Curvature and Complexity:

Better lower bounds for geodesically convex optimization

COLT 2023

Chris Criscitiello
Nicolas Boumal
OPTIM, Chair of Continuous Optimization
Institute of Mathematics, EPFL

EPFL

Setting: g-convex optimization

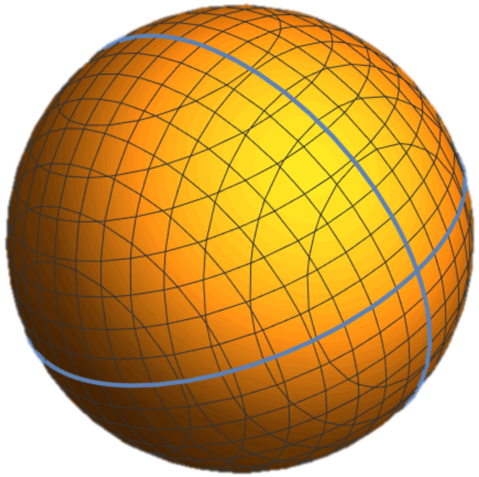
$$\min_{x \in \mathcal{M}} f(x) \qquad \textbf{(P)}$$

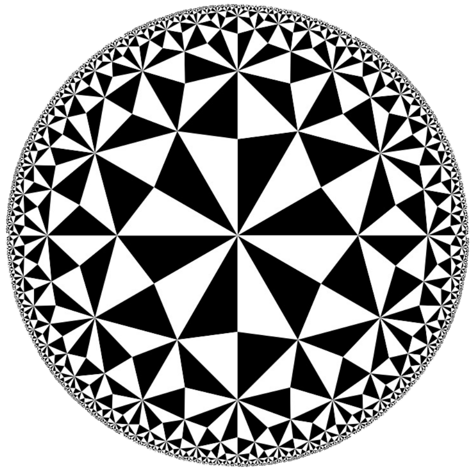
$\mathcal{M}$ is a Riemannian manifold

Setting: g-convex optimization

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

$\mathcal{M}$ is a Riemannian manifold

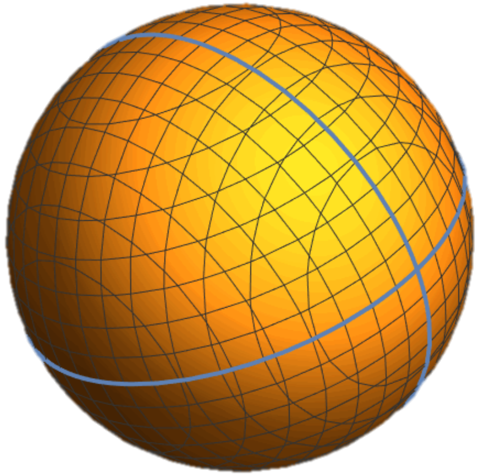


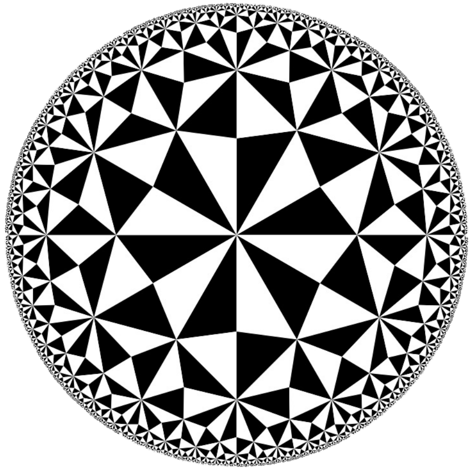


Setting: g-convex optimization

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

$\mathcal{M}$ is a Riemannian manifold





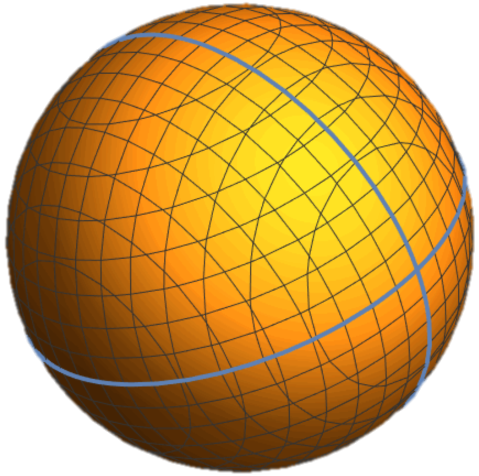
f is *geodesically* convex (g-convex)

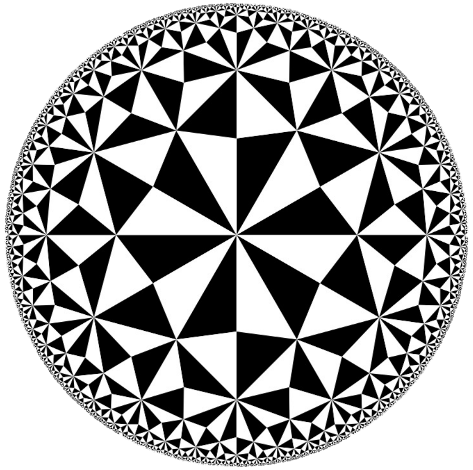
Def: For every geodesic γ, the 1D function $t \mapsto f(\gamma(t))$ is convex.

Setting: g-convex optimization

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

$\mathcal{M}$ is a Riemannian manifold





f is *geodesically* convex (g-convex)

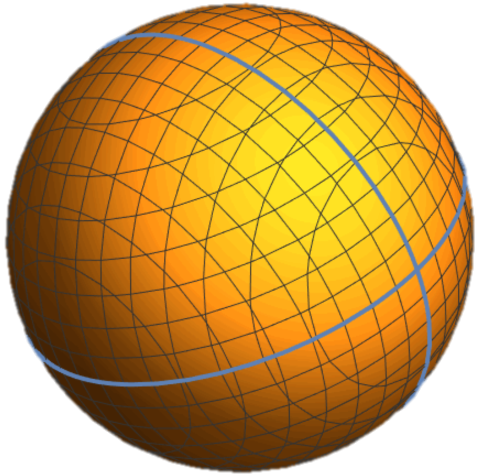
Def: For every geodesic γ, the 1D function $t \mapsto f(\gamma(t))$ is convex.

