Horo-Convex Optimization on Hadamard Manifolds

EUROPT 2024

Chris Criscitiello (EPFL)

Jungbin Kim (Seoul National University)

EPFL

Setting

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

$\mathcal{M}$ is a Hadamard manifold

f is geodesically convex (g-convex) or horospherically convex (horo-convex)

Setting

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

$\mathcal{M}$ is a Hadamard manifold

f is geodesically convex (g-convex) or horospherically convex (horo-convex)

- Setting + motivation
- Issues with g-convexity
- Horo-convexity, its nice properties and limitations
  - Nec&suff interpolation conditions, complexity guarantees, etc.
  - Generalization of (Euclidean) convexity

Setting

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

$\mathcal{M}$ is a Hadamard manifold

f is geodesically convex (g-convex) or horospherically convex (horo-convex)

- Setting + motivation
- Issues with g-convexity
- Horo-convexity, its nice properties and limitations
  - Nec&suff interpolation conditions, complexity guarantees, etc.
  - Generalization of (Euclidean) convexity

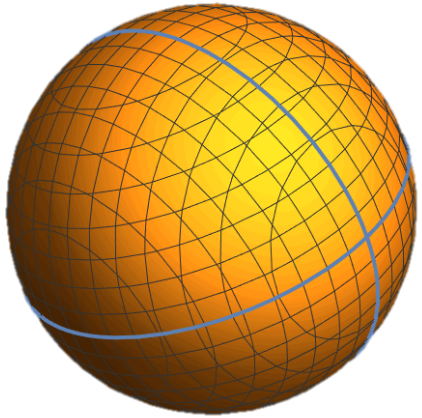
Setting: $\mathcal{M}$ is a Hadamard manifold

$\mathcal{M}$ is a complete, simply connected Riemannian manifold
with non-positive curvature.

Setting: $\mathcal{M}$ is a Hadamard manifold

$\mathcal{M}$ is a complete, simply connected Riemannian manifold with non-positive curvature.

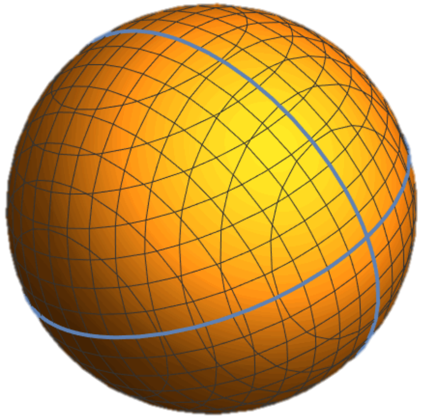
Non-example: Sphere



Setting: $\mathcal{M}$ is a Hadamard manifold

$\mathcal{M}$ is a complete, simply connected Riemannian manifold with non-positive curvature.

Non-example: Sphere

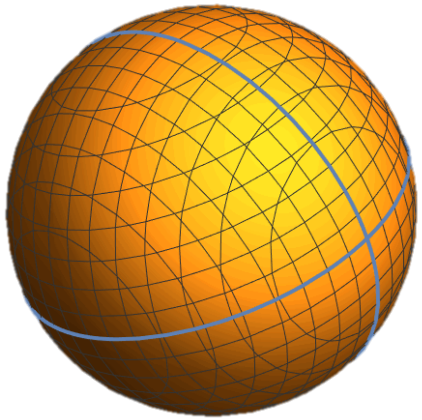


Euclidean space: $\mathcal{M} = \mathbb{R}^d$

Setting: $\mathcal{M}$ is a Hadamard manifold

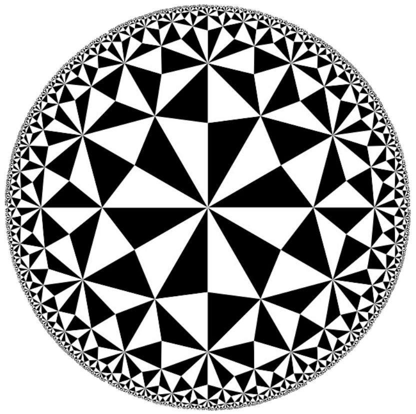
$\mathcal{M}$ is a complete, simply connected Riemannian manifold with non-positive curvature.

