Open Problem:

Polynomial linearly-convergent method for geodesically convex optimization?

COLT 2023

Chris Criscitiello — Institute of Mathematics, EPFL

David Martínez-Rubio — Zuse Institute Berlin and Technische Universität

Nicolas Boumal — Institute of Mathematics, EPFL

Lipschitz g-convex optimization

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\mathrm{ref}}, r) \subset \mathcal{M}$$

Lipschitz g-convex optimization

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\mathrm{ref}}, r) \subset \mathcal{M}$$

$\mathcal{M}$ is a d-dimensional Riemannian manifold with $|\text{curvature}| \leq K$

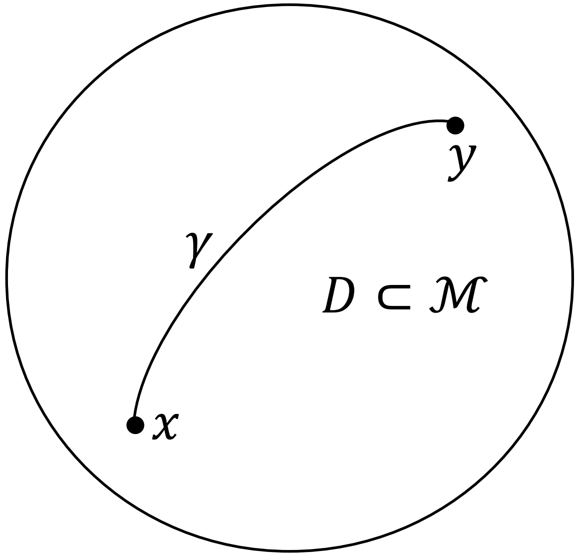
Lipschitz g-convex optimization

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\mathrm{ref}}, r) \subset \mathcal{M}$$

$\mathcal{M}$ is a d-dimensional Riemannian manifold with $|\text{curvature}| \le K$

$f : D \to \mathbb{R}$ is geodesically convex (g-convex):

- The 1D fct $t \mapsto f\big(\gamma(t)\big)$ is convex for any geodesic γ in D



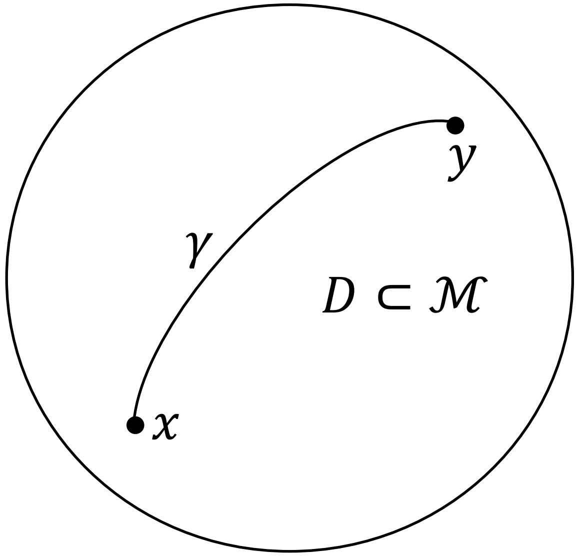
Lipschitz g-convex optimization

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\mathrm{ref}}, r) \subset \mathcal{M}$$

$\mathcal{M}$ is a d-dimensional Riemannian manifold with $|\mathrm{curvature}| \leq K$

$f : D \to \mathbb{R}$ is geodesically convex (g-convex):

- The 1D fct $t \mapsto f\big(\gamma(t)\big)$ is convex for any geodesic γ in D



$f : D \to \mathbb{R}$ is M-Lipschitz:

- $|f(x) - f(y)| \leq M \,\mathrm{dist}(x, y)$, for all $x, y \in D$

Ellipsoid method

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$$

If $\mathcal{M} = \mathbb{R}^d$, amounts to Lipschitz convex optimization.

Ellipsoid method

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$$

If $\mathcal{M} = \mathbb{R}^d$, amounts to Lipschitz convex optimization.

Ellipsoid method: iterative $x_k \to x_{k+1}$,

maintains ellipsoids $E_{k+1} \subset E_k$

Ellipsoid method

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$$

If $\mathcal{M} = \mathbb{R}^d$, amounts to Lipschitz convex optimization.

Ellipsoid method: iterative $x_k \to x_{k+1}$,

maintains ellipsoids $E_{k+1} \subset E_k$

Two key properties:

- (Linear convergence) uses at most $O(\text{poly}(d)\log(\epsilon^{-1}))$ subgradient queries to find a point with target accuracy ϵ, and

Ellipsoid method

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$$

If $\mathcal{M} = \mathbb{R}^d$, amounts to Lipschitz convex optimization.

Ellipsoid method: iterative $x_k \to x_{k+1}$,

maintains ellipsoids $E_{k+1} \subset E_k$

Two key properties:

- (Linear convergence) uses at most $O(\text{poly}(d) \log(\epsilon^{-1}))$ subgradient queries to find a point with target accuracy ϵ, and
- (Polynomial per-query complexity) requires only $O(\text{poly}(d))$ arithmetic operations per query x_k

Ellipsoid method

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$$

If $\mathcal{M} = \mathbb{R}^d$, amounts to Lipschitz convex optimization.

Ellipsoid method: iterative $x_k \to x_{k+1}$,

maintains ellipsoids $E_{k+1} \subset E_k$

Full convexity not needed.
Just need halfspace at each iteration.