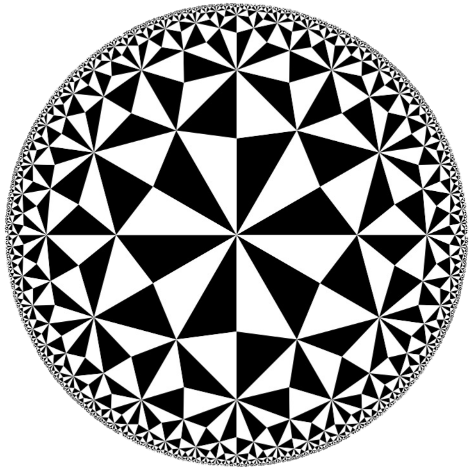
Question: complexity of **(P)** & dependence on curvature of $\mathcal{M}$?

Setting: g-convex optimization

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

$\mathcal{M}$ 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.

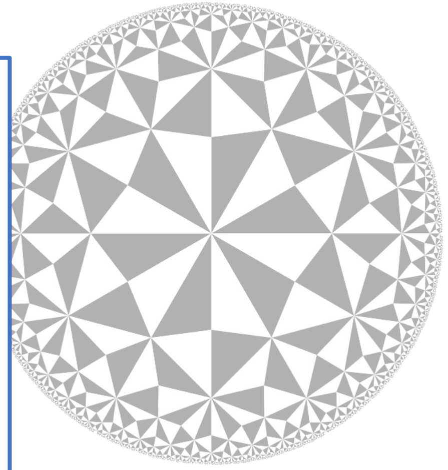
Question: complexity of **(P)** & dependence on curvature of $\mathcal{M}$?

Here, interested in $\mathcal{M}$ with curvature ≤ 0

Setting: g-convex optimization

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

$\mathcal{M}$ is a Riemannian manifold



f is *geodesic*

Def: For every

function

Question: co $\mathcal{M}$?

Most positively curved spaces do not carry nonconstant global g-convex functions.

Applications of g-convexity usually have K $\leq$ 0:

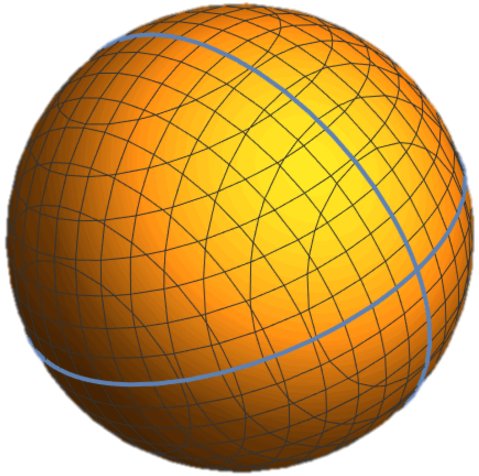
- *Operator scaling*
- *Robust covariance estimation*
- *Matrix normal models*
- *Intrinsic medians (e.g., for phylogenetics)*

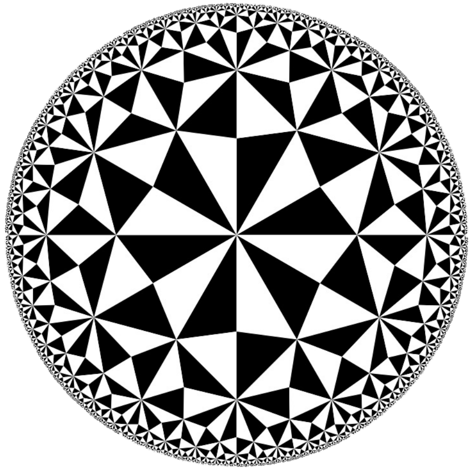
Here, interested in $\mathcal{M}$ with curvature ≤ 0

Setting: g-convex optimization

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

$\mathcal{M}$ 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.

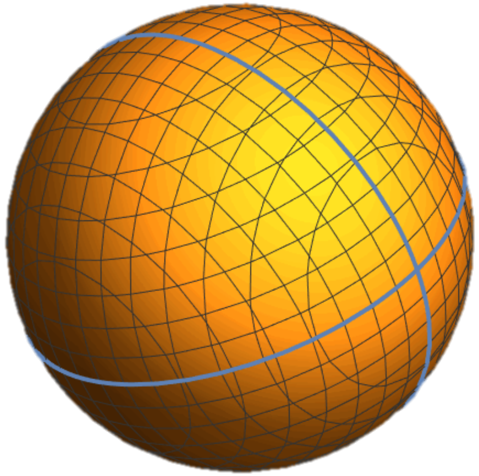
Question: complexity of **(P)** & dependence on curvature of $\mathcal{M}$?

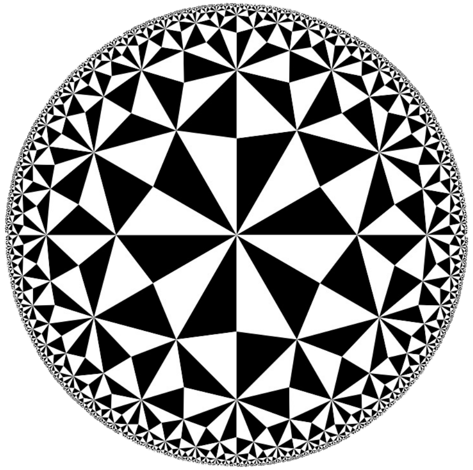
Here, interested in $\mathcal{M}$ with curvature ≤ 0

Setting: g-convex optimization

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

$\mathcal{M}$ 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 $\mathcal{M}$?