Non-example: Sphere



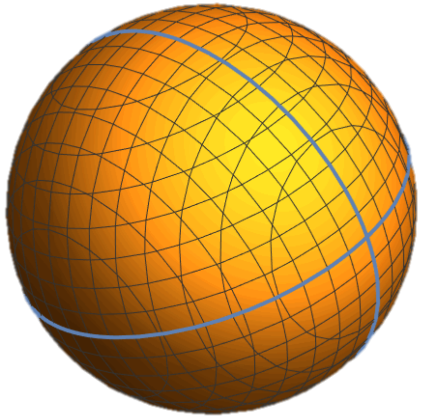
Euclidean space: $\mathcal{M} = \mathbb{R}^d$

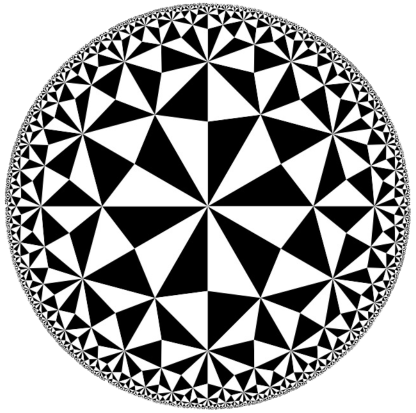
Hyperbolic space



Setting: $\mathcal{M}$ is a Hadamard manifold

$\mathcal{M}$ is a complete, simply connected Riemannian manifold with non-positive curvature.





Non-example: Sphere

Euclidean space: $\mathcal{M} = \mathbb{R}^d$

Hyperbolic space

Positive definite matrices: $\mathcal{M} = \{\Sigma \in \mathbf{R}^{d\times d} : \Sigma = \Sigma^\top \text{ and } \Sigma \succ 0\}$ with affine-invariant metric $\langle X, Y\rangle_\Sigma = \mathrm{Tr}(\Sigma^{-1} X \Sigma^{-1} Y)$.

Applications of g-convexity

Minimal enclosing ball (MEB) of set $C \subseteq \mathcal{M}$ (e.g., Arnaudon + Nielsen)

$$\min_{x \in \mathcal{M}} \max_{y \in C} \operatorname{dist}(x, y)$$

Applications of g-convexity

Minimal enclosing ball (MEB) of set $C \subseteq \mathcal{M}$ (e.g., Arnaudon + Nielsen)

$$\min_{x \in \mathcal{M}} \max_{y \in C} \operatorname{dist}(x, y)$$

Fréchet means and medians (e.g., Karcher): $\min_{x \in \mathcal{M}} \sum_i \operatorname{dist}(x, x_i)^p$

Applications of g-convexity

Minimal enclosing ball (MEB) of set $C \subseteq \mathcal{M}$ (e.g., Arnaudon + Nielsen)

$$\min_{x \in \mathcal{M}} \max_{y \in C} \operatorname{dist}(x, y)$$

Fréchet means and medians (e.g., Karcher): $\min_{x \in \mathcal{M}} \sum_i \operatorname{dist}(x, x_i)^p$

Tyler's M-estimator for robust covariance estimation (Weisel + Zhang, Franks + Moitra):

$$\min_{\Sigma \in \mathrm{PD}(d)} \sum_i \log\det\Sigma + d \log\left(x_i^\top \Sigma^{-1} x_i\right)$$

Applications of g-convexity

Minimal enclosing ball (MEB) of set $C \subseteq \mathcal{M}$ (e.g., Arnaudon + Nielsen)

$$\min_{x \in \mathcal{M}} \max_{y \in C} \operatorname{dist}(x, y)$$

Fréchet means and medians (e.g., Karcher): $\min_{x \in \mathcal{M}} \sum_i \operatorname{dist}(x, x_i)^p$

Tyler's M-estimator for robust covariance estimation (Weisel + Zhang, Franks + Moitra):

