every geodesic γ, the 1D function $t \mapsto f(\gamma(t))$ is convex.

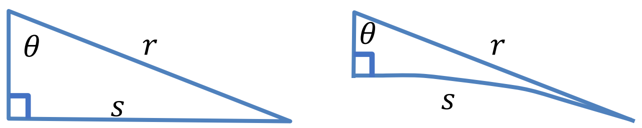
Question: complexity of **(P)** & dependence on curvature of $\mathcal{M}$?

2 Curvature

	$K = 0$	$K < 0$
Constant curvature:	Euclidean space $\mathbb{R}^d$	Hyperbolic space $\mathbb{H}^d$ ($K = -1$)
Volume of ball of radius r:	$\mathrm{Vol} \sim r^d = e^{d \log(r)}$	$\mathrm{Vol} \sim e^{dr}$
Pythagorean thm:	$s^2 = (1 - \epsilon^2) r^2$	$s^2 \approx \left(1 - \frac{\epsilon^2}{\zeta}\right) r^2$

$\zeta = \frac{r\sqrt{|K|}}{\tanh(r\sqrt{|K|})} \sim 1 + r\sqrt{|K|}$

$\epsilon = \cos(\theta)$



Applications usually have K ≤ 0: *Operator scaling, Robust covariance estimation, Matrix normal models, Intrinsic medians (e.g., for phylogenetics)*

3 Complexity: the computational problem

$\mathcal{M} = \mathbb{H}^d$ is a hyperbolic space of curvature $K = -1$ (simplifying assumption)

f attains a global minimizer x^* in a known radius-r geodesic ball $B(x_{\mathrm{ref}}, r)$

f is globally g-convex and regular (consider four function classes):

(P1) Lipschitz & Low-dimensional (d fixed)
(P2) Lipschitz $|f(x) - f(y)| \le M\,\mathrm{dist}(x, y)$
} $f(x) - f(x^*) \le \epsilon \cdot Mr$

(P3) Smooth $|\nabla f(x) - P_{y \to x} \nabla f(y)| \le L\,\mathrm{dist}(x, y)$
(P4) Smooth & strongly g-convex
} $f(x) - f(x^*) \le \epsilon \cdot \frac{1}{2} L r^2$

Black-box model: Algorithm can query an oracle at x to get $f(x), \nabla f(x)$
Worst-case complexity: Least number of oracle queries to find an ϵ-optimal point x, for all functions in the class.

- For (P2) & (P3): Depends on both ϵ and $\zeta \sim 1 + r\sqrt{|K|}$

4 Main results

g-convex setting	Lower bound	Upper bound	Algorithm
(P1) Lipschitz, lo-dim	$\tilde{\Omega}(\zeta + d)$	$O(\zeta d^2)$	Center of gravity (Rusciano'19)
(P2) Lipschitz	$\tilde{\Omega}\left(\zeta + \frac{1}{\zeta^2 \epsilon^2}\right)$	$O\left(\frac{\zeta}{\epsilon^2}\right)$	Subgradient descent (Zhang & Sra'16)
(P3) Smooth	$\tilde{\Omega}\left(\zeta + \frac{1}{\zeta\sqrt{\epsilon}}\right)$	$\tilde{O}\left(\sqrt{\zeta/\epsilon}\right)$	RNAG-C (Kim & Yang'22 / Martinez-Rubio et al.'22)
(P4) Smooth, strongly g-convex	$\tilde{\Omega}(\zeta^* + \sqrt{\kappa})$	$\tilde{O}(\sqrt{\zeta\kappa})$	RNAG-SC (Kim & Yang'22 / Martinez-Rubio et al.'22)

*Hamilton & Moitra'21 / C & Boumal'22

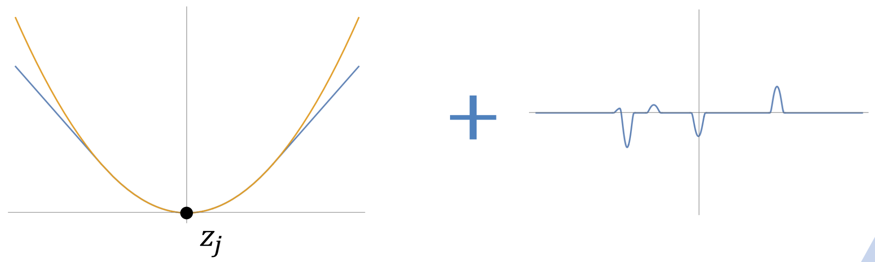
Note: $\kappa \ge \zeta$ and $\frac{1}{\epsilon} \ge \zeta$