Two key properties:

- (Linear convergence) uses at most $O(\text{poly}(d)\log(\epsilon^{-1}))$ subgradient queries to find a point with target accuracy ϵ, and
- (Polynomial per-query complexity) requires only $O(\text{poly}(d))$ arithmetic operations per query x_k

Open Problem

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$$

$f: D \to \mathbb{R}$ is g-convex and Lipschitz

Open Problem

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$$

$f: D \to \mathbb{R}$ is g-convex and Lipschitz

Q: Is there a first-order deterministic algorithm with the following properties?

- (Linear convergence) uses at most $O(\text{poly}(d, \zeta) \log(\epsilon^{-1}))$ subgradient queries to find a point with target accuracy ϵ, and
- (Polynomial per-query complexity) requires only $O(\text{poly}(d, \zeta))$ arithmetic operations per query?

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

Open Problem

Q: Is there a first-order deterministic algorithm with the following properties?

- (Linear convergence) uses at most $O(\text{poly}(d, \zeta) \log(\epsilon^{-1}))$ subgradient queries to find a point with target accuracy ϵ, i.e., $f(x) - f^* \leq \epsilon \cdot Mr$, and
- (Polynomial per-query complexity) requires only $O(\text{poly}(d, \zeta))$ arithmetic operations per query?

Stated before by Allen-Zhu et al.'18 and Rusciano'19.

Open Problem

Q: Is there a first-order deterministic algorithm with the following properties?

- (Linear convergence) uses at most $O(\text{poly}(d, \zeta)\log(\epsilon^{-1}))$ subgradient queries to find a point with target accuracy ϵ, i.e., $f(x) - f^* \leq \epsilon \cdot Mr$, and
- (Polynomial per-query complexity) requires only $O(\text{poly}(d, \zeta))$ arithmetic operations per query?

Stated before by Allen-Zhu et al.'18 and Rusciano'19.

Our contribution:

- Partial solution: the case of constant curvature (hyperbolic spaces, hemispheres)
- Get others interested in it!

Open Problem

Q: Is there a first-order deterministic algorithm with the following properties?

- (Linear convergence) uses at most $O(\text{poly}(d, \zeta)\log(\epsilon^{-1}))$ subgradient queries to find a point with target accuracy ϵ, i.e., $f(x) - f^* \leq \epsilon \cdot Mr$, and
- (Polynomial per-query complexity) requires only $O(\text{poly}(d, \zeta))$ arithmetic operations per query?

Answer should be **YES**, right?

Open Problem

Q: Is there a first-order deterministic algorithm with the following properties?

- (Linear convergence) uses at most $O(\text{poly}(d, \zeta)\log(\epsilon^{-1}))$ subgradient queries to find a point with target accuracy ϵ, i.e., $f(x) - f^* \leq \epsilon \cdot Mr$, and
- (Polynomial per-query complexity) requires only $O(\text{poly}(d, \zeta))$ arithmetic operations per query?

Answer should be **YES**, right?

- Partial solution: the case of constant curvature (hyperbolic spaces, hemispheres)

Open Problem

Q: Is there a first-order deterministic algorithm with the following properties?

- (Linear convergence) uses at most $O(\mathrm{poly}(d, \zeta) \log(\epsilon^{-1}))$ subgradient queries to find a point with target accuracy ϵ, i.e., $f(x) - f^* \leq \epsilon \cdot Mr$, and
- (Polynomial per-query complexity) requires only $O(\mathrm{poly}(d, \zeta))$ arithmetic operations per query?

Answer should be **YES**, right?

- Partial solution: the case of constant curvature (hyperbolic spaces, hemispheres)
- No information-theoretic bottleneck (centerpoint method, Rusciano'19)

Open Problem

Q: Is there a first-order deterministic algorithm with the following properties?

- (Linear convergence) uses at most $O(\mathrm{poly}(d, \zeta)\log(\epsilon^{-1}))$ subgradient queries to find a point with target accuracy ϵ, i.e., $f(x) - f^* \leq \epsilon \cdot Mr$, and
- (Polynomial per-query complexity) requires only $O(\mathrm{poly}(d, \zeta))$ arithmetic operations per query?

Answer should be **YES**, right?

- Partial solution: the case of constant curvature (hyperbolic spaces, hemispheres)
- No information-theoretic bottleneck (centerpoint method, Rusciano'19)

Note: Such algorithms exist under additional assumptions (strong g-convexity, smoothness, 2nd-order robustness) [Zhang & Sra'16, Allen-Zhu et al.'18, Hirai et al.'23]

Motivations

Ellipsoid method of fundamental theoretical importance in convex opt

Motivations

Ellipsoid method of fundamental theoretical importance in convex opt

Computing geometric median

- computational anatomy (Fletcher et al.'09)
- phylogenetics (Bacak'14)

Motivations

Ellipsoid method of fundamental theoretical importance in convex opt

Computing geometric median

- computational anatomy (Fletcher et al.'09)
- phylogenetics (Bacak'14)

Operator scaling (Gurvits'04, Burgisser'19, ...)

- robust covariance estimation
- matrix normal models
- variant on polynomial identity testing, etc.