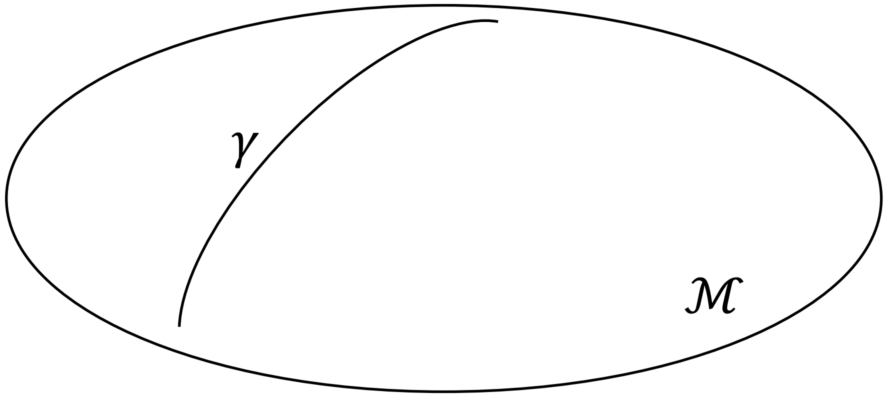
$$\min_{\Sigma \in \mathrm{PD}(d)} \sum_i \log\det\Sigma + d \log\left(x_i^\top \Sigma^{-1} x_i\right)$$

Operator scaling, tensor scaling, Horn's problem, Brascamp-Lieb constants, …

- "Towards a theory of non-commutative optimization" -- Bürgisser, Franks, Garg, Oliveira, Walter, Wigderson

Setting: f is g-convex

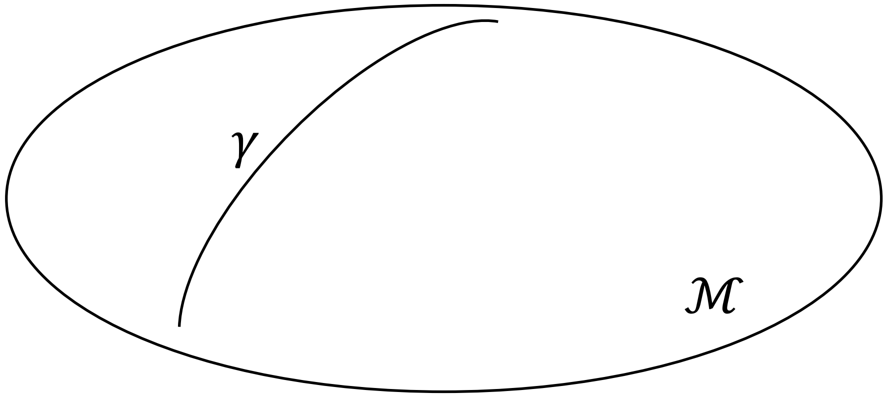
$$\min_{x \in \mathcal{M}} f(x)$$



Geodesic: the analog of straight lines on manifolds

Setting: f is g-convex

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

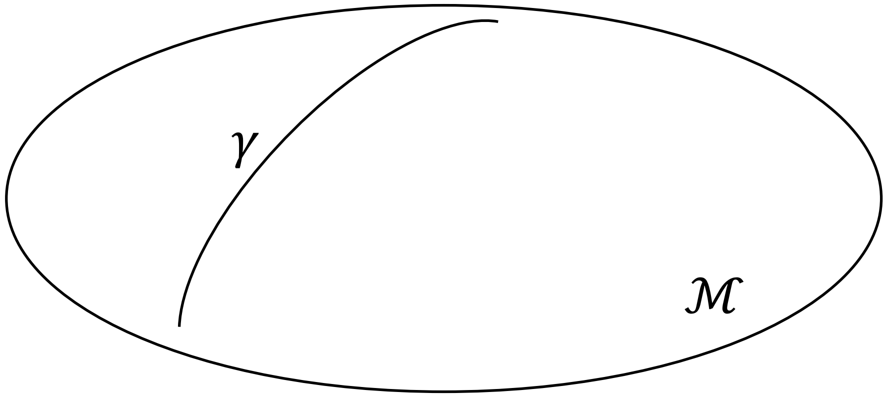


Geodesic: the analog of straight lines on manifolds

f is g-convex if its restriction $f \circ \gamma$ to any geodesic $t \mapsto \gamma(t)$ is convex

Setting: f is g-convex

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



Geodesic: the analog of straight lines on manifolds

f is g-convex if its restriction $f \circ \gamma$ to any geodesic $t \mapsto \gamma(t)$ is convex