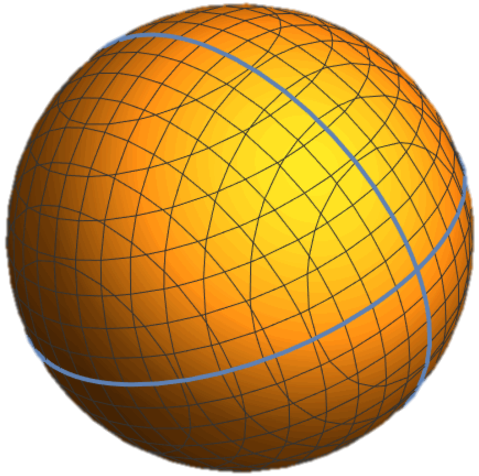
Here, interested in $\mathcal{M}$ with curvature ≤ 0

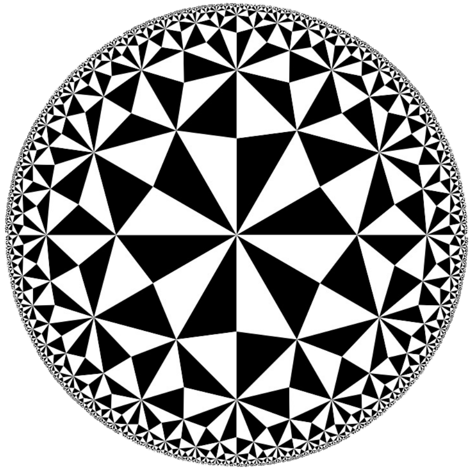
Assume $\mathcal{M}$ has constant curvature (simplifying assumption)

Setting: g-convex optimization

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

$\mathcal{M}$ 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 $\mathcal{M}$?

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

Curvature

$K = 0$

$K < 0$

Constant curvature:

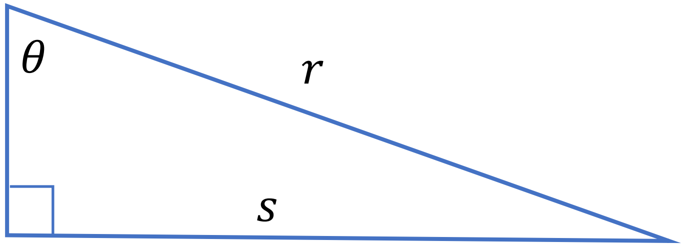
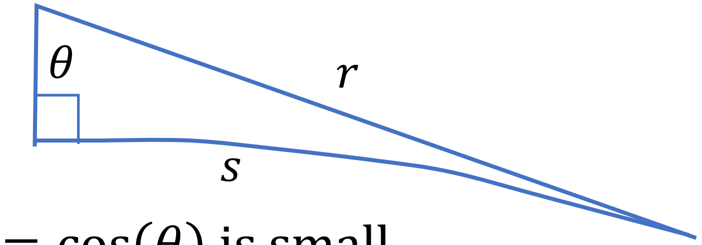
Euclidean space $\mathbb{R}^d$

Hyperbolic space $\mathbb{H}^d$ $(K = -1)$

Curvature

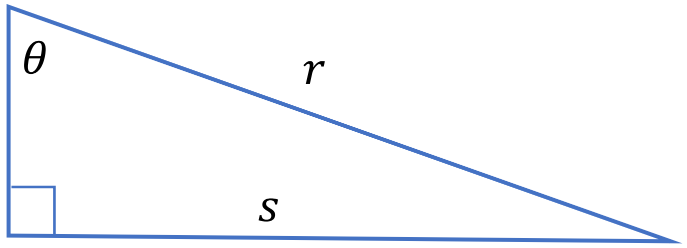
	$K = 0$	$K < 0$
Constant curvature:	Euclidean space $\mathbb{R}^d$	Hyperbolic space $\mathbb{H}^d$ $(K = -1)$
Volume of ball of radius r:	$\text{Vol} \sim r^d = e^{d \log(r)}$	$\text{Vol} \sim e^{dr}$

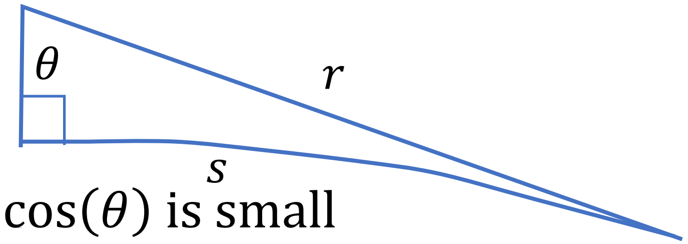
Curvature

	$K = 0$	$K < 0$
Constant curvature:	Euclidean space $\mathbb{R}^d$	Hyperbolic space $\mathbb{H}^d$ $(K = -1)$
Volume of ball of radius r:	$\text{Vol} \sim r^d = e^{d \log(r)}$	$\text{Vol} \sim e^{dr}$
Pythagorean thm:		

$\epsilon = \cos(\theta)$ is small

Curvature





$\epsilon = \cos(\theta)$ is small

$K = 0$

$K < 0$

Constant curvature:

Euclidean space $\mathbb{R}^d$

Hyperbolic space $\mathbb{H}^d$ $(K = -1)$

Volume of ball of radius r:

$\mathrm{Vol} \sim r^d = e^{d\log(r)}$

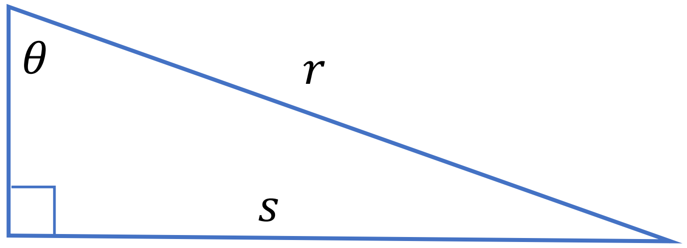
$\mathrm{Vol} \sim e^{dr}$

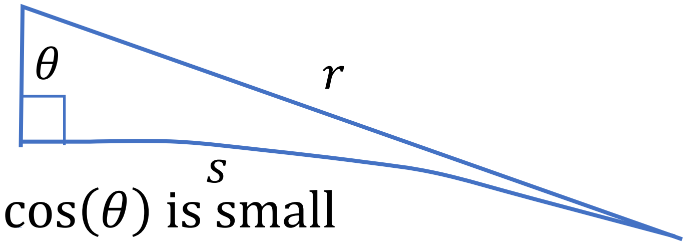
Pythagorean thm:

$s^2 = (1 - \epsilon^2)r^2$

$s^2 \approx \left(1 - \frac{\epsilon^2}{\zeta}\right)r^2$

Curvature





$\epsilon = \cos(\theta)$ is small

	$K = 0$	$K < 0$
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:	$s^2 = (1 - \epsilon^2) r^2$	$s^2 \approx \left(1 - \frac{\epsilon^2}{\zeta}\right) r^2$

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

Complexity: the computational problem

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

Complexity: the computational problem

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

f is globally g-convex and in one of four function classes:

Complexity: the computational problem

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

f is globally g-convex and in one of four function classes:

- (P1) Lipschitz & Low-dimensional (d fixed)
- (P2) Lipschitz $|f(x) - f(y)| \le M \operatorname{dist}(x, y)$
- (P3) Smooth $\left|\nabla f(x) - P_{y \to x} \nabla f(y)\right| \le L \operatorname{dist}(x, y)$
- (P4) Smooth & strongly g-convex

Complexity: the computational problem

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

f is globally g-convex and in one of four function classes:

- (P1) Lipschitz & Low-dimensional (d fixed)
- (P2) Lipschitz $|f(x) - f(y)| \leq M \operatorname{dist}(x, y)$
- (P3) Smooth $\left|\nabla f(x) - P_{y \to x} \nabla f(y)\right| \leq L \operatorname{dist}(x, y)$
- (P4) Smooth & strongly g-convex

Task: find an ϵ-optimal point x, e.g., $f(x) - f(x^*) \leq \epsilon \cdot Mr$

First order black-box model

Complexity: the computational problem

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

f is globally g-convex and in one of four function classes:

- (P1) Lipschitz & Low-dimensional (d fixed)
- (P2) Lipschitz $|f(x) - f(y)| \leq M \operatorname{dist}(x, y)$
- (P3) Smooth $\left|\nabla f(x) - P_{y \to x} \nabla f(y)\right| \leq L \operatorname{dist}(x, y)$
- (P4) Smooth & strongly g-convex

Task: find an ϵ-optimal point x, e.g., $f(x) - f(x^*) \leq \epsilon \cdot Mr$

First order black-box model

Worst-case complexity for (P2) & (P3): can depend on both ϵ and $\zeta \sim 1 + r\sqrt{|K|}$

Main results

$\zeta \sim 1 + r\sqrt{|K|}$

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

Main results

$\zeta \sim 1 + r\sqrt{|K|}$

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

(Hamilton & Moitra'21 / C & Boumal'22)

Main results

$\zeta \sim 1 + r\sqrt{|K|}$

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

(Hamilton & Moitra'21 / C & Boumal'22)

Main results

$\zeta \sim 1 + r\sqrt{|K|}$

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

(Hamilton & Moitra'21 / C & Boumal'22)

Curvature dependence in all upper bounds is unavoidable!

Main results

$\zeta \sim 1 + r\sqrt{|K|}$

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

(Hamilton & Moitra'21 / C & Boumal'22)

Slight modification of hard function in C & Boumal'22

Volume of ball of radius r: Vol $\sim e^{dr}$

Main results

$\zeta \sim 1 + r\sqrt{|K|}$

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

Main results

$\zeta \sim 1 + r\sqrt{|K|}$

g-convex setting	Lower bound	Upper bound	Algorithm
(P1) Lipschitz, lo-dim	$\tilde{\Omega}(\zeta + d)$	$O(\zeta d^2)$	Centerpoint method (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)

Do not match!

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

Additive

Multiplicative

Main results

$\zeta \sim 1 + r\sqrt{|K|}$

g-convex setting	Lower bound	Upper bound	Algorithm
(P1) Lipschitz, lo-dim	$\tilde{\Omega}(\zeta + d)$	$O(\zeta d^2)$	Centerpoint method (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)

Cutting-planes game	$\tilde{\Omega}(\zeta d)$	$O(\zeta d^2)$	Centerpoint method (Rusciano'19)

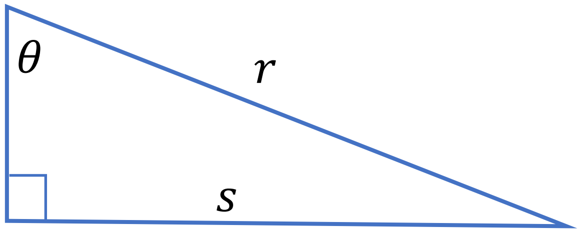
$\Omega\left(\frac{\zeta}{\epsilon^2}\right)$ lower bound for subgradient descent (with Polyak step size)

Main results

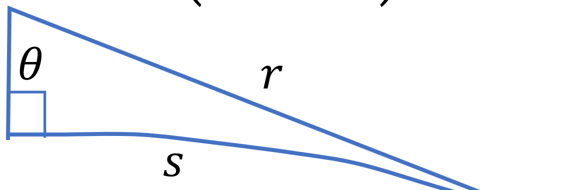
$\zeta \sim 1 + r\sqrt{|K|}$

g-convex setting	Lower bound	Upper bound	Algorithm
(P1) Lipschitz, lo-dim	$\tilde{\Omega}(\zeta + d)$	$O(\zeta d^2)$	Centerpoint method (Rusciano'19)
(P2) Lipschitz	$\tilde{\Omega}(\zeta + \frac{1}{\ldots})$	$O(\frac{\zeta}{\ldots})$	Subgradient descent ...6)
(P3) Sr...			... & Yang'22 / ...o et al.'22)
(P4) Sr... strong...			...m & Yang'22 ...bio et al.'22)
Cutting... game			...method (Rusciano'19)

$$s^2 = (1 - \epsilon^2) r^2$$



$$s^2 \approx \left(1 - \frac{\epsilon^2}{\zeta}\right) r^2$$



$\epsilon = \cos(\theta)$ is small

$\Omega\left(\frac{\zeta}{\epsilon^2}\right)$ lower bound for subgradient descent (with Polyak step size)

Main results

$\zeta \sim 1 + r\sqrt{|K|}$

g-convex setting	Lower bound	Upper bound	Algorithm
(P1) Lipschitz, lo-dim	$\tilde{\Omega}(\zeta + d)$	$O(\zeta d^2)$	Centerpoint method (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)