Pos def matrices with affine-invariant metric, $\mathcal{M} = SL(n)/SO(n)$

Constant curvature

$\min_{x \in D} f(x), \;\; D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$

Goal: x with $f(x) - f^* \leq \epsilon \cdot Mr$

Constant curvature

$\min_{x \in D} f(x), \;\; D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$

Goal: x with $f(x) - f^* \leq \epsilon \cdot Mr$

Suffices to solve this problem with just $r = \frac{1}{\sqrt{K}}$

Constant curvature

Suffices to solve this problem with just $r = \frac{1}{\sqrt{K}}$

Algo

Define $R = \frac{1}{\sqrt{K}}$, $\quad x_0 = x_{\text{ref}}$,

For $k = 0, \ldots, T = \lceil \zeta \log(\epsilon^{-1}) \rceil$

$$x_{k+1} = \frac{\epsilon}{4}\text{-approx solution of subproblem } \min_{\bar{B}(x_k, R) \cap D} f$$

Constant curvature

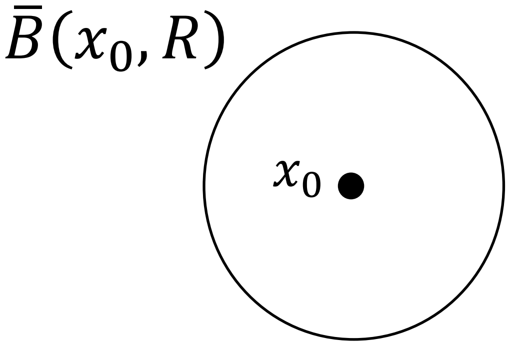
Suffices to solve this problem with just $r = \frac{1}{\sqrt{K}}$

Algo

Define $R = \frac{1}{\sqrt{K}}$, $\quad x_0 = x_{\text{ref}}$,

For $k = 0, \dots, T = \lceil \zeta \log(\epsilon^{-1}) \rceil$

$$x_{k+1} = \frac{\epsilon}{4}\text{-approx solution of subproblem } \min_{\bar{B}(x_k, R) \cap D} f$$



Constant curvature

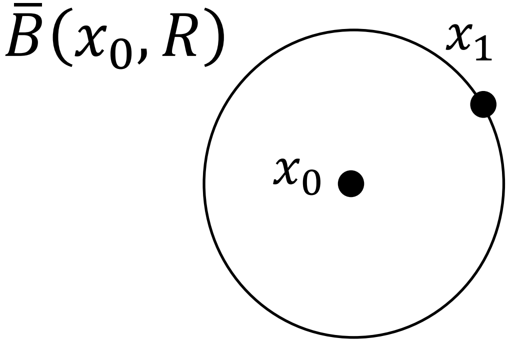
Suffices to solve this problem with just $r = \frac{1}{\sqrt{K}}$

Algo

Define $R = \frac{1}{\sqrt{K}}$, $\quad x_0 = x_{\text{ref}}$,

For $k = 0, \ldots, T = \lceil \zeta \log(\epsilon^{-1}) \rceil$

$$x_{k+1} = \frac{\epsilon}{4}\text{-approx solution of subproblem } \min_{\bar{B}(x_k, R) \cap D} f$$



Constant curvature

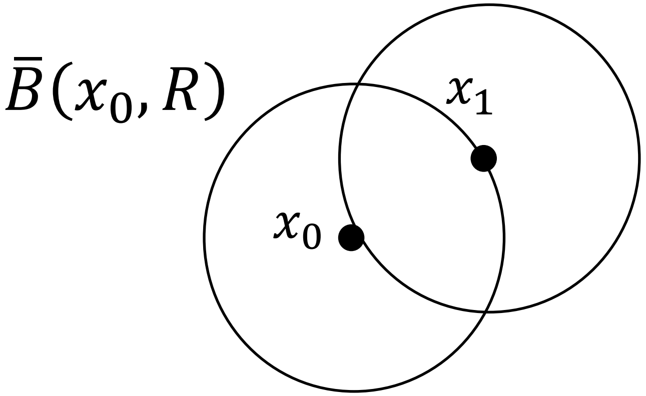
Suffices to solve this problem with just $r = \frac{1}{\sqrt{K}}$

Algo

Define $R = \frac{1}{\sqrt{K}}$, $\quad x_0 = x_{\text{ref}}$,

For $k = 0, \dots, T = \lceil \zeta \log(\epsilon^{-1}) \rceil$

$\quad x_{k+1} = \frac{\epsilon}{4}$-approx solution of subproblem $\min_{\bar{B}(x_k, R) \cap D} f$



Constant curvature

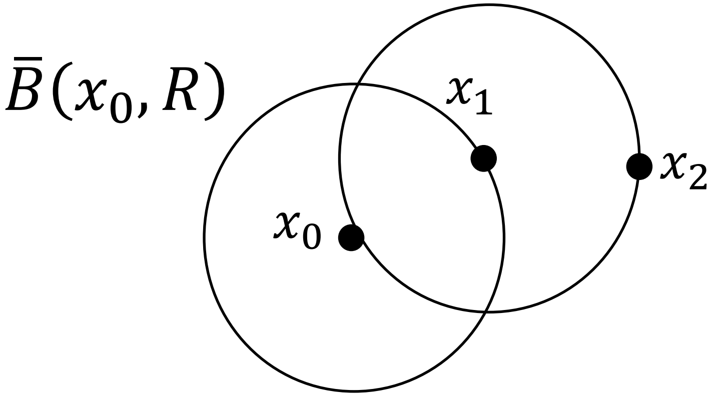
Suffices to solve this problem with just $r = \frac{1}{\sqrt{K}}$