Alternative characterization: for all $x, y \in \mathcal{M}$

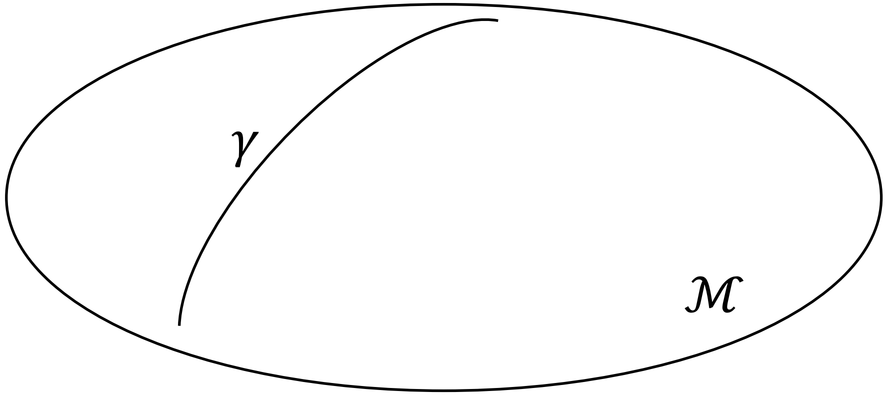
$$f(y) \geq f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$$

Exponential map: for $x \in \mathcal{M}$ and $v \in \mathrm{T}_x\mathcal{M}$,

$$\gamma(t) = \exp_x(tv) \text{ is the geodesic with } \gamma(0) = x,\ \gamma'(0) = v.$$

Setting: f is g-convex

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



Geodesic: the analog of straight lines on manifolds

f is g-convex if its restriction $f \circ \gamma$ to any geodesic $t \mapsto \gamma(t)$ is g-convex

Alternative characterization: for all $x, y \in \mathcal{M}$

$$f(y) \geq f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$$

Exponential map: for $x \in \mathcal{M}$ and $v \in \mathrm{T}_x\mathcal{M}$,

$$\gamma(t) = \exp_x(tv) \text{ is the geodesic with } \gamma(0) = x,\ \gamma'(0) = v.$$

A few “issues” with g-convexity

Alternative characterization: for all $x, y \in \mathcal{M}$

$$f(y) \geq f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$$

Each “basis” function $y \mapsto f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$ is NOT g-convex!

A few "issues" with g-convexity

Each "basis" function $y \mapsto f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$ is NOT g-convex!

A few “issues” with g-convexity

Each “basis” function $y \mapsto f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$ is NOT g-convex!

No necessary and sufficient interpolation conditions

A few “issues” with g-convexity

Each “basis” function $y \mapsto f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$ is NOT g-convex!

No necessary and sufficient interpolation conditions

G-convex hulls of even 3 points are not even closed, for “generic” manifolds (Lytchak+Petrunin 2021)

A few “issues” with g-convexity

Each “basis” function $y \mapsto f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$ is NOT g-convex!

No necessary and sufficient interpolation conditions

G-convex hulls of even 3 points are not even closed, for “generic” manifolds (Lytchak+Petrunin 2021)

Complexity of all algorithms depend on curvature ζ of the manifold, e.g.:

- Sub-gradient descent: $O(\zeta/\epsilon^2)$
- Nesterov acceleration: $O\left(\sqrt{\zeta/\epsilon}\right)$

(To find a point with relative accuracy $\epsilon \in (0,1)$.)

$$\zeta = \frac{r\sqrt{-K}}{\tanh(r\sqrt{-K})}$$

K = sectional curvature

A few "issues" with g-convexity

Each "basis" function $y \mapsto f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$ is NOT g-convex!

No necessary and sufficient interpolation conditions

G-convex hulls of even 3 points are not even closed, for "generic" manifolds (Lytchak+Petrunin 2021)

Complexity of all algorithms depend on curvature ζ of the manifold, e.g.:

- Sub-gradient descent: $O(\zeta/\epsilon^2)$
- Nesterov acceleration: $O\left(\sqrt{\zeta/\epsilon}\right)$

(To find a point with relative accuracy $\epsilon \in (0,1)$.)

Can really be important!

Tensor scaling, Horn's problem, Brascamp-Lieb constants, …

A few “issues” with g-convexity

Each “basis” function $y \mapsto f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$ is NOT g-convex!

No ne

G-con s

(Lytch

Next:

- We propose and study a function class which generalizes Euclidean convexity, and resolves all these issues
- Downside: the function class is smaller, so covers less examples

Complexity of all algorithms depend on curvature ζ of the manifold, e.g.:

- Sub-gradient descent: $O(\zeta/\epsilon^2)$
- Nesterov acceleration: $O\left(\sqrt{\zeta/\epsilon}\right)$

(To find a point with relative accuracy $\epsilon \in (0,1)$.)