Cutting-planes game	$\tilde{\Omega}(\zeta d)$	$O(\zeta d^2)$	Centerpoint method (Rusciano'19)

$\Omega\left(\frac{\zeta}{\epsilon^2}\right)$ lower bound for subgradient descent (with Polyak step size)

Lipschitz lower bound (P2)

In Euclidean space: $f(x) = \max_i \{F_i + \langle g_i, x - y_i \rangle\}$

Lipschitz lower bound (P2)

In Euclidean space: $f(x) = \max_i \{F_i + \langle g_i, x - y_i \rangle\}$

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

Lipschitz lower bound (P2)

In Euclidean space: $f(x) = \max_i \{F_i + \langle g_i, x - y_i \rangle\}$

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

Hyperbolic halfspaces: $H_i = \left\{x \in \mathcal{M} : \left\langle g_i, \log_{y_i}(x) \right\rangle \leq 0\right\}$

Lipschitz lower bound (P2)

In Euclidean space: $f(x) = \max_i \{F_i + \langle g_i, x - y_i \rangle\}$

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

Hyperbolic halfspaces: $H_i = \left\{x \in \mathcal{M} : \left\langle g_i, \log_{y_i}(x) \right\rangle \leq 0\right\}$

Idea: use max of distance to halfspaces: $f(x) = \max_i \{F_i + \mathrm{dist}(x, H_i)\}$

Lipschitz lower bound (P2)

In Euclidean space: $f(x) = \max_i \{F_i + \langle g_i, x - y_i \rangle\}$

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

Hyperbolic halfspaces: $H_i = \left\{x \in \mathcal{M}: \left\langle g_i, \log_{y_i}(x) \right\rangle \leq 0\right\}$

Idea: use max of distance to halfspaces: $f(x) = \max_i \{F_i + \mathrm{dist}(x, H_i)\}$

Choosing halfspaces in right way yields lower bound $\widetilde{\Omega}\left(\frac{1}{\zeta^2 \epsilon^2}\right)$.

Smooth lower bound (P3)

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

Smooth lower bound (P3)

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

Smooth lower bound (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 and Nemirovski'15):

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

Smooth lower bound (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 and 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.

Yields lower bound $\tilde{\Omega}\left(\frac{1}{\zeta\sqrt{\epsilon}}\right)$.

Tight lower bound for subgradient descent (P2)

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

$$f(x) = \mathrm{dist}(x, x^*) +$$

Tight lower bound for subgradient descent (P2)

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

$$f(x) = \mathrm{dist}(x, x^*) + \max_i \left\{ \frac{1}{4\epsilon} \mathrm{dist}(x, L_i) \right\}$$

L_i are cleverly chosen hyperbolic halfspaces.

Yields $\Omega\left(\frac{\zeta}{\epsilon^2}\right)$ lower bound for subgradient descent (with Polyak step size).

Tight lower bound for subgradient descent (P2)

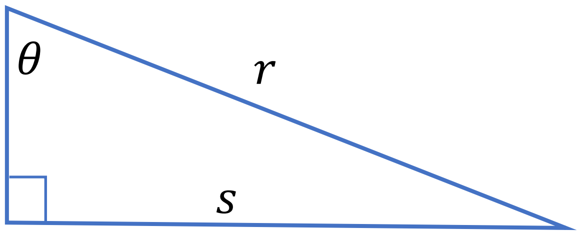
New worst function in the world (*not* an extension of a Euclidean proof):

$$f(x) = \text{dist}(x, x^*) + \max_i \left\{ \frac{1}{4\epsilon} \text{dist}(x, L_i) \right\}$$

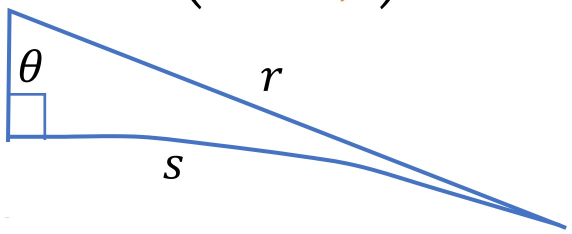
L_i are cleverly chosen hyperbolic halfspaces.

Yields $\Omega\left(\frac{\zeta}{\epsilon^2}\right)$ lower bound for subgradient descent (with Polyak step size).