Algo

Define $R = \frac{1}{\sqrt{K}}$, $\quad x_0 = x_{\text{ref}}$,

For $k = 0, \dots, T = \lceil \zeta \log(\epsilon^{-1}) \rceil$

$$x_{k+1} = \frac{\epsilon}{4}\text{-approx solution of subproblem } \min_{\bar{B}(x_k, R) \cap D} f$$



Constant curvature

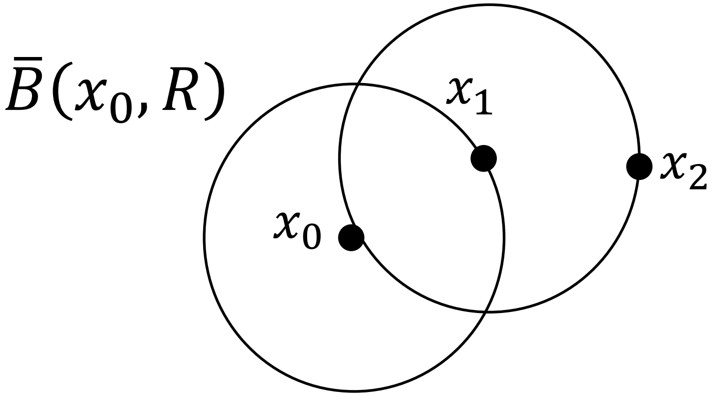
Suffices to solve this problem with just $r = \frac{1}{\sqrt{K}}$

Algo

Define $R = \frac{1}{\sqrt{K}}$, $\quad x_0 = x_{\text{ref}}$,

For $k = 0, \ldots, T = \lceil \zeta \log(\epsilon^{-1}) \rceil$

$$x_{k+1} = \frac{\epsilon}{4}\text{-approx solution of subproblem } \min_{\bar{B}(x_k, R) \cap D} f$$



- MR'23, Theorem 7
- Works for any Riemannian Manifold!

Constant curvature

$$\min_{x \in \bar{B}(x_k, R) \cap D} f(x), \text{ with } R = \frac{1}{\sqrt{K}}$$

Goal: $f(x_{k+1}) - \min_{\bar{B}(x_k, R) \cap D} f \leq \frac{\epsilon}{4} \cdot MR$

Assume $\mathcal{M} = \mathbb{H}^d$ is a hyperbolic space (constant curvature -1)

Constant curvature

$$\min_{x \in \bar{B}(x_k, R) \cap D} f(x), \text{ with } R = \frac{1}{\sqrt{K}}$$

Goal: $$f(x_{k+1}) - \min_{\bar{B}(x_k, R) \cap D} f \leq \frac{\epsilon}{4} \cdot MR$$

Assume $\mathcal{M} = \mathbb{H}^d$ is a hyperbolic space (constant curvature -1)

Key tool: Geodesic map

Constant curvature

$\min_{x \in \bar{B}(x_k, R) \cap D} f(x)$, with $R = \frac{1}{\sqrt{K}}$

Goal: $f(x_{k+1}) - \min_{\bar{B}(x_k, R) \cap D} f \leq \frac{\epsilon}{4} \cdot MR$

Assume $\mathcal{M} = \mathbb{H}^d$ is a hyperbolic space (constant curvature -1)

Key tool: Geodesic map

- Sends halfspaces in $\mathbb{H}^d$ to halfspaces in $B(0,1) \subset \mathbb{R}^d$

Constant curvature

$$\min_{x \in \bar{B}(x_k, R) \cap D} f(x), \text{ with } R = \frac{1}{\sqrt{K}}$$

Goal: $f(x_{k+1}) - \min_{\bar{B}(x_k,R)\cap D} f \leq \frac{\epsilon}{4} \cdot MR$

Assume $\mathcal{M} = \mathbb{H}^d$ is a hyperbolic space (constant curvature -1)

Key tool: Geodesic map with base point x_k is a diffeomorphism

$$\phi_k: \mathbb{H}^d \to B(0,1) \subset \mathbb{R}^d$$

with $\phi(x_k) = 0$ and which maps halfspaces in $\mathbb{H}^d$ to halfspaces in $B(0,1) \subset \mathbb{R}^d$

$$\{x \in \mathbb{H}^d : \langle g, \log_y(x) \rangle \leq 0\} \leftrightarrow \{\tilde{x} \in B(0,1) : \tilde{g}^\top(\tilde{x} - \tilde{y}) \leq 0\}$$

$$\tilde{x} = \phi_k(x), \tilde{y} = \phi_k(y)$$
$$\tilde{g} = d\phi_k(x)[g]$$

Constant curvature

$\min_{x \in \bar{B}(x_k,R) \cap D} f(x)$, with $R = \frac{1}{\sqrt{K}}$

Goal: $f(x_{k+1}) - \min_{\bar{B}(x_k,R) \cap D} f \leq \frac{\epsilon}{4} \cdot MR$

Assume $\mathcal{M} = \mathbb{H}^d$ is a hyperbolic space (constant curvature -1)

Key tool: Geodesic map with base point x_k is a diffeomorphism

$$\phi_k: \mathbb{H}^d \to B(0,1) \subset \mathbb{R}^d$$

with $\phi(x_k) = 0$ and which maps halfspaces in $\mathbb{H}^d$ to halfspaces in $B(0,1) \subset \mathbb{R}^d$

$$\left\{x \in \mathbb{H}^d : \left\langle g, \log_y(x) \right\rangle \leq 0\right\} \leftrightarrow \{\tilde{x} \in B(0,1) : \tilde{g}^\top(\tilde{x} - \tilde{y}) \leq 0\}$$

Geodesics map of $\mathbb{H}^d$ given by Beltrami Klein model (*explicit formula*)

Constant curvature

Pull back Riemannian problem to Euclidean space via geodesic map

$$\min_{\tilde{x} \in B(0,\tilde{R}) \cap \phi_k^{-1}(D)} \left(f \circ \phi_k^{-1}\right)(\tilde{x})$$

Constant curvature

Pull back Riemannian problem to Euclidean space via geodesic map

$$\min_{\tilde{x} \in B(0,\tilde{R}) \cap \phi_k^{-1}(D)} \left(f \circ \phi_k^{-1}\right)(\tilde{x})$$

Solve transformed subproblem with usual ellipsoid method!