Can really be important!

Tensor scaling, Horn’s problem, Brascamp-Lieb constants, ...

Horo-convexity

Starting observations:

- Each "basis" function $y \mapsto f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$ is NOT g-convex
- μ-strongly convex functions are suprema of squared distance functions:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|y - x\|^2$$

$$= f(x) - \frac{1}{2\mu} \|\nabla f(x)\|^2 + \frac{\mu}{2} \|y - x^{++}\|^2 \qquad \text{where } x^{++} = x - \frac{1}{\mu} \nabla f(x).$$

- Squared distance $x \mapsto \frac{\mu}{2} \text{dist}^2(x, p)$ is μ-strongly g-convex.

Horo-convexity

Starting observations:

- Each "basis" function $y \mapsto f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$ is NOT g-convex
- μ-strongly convex functions are suprema of squared distance functions:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2$$

$$= f(x) - \frac{1}{2\mu}\|\nabla f(x)\|^2 + \frac{\mu}{2}\|y - x^{++}\|^2 \qquad \text{where } x^{++} = x - \frac{1}{\mu}\nabla f(x).$$

- Squared distance $x \mapsto \frac{\mu}{2}\text{dist}^2(x, p)$ is μ-strongly g-convex.

(**Definition**) For $\mu > 0$, f is μ-strongly horo-convex if it is a supremum of squared distance functions of the form

$$\mathrm{y} \mapsto (const) + \frac{\mu}{2}\text{dist}^2(y, x).$$

Horo-convexity

Starting observations:

- Each "basis" function $y \mapsto f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$ is NOT g-convex
- μ-strongly convex functions are suprema of squared distance functions:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|y - x\|^2$$

$$= f(x) - \frac{1}{2\mu} \|\nabla f(x)\|^2 + \frac{\mu}{2} \|y - x^{++}\|^2 \qquad \text{where } x^{++} = x - \frac{1}{\mu} \nabla f(x).$$

- Squared distance $x \mapsto \frac{\mu}{2} \text{dist}^2(x, p)$ is μ-strongly g-convex.

(**Definition**) For $\mu > 0$, f is μ-strongly horo-convex if it is a supremum of squared distance functions of the form

$$\mathrm{y} \mapsto (const) + \frac{\mu}{2} \text{dist}^2(y, x).$$

If f is differentiable, f is μ-strongly horo-convex iff

$$f(y) \geq f(x) - \frac{1}{2\mu} \|\nabla f(x)\|^2 + \frac{\mu}{2} \text{dist}^2(y, x^{++})$$

where $x^{++} = \exp_x\left(-\frac{1}{\mu} \nabla f(x)\right)$.

Horo-convexity

Starting observations:

- Each "basis" function $y \mapsto f(x) + \langle \nabla f(x), \exp_x^{-1}(y) \rangle$ is NOT g-convex
- μ-strongly convex functions are suprema of squared distance functions:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2$$

$$= f(x) - \frac{1}{2\mu}\|\nabla f(x)\|^2 + \frac{\mu}{2}\|y - x^{++}\|^2 \qquad \text{where } x^{++} = x - \frac{1}{\mu}\nabla f(x).$$

- Squared distance $x \mapsto \frac{\mu}{2}\mathrm{dist}^2(x, p)$ is μ-strongly g-convex.

(**Definition**) For $\mu > 0$, f is μ-strongly horo-convex if it is a supremum of squared distance functions of the form

$$y \mapsto (const) + \frac{\mu}{2}\mathrm{dist}^2(y, x).$$

If f is differentiable, f is μ-strongly horo-convex iff

$$f(y) \geq f(x) - \frac{1}{2\mu}\|\nabla f(x)\|^2 + \frac{\mu}{2}\mathrm{dist}^2(y, x^{++})$$

where $x^{++} = \exp_x\left(-\frac{1}{\mu}\nabla f(x)\right)$.

$\mu = 0$? Take the limit!