$$s^2 = (1 - \epsilon^2) r^2$$



$$s^2 \approx \left(1 - \frac{\epsilon^2}{\zeta}\right) r^2$$



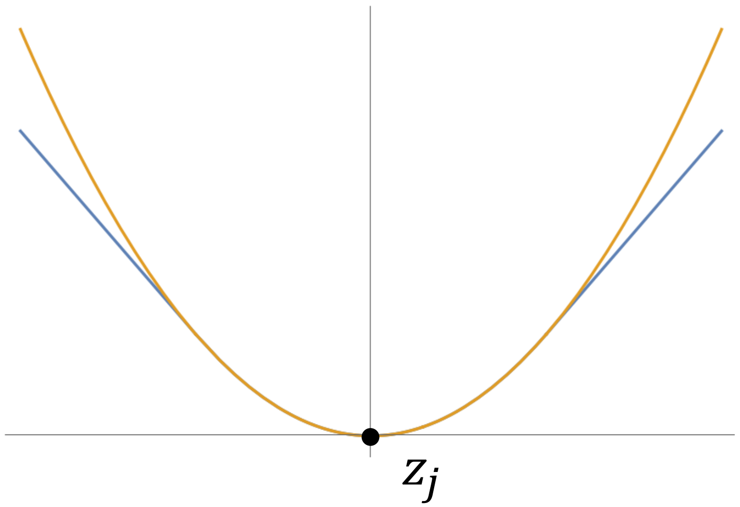
$\epsilon = \cos(\theta)$ is small

Appendix

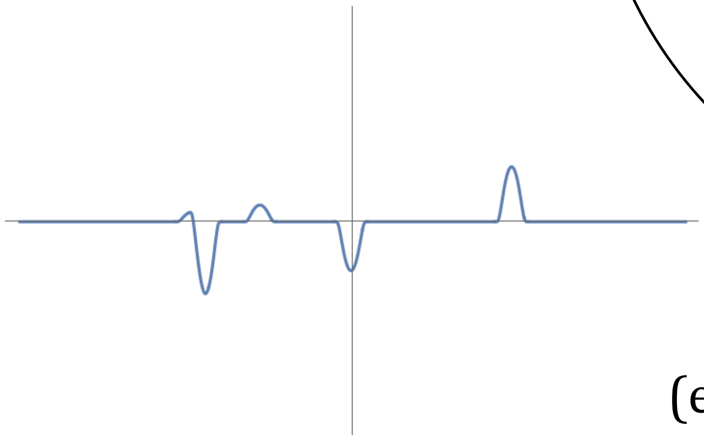
Sec 3

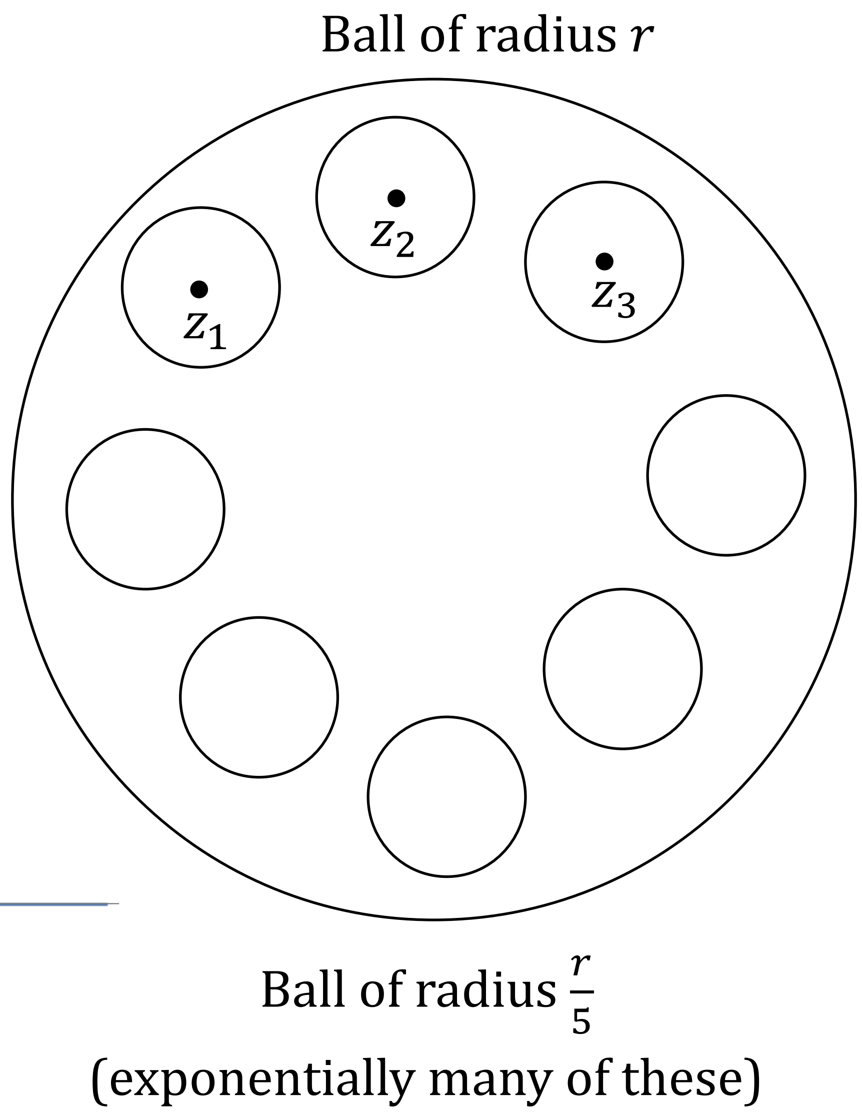
Slight modification of hard function in C & Boumal'21

Volume of ball of radius r: Vol $\sim e^{dr}$



+





Lipschitz lower bound (P2) – Sec 4

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.

- Every convex function equals a sup of affine functions, which are themselves convex.
- Every g-convex function equals a sup of $x \mapsto F_i + \langle g_i, \log_{y_i}(x) \rangle$, but these are not g-convex.