Constant curvature

Pull back Riemannian problem to Euclidean space via geodesic map

$$\min_{\tilde{x} \in B(0,\tilde{R}) \cap \phi_k^{-1}(D)} \left(f \circ \phi_k^{-1}\right)(\tilde{x})$$

Solve transformed subproblem with usual ellipsoid method!

Why works?

- Ellipsoid method just needs halfspaces

Constant curvature

Pull back Riemannian problem to Euclidean space via geodesic map

$$\min_{\tilde{x} \in B(0,\tilde{R}) \cap \phi_k^{-1}(D)} \left(f \circ \phi_k^{-1}\right)(\tilde{x})$$

Solve transformed subproblem with usual ellipsoid method!

Why works?

- Ellipsoid method just needs halfspaces
- Use geodesic map to map hyperbolic halfspaces (from g-convexity) to Euclidean halfspaces

Constant curvature

Pull back Riemannian problem to Euclidean space via geodesic map

$$\min_{\tilde{x} \in B(0,\tilde{R}) \cap \phi_k^{-1}(D)} \left(f \circ \phi_k^{-1}\right)(\tilde{x})$$

Solve transformed subproblem with usual ellipsoid method!

Why works?

- Ellipsoid method just needs halfspaces
- Use geodesic map to map hyperbolic halfspaces (from g-convexity) to Euclidean halfspaces
- $R = 1/\sqrt{K}$ is small, so the distortion is small (e.g., for Lipschitzness)

Constant curvature

Pull back Riemannian problem to Euclidean space via geodesic map

$$\min_{\tilde{x} \in B(0,\tilde{R}) \cap \phi_k^{-1}(D)} \left(f \circ \phi_k^{-1}\right)(\tilde{x})$$

Solve transformed subproblem with usual ellipsoid method!

Why works?

- Ellipsoid method just needs halfspaces
- Use geodesic map to map hyperbolic halfspaces (from g-convexity) to Euclidean halfspaces
- $R = 1/\sqrt{K}$ is small, so the distortion is small (e.g., for Lipschitzness)

Takes $O(d^2 \log(\epsilon^{-1}))$ queries, each requiring $O(d^2)$ arithmetic operations.

How to go beyond constant curvature?

Beltrami's theorem: The only Riemannian manifolds which admit geodesic maps to Euclidean space are those of constant curvature.

How to go beyond constant curvature?

Beltrami's theorem: The only Riemannian manifolds which admit geodesic maps to Euclidean space are those of constant curvature.

Maybe replace geodesic maps with exponential map and use comparison theorems? Not clear …

Other possible approaches

Difficulty: there is no convenient notion of ellipsoid on general manifolds

Other possible approaches

Difficulty: there is no convenient notion of ellipsoid on general manifolds

One idea: Maintain ellipsoids in tangent spaces.

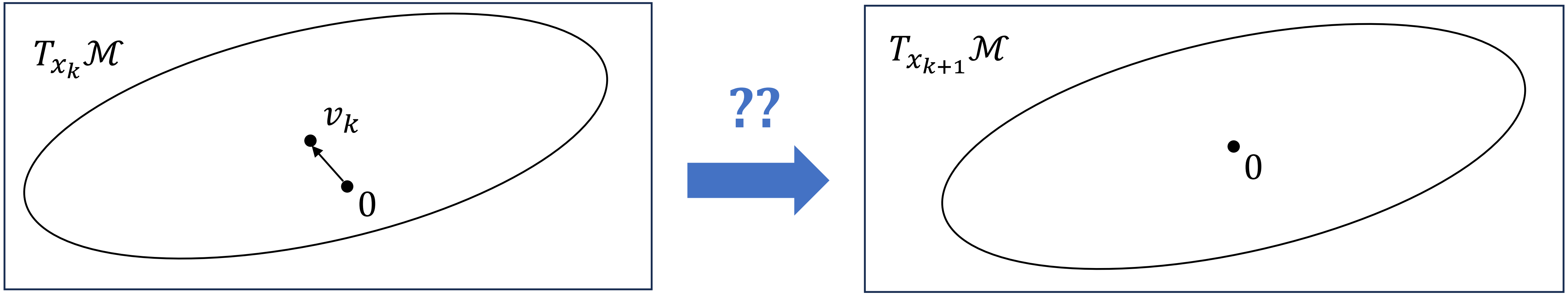
Other possible approaches

Difficulty: there is no convenient notion of ellipsoid on general manifolds

One idea: Maintain ellipsoids in tangent spaces.

How to transfer ellipsoids between tangent spaces?