Consequences:
- Provides first lower bounds depending on ϵ, κ, d
- Curvature terms in all upper bounds are unavoidable.
- "Full" acceleration is impossible for smooth g-convex opt (P3)

But, lower bounds do not match upper bounds! To address this:
- **Tight $\Omega\left(\frac{\zeta}{\epsilon^2}\right)$ lower bound for subgradient descent** (with Polyak step sizes) for (P2)
- $\tilde{\Omega}(\zeta d)$ **lower bound for cutting-planes game**, a proxy for (P1)

5 Lower bound $\tilde{\Omega}(\zeta)$ for (P1), (P2), (P3)

Slight modification of hard function in C & Boumal'22



Key geometric fact: Volume of ball of radius r scales like e^{dr}
- Pack $B(x_{\mathrm{ref}}, r)$ with *exponentially many* balls $B\left(z_j, \frac{r}{5}\right)$
- Initially, each z_j is a potential value for x^*

6 Lower bound $\Omega\left(\frac{1}{\zeta^2 \epsilon^2}\right)$ for (P2)

In Euclidean space: $f(x) = \max_{i=1,\ldots,d}\{\langle s_i e_i, x\rangle\}, s_i \in \{-1, +1\}$

Fundamental difficulty: No good notion of linear functions on manifolds.

Totally geodesic submanifolds: $S_i = \{x \in \mathcal{M} = \mathbb{H}^d : \langle g_i, \log_{y_i}(x)\rangle = 0\}$
- Generalization of affine subspace

Idea: use max of distance to totally geodesic submanifolds:

$$f(x) = \max_{i=1,\ldots,d}\{\mathrm{dist}(x, S_i^{s_i})\}, s_i \in \{-1, +1\}$$

7 Lower bound $\tilde{\Omega}\left(\frac{1}{\zeta\sqrt{\epsilon}}\right)$ for (P3)

~~Worst function in world: $f(x) = x_{(1)}^2 - 2x_{(1)} + \sum_i (x_{(i)} - x_{(i+1)})^2 + x_{(k)}^2$~~

Smooth the Lipschitz lower bound with Moreau envelope (Guzman & Nemirovski'15):

$$f_\lambda(x) = \inf_{y \in \mathcal{M}}\left\{f(y) + \frac{1}{2\lambda}\mathrm{dist}^2(x, y)\right\}$$

Surprising because no good notion of Fenchel dual on manifolds.

8 Lower bound $\Omega\left(\frac{\zeta}{\epsilon^2}\right)$ for subgrad descent (P2)

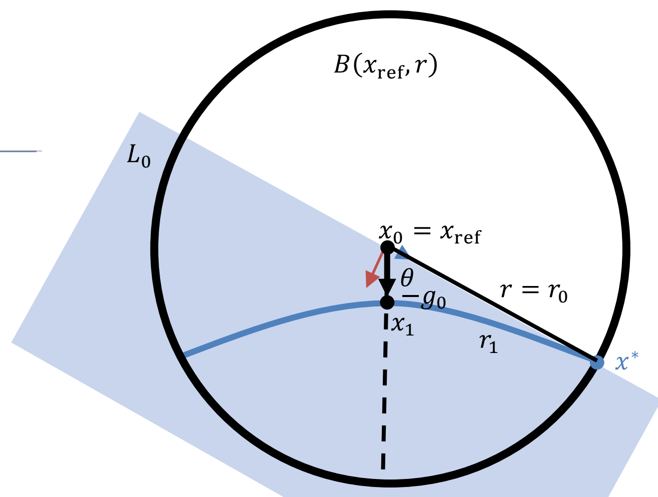
New worst function in the world (*not* an extension of a Euc. proof):

$$f(x) = \mathrm{dist}(x, x^*) + \max_{i=0,\ldots,d-2}\left\{\frac{1}{4\epsilon}\mathrm{dist}(x, L_i)\right\}$$

L_i are cleverly chosen hyperbolic halfspaces.

Key Geometric fact: Pythagorean theorem $r_1^2 \approx \left(1 - \frac{\epsilon^2}{\zeta}\right) r^2$

Construction for (8)



Construction for (6)