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

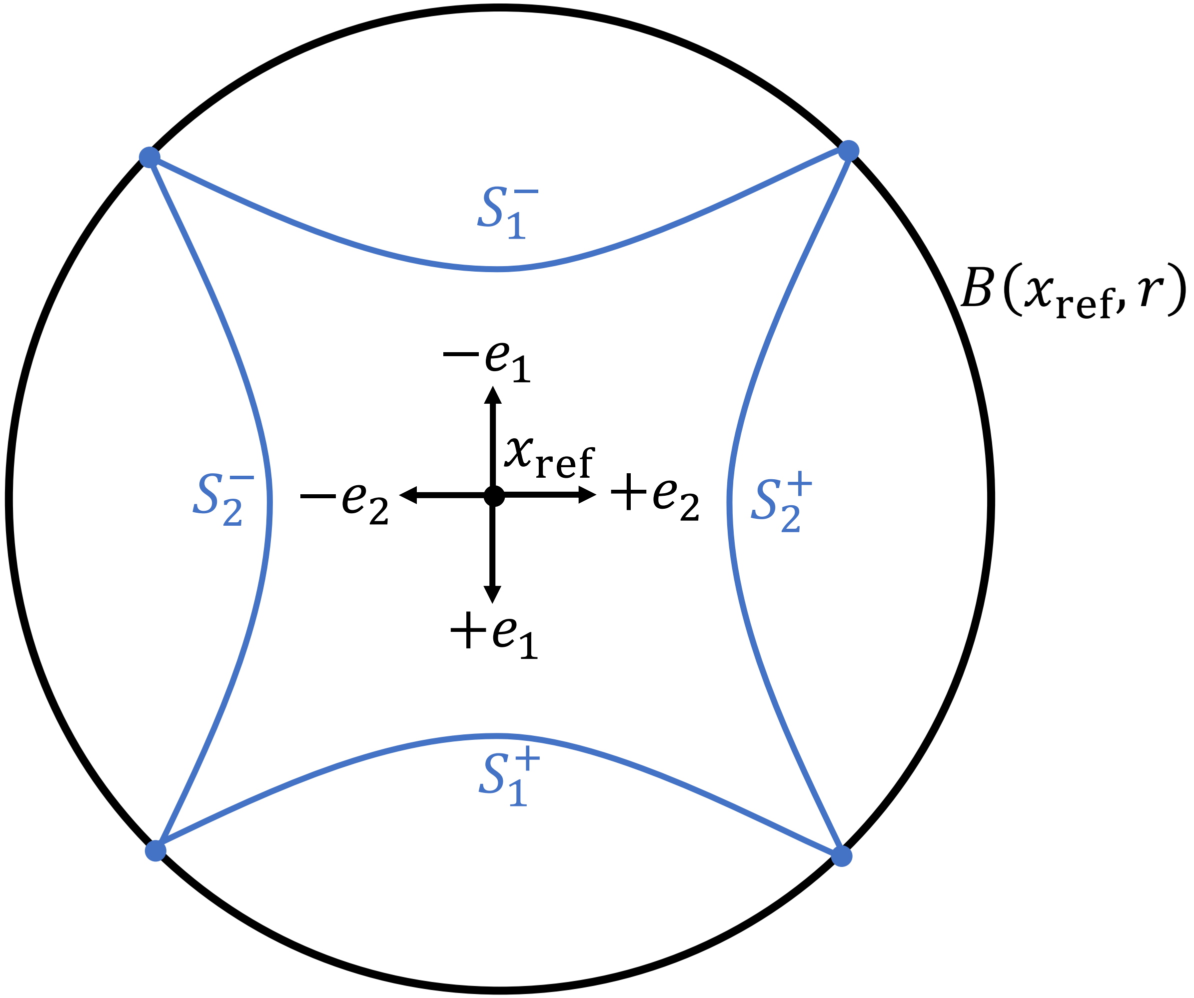
- Generalization of affine subspace

Idea: use max of distance to totally geodesic submanifolds:

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

Choosing halfspaces in right way yields lower bound $\tilde{\Omega}\left(\frac{1}{\zeta^2 \epsilon^2}\right)$.

Lipschitz lower bound (P2) – Sec 4



Smooth lower bound (P3) – Sec 5

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

Smooth Lipschitz lower bound with Moreau envelope (Guzman and 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.

Yields lower bound $\widetilde{\Omega}\left(\frac{1}{\zeta\sqrt{\epsilon}}\right)$.

Tight lower bound for subgradient descent (P2) – Sec 6

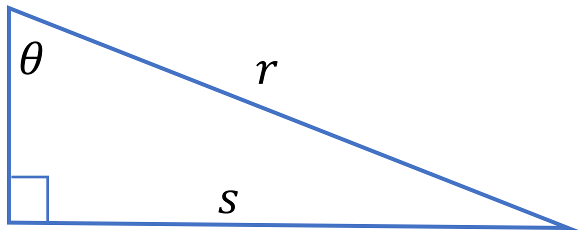
New worst function in the world (*not* an extension of a Euclidean proof):

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

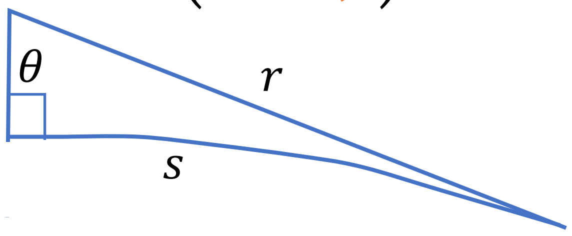
L_i are cleverly chosen hyperbolic halfspaces.

Yields $\Omega\left(\frac{\zeta}{\epsilon^2}\right)$ lower bound for subgradient descent (with Polyak step size).