$x_{k+1} = \exp_{x_k}(v_k)$



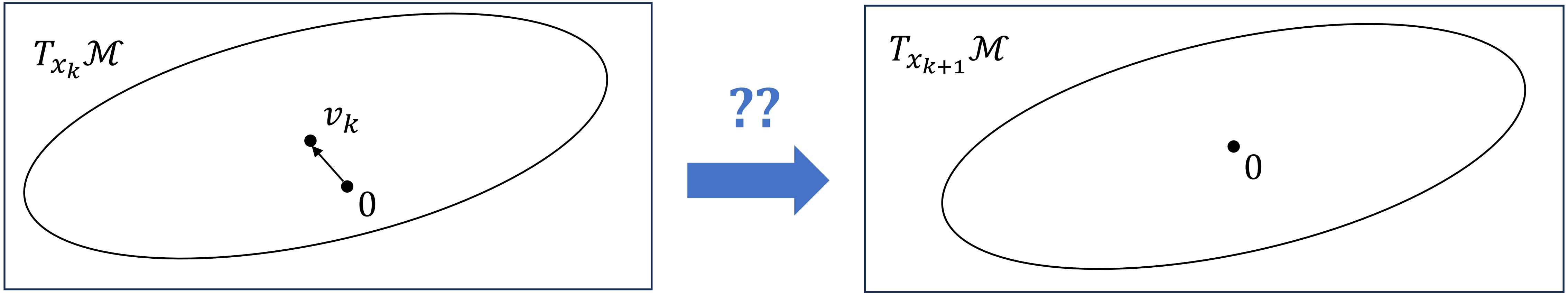
Other possible approaches

Difficulty: there is no convenient notion of ellipsoid on general manifolds

One idea: Maintain ellipsoids in tangent spaces.

How to transfer ellipsoids between tangent spaces?

$x_{k+1} = \exp_{x_k}(v_k)$



Kim and Yang'22 introduced a way of transferring balls between tangent spaces.

It is not clear how to generalize their results to ellipsoids.

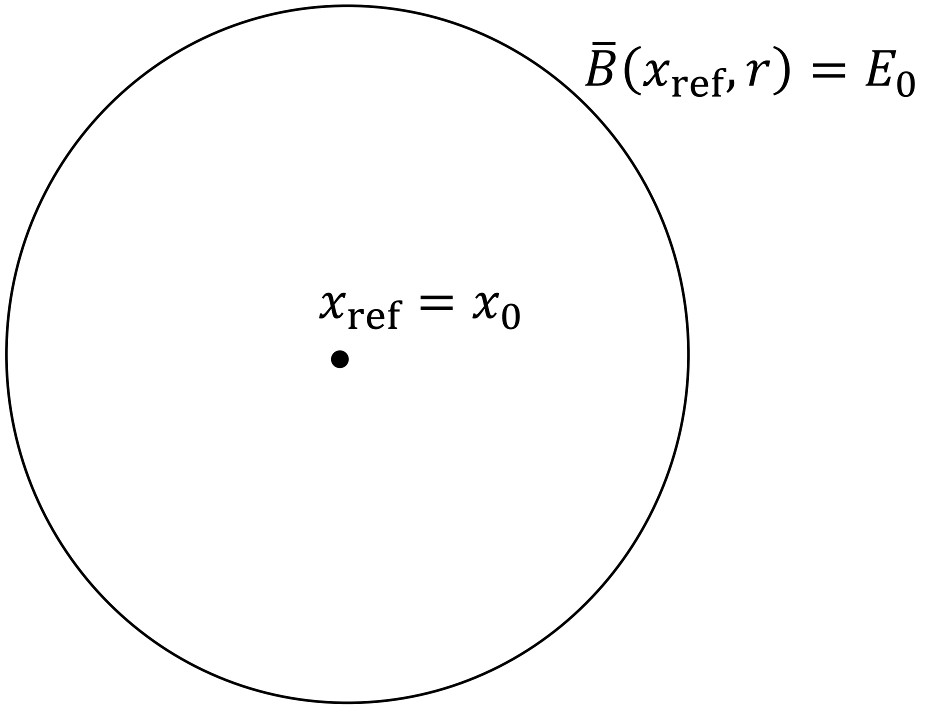
Appendix

Ellipsoid method

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\mathrm{ref}}, r) \subset \mathcal{M}$$

If $\mathcal{M} = \mathbb{R}^d$, amounts to Lipschitz convex optimization.

Ellipsoid method:

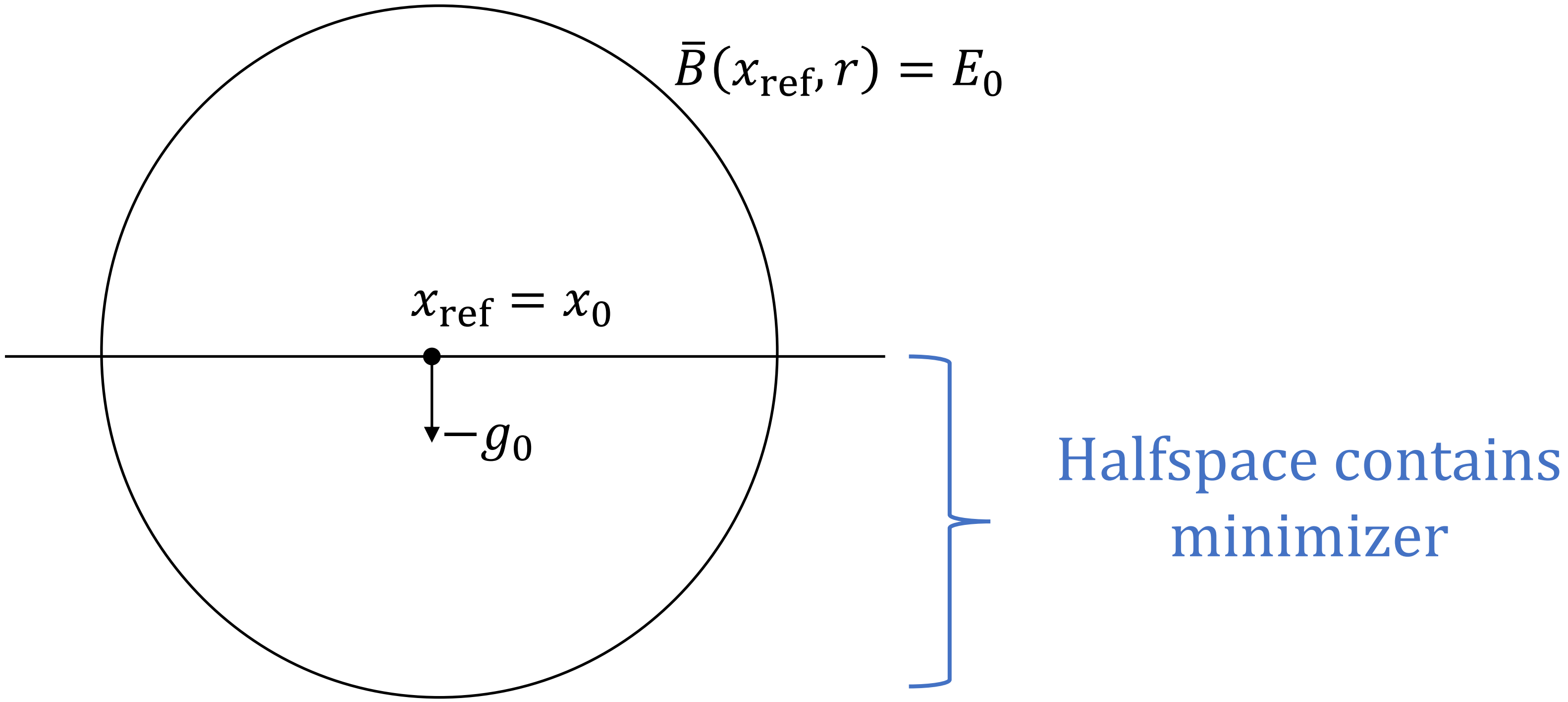


Ellipsoid method

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$$

If $\mathcal{M} = \mathbb{R}^d$, amounts to Lipschitz convex optimization.

Ellipsoid method:

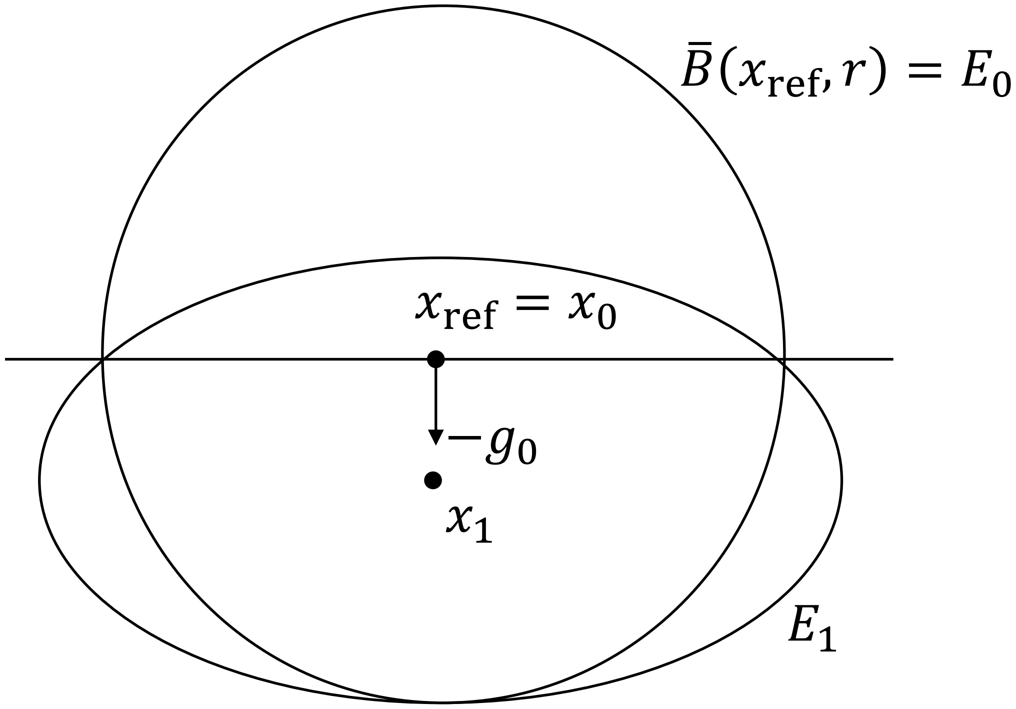


Ellipsoid method

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$$

If $\mathcal{M} = \mathbb{R}^d$, amounts to Lipschitz convex optimization.