$$s^2 = (1 - \epsilon^2) r^2$$

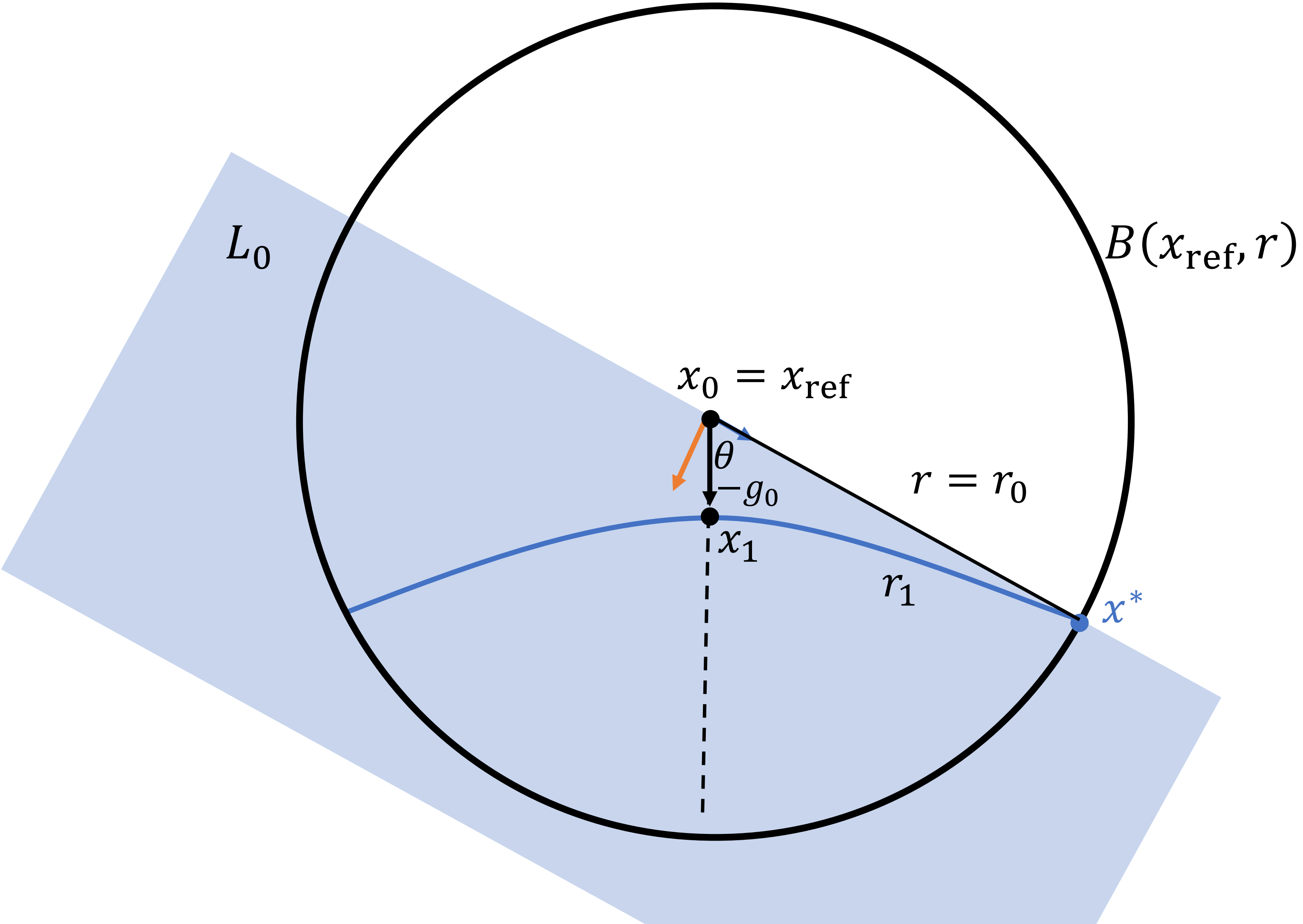


$$s^2 \approx \left(1 - \frac{\epsilon^2}{\zeta}\right) r^2$$



$\epsilon = \cos(\theta)$ is small

Tight lower bound for subgradient descent (P2) – Sec 6



Cutting-planes game (proxy for (P1)) – Sec 7

Cutting planes game:

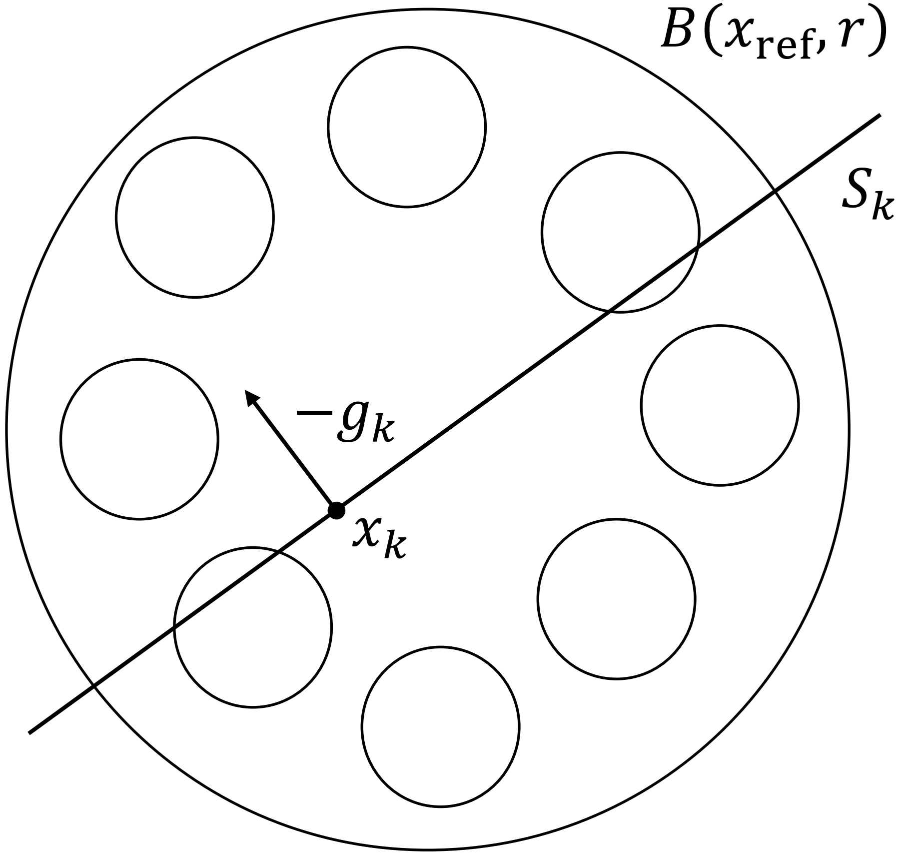
- Initially given $B(x_{\text{ref}}, r)$ containing x^*
- Upon query x_k, oracle returns tangent vector g_k such that $\langle g_k, \log_{x_k}(x^*) \rangle \le 0$
- Task: Find x_k so that $B(x^*, \epsilon r)$ intersects $S_k = \{x \in \mathcal{M} : \langle g_k, \log_{x_k}(x) \rangle = 0\}$

Usual lower bounds proof (tessellation of cube) does not generalize (Nesterov'04).

Width-bounded separators (Kisfaludi-Back'20, Fu'11):

- Given a collect of balls and a point x_k, there exists a totally geodesic submanifold S_k through x_k which intersects a small fraction of these balls

Yields $\tilde{\Omega}(\zeta d)$ lower bound for cutting-plane schemes, e.g., center-of-gravity method.



G-convex interpolation – Sec 8

Interpolation is crucial for building lower bounds (Taylor et al.'16).

A collection of function values and tangent vectors $(F_i, x_i, g_i)_{i=1}^N$ is interpolated by a g-convex function f if $f(x_i) = F_i$ and $g_i \in \partial f(x_i)$ for all i.

A convex function interpolates $(F_i, x_i, g_i)_{i=1}^N$ if and only if

$$F_j \geq F_i + \langle g_i, x_j - x_i \rangle \text{ for all } i, j$$

For g-convex functions the analogous naïve necessary conditions are *not* sufficient for interpolation even for just 3 points:

- There exists $(F_i, x_i, g_i)_{i=1}^3$ such that $F_j \geq F_i + \left\langle g_i, \log_{x_j}(x_j) \right\rangle$ for all i, j, yet this data cannot be interpolated by a g-convex function.