Ellipsoid method:



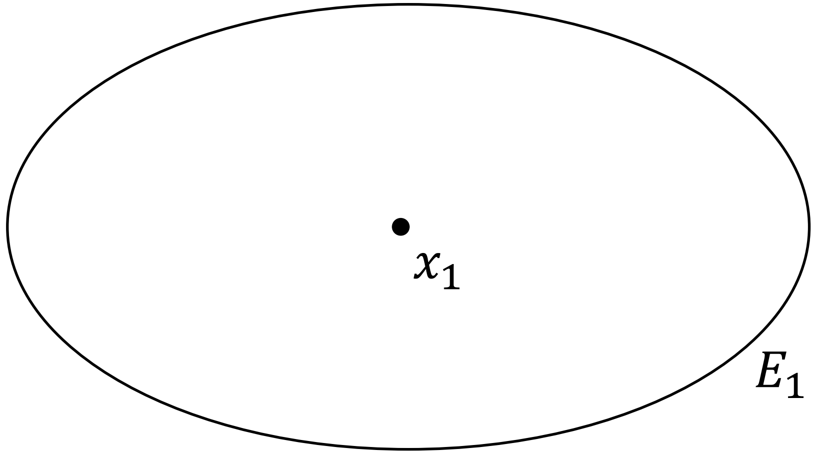
E_1 is minimum-volume ellipsoid containing intersection of halfspace and ball E_0

Ellipsoid method

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$$

If $\mathcal{M} = \mathbb{R}^d$, amounts to Lipschitz convex optimization.

Ellipsoid method:

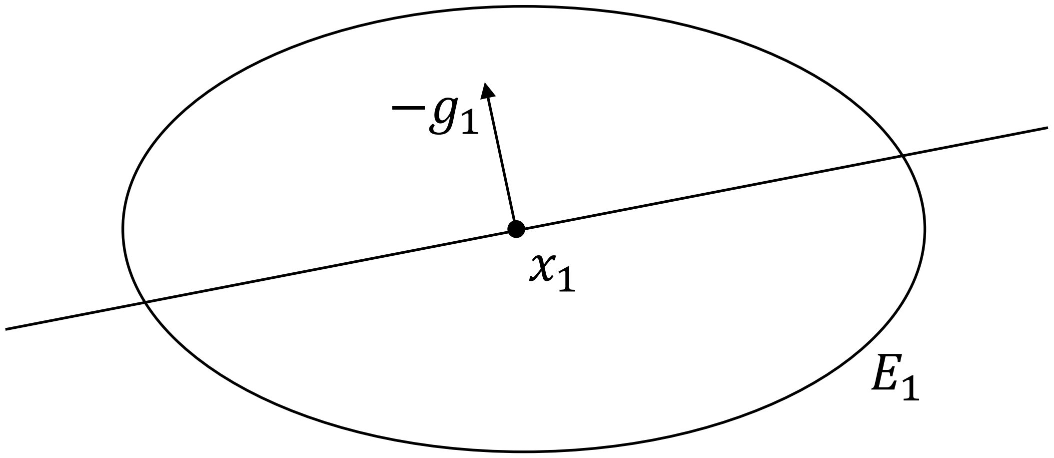


Ellipsoid method

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\mathrm{ref}}, r) \subset \mathcal{M}$$

If $\mathcal{M} = \mathbb{R}^d$, amounts to Lipschitz convex optimization.

Ellipsoid method:

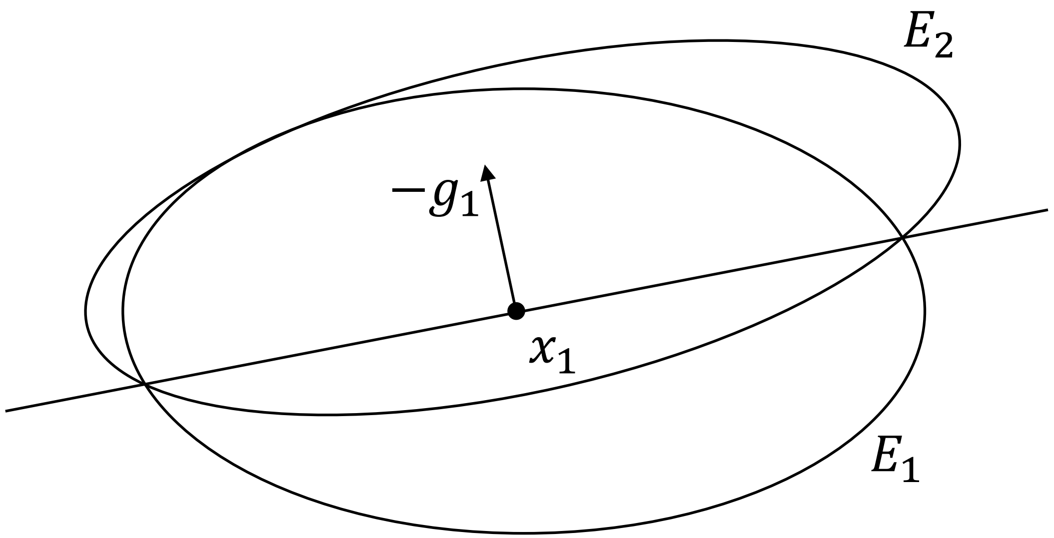


Ellipsoid method

$$\min_{x \in D} f(x), \qquad D = \bar{B}(x_{\text{ref}}, r) \subset \mathcal{M}$$

If $\mathcal{M} = \mathbb{R}^d$, amounts to Lipschitz convex optimization.

Ellipsoid method:



E_2 is minimum-volume ellipsoid containing intersection of halfspace and E_1