Horo-convexity

(**Definition**) Given any unit-speed geodesic γ, the corresponding Busemann function is

$$B_\gamma(x) = \lim_{t\to\infty} (\text{dist}(\gamma(t), x) - t)$$

Horo-convexity

(**Definition**) Given any unit-speed geodesic γ, the corresponding Busemann function is

$$B_\gamma(x) = \lim_{t\to\infty} (\mathrm{dist}(\gamma(t), x) - t)$$

- Very well-studied with remarkable properties – g-convex, 1-Lipschitz, gradient flow is geodesics, and more! (Ballman 1995, Bridson+Haefliger 2013)
- Generalizes affine functions! (with gradient norm = 1)

Horo-convexity

(**Definition**) Given any unit-speed geodesic γ, the corresponding Busemann function is

$$B_\gamma(x) = \lim_{t \to \infty} (\text{dist}(\gamma(t), x) - t)$$

- Very well-studied with remarkable properties – g-convex, 1-Lipschitz, gradient flow is geodesics, and more! (Ballman 1995, Bridson+Haefliger 2013)
- Generalizes affine functions! (with gradient norm = 1)

(**Definition**) f is horo-convex if it is a supremum of scaled Busemann functions of the form

$$x \mapsto a\, B_\gamma(x), a \geq 0$$

Horo-convexity

(**Definition**) Given any unit-speed geodesic γ, the corresponding Busemann function is

$$B_\gamma(x) = \lim_{t\to\infty}(\mathrm{dist}(\gamma(t), x) - t)$$

- Very well-studied with remarkable properties – g-convex, 1-Lipschitz, gradient flow is geodesics, and more! (Ballman 1995, Bridson+Haefliger 2013)
- Generalizes affine functions! (with gradient norm = 1)

(**Definition**) f is horo-convex if it is a supremum of scaled Busemann functions of the form

$$x \mapsto a\, B_\gamma(x), a \geq 0$$

If f is differentiable, f is horo-convex iff

$$f(y) \geq f(x) + B_{x,[\nabla f(x)]}(y).$$

Horo-convexity

(**Definition**) Given any unit-speed geodesic γ, the corresponding Busemann function is

$$B_\gamma(x) = \lim_{t\to\infty} (\mathrm{dist}(\gamma(t), x) - t)$$

- Very well-studied with remarkable properties – g-convex, 1-Lipschitz, gradient flow is geodesics, and more! (Ballman 1995, Bridson+Haefliger 2013)
- Generalizes affine functions! (with gradient norm = 1)

(**Definition**) f is horo-convex if it is a supremum of scaled Busemann functions of the form

$$x \mapsto a\, B_\gamma(x), a \geq 0$$

If f is differentiable, f is horo-convex iff

$$f(y) \geq f(x) + B_{x,[\nabla f(x)]}(y).$$

Horoballs are the sublevel sets of Busemann functions.

- They are distinct from geodesic half-spaces!

Horo-convexity

(**Definition**) Given any unit-speed geodesic γ, the corresponding Busemann function is

$$B_\gamma(x) = \lim (\mathrm{dist}(\gamma(t), x) - t)$$

"Horoballs and the subgradient method" -- Lewis, Lopez-Acedo, Nicolae 2024

"Horoball Hulls and Extents in Positive Definite Space" -- Fletcher et al. 2011

"Abstract convexity and global optimization" -- Rubinov

- Very wel … nt flow is geodesic
- Generali

(Definitio … s of the form

If f is diffe

$$f(y) \geq f(x) + B_{x,[\nabla f(x)]}(y).$$

Horoballs are the sublevel sets of Busemann functions.

- They are distinct from geodesic half-spaces!

A few properties of horo-convexity

When $\mathcal{M} = \mathbb{R}^d$, is just Euclidean convexity.

All horo-convex functions are g-convex (but not vice versa!)

A few properties of horo-convexity

When $\mathcal{M} = \mathbb{R}^d$, is just Euclidean convexity.

All horo-convex functions are g-convex (but not vice versa!)

Distances $x \mapsto \mathrm{dist}^p(x, y)$ are horo-convex, $p \geq 1$.

A few properties of horo-convexity

When $\mathcal{M} = \mathbb{R}^d$, is just Euclidean convexity.

All horo-convex functions are g-convex (but not vice versa!)

Distances $x \mapsto \mathrm{dist}^p(x, y)$ are horo-convex, $p \geq 1$.

Moreau envelope of horo-convex function gives a smooth horoconvex function (with all the usual properties)

$$f_t(x) = \min_{y \in \mathcal{M}} \left\{ f(y) + \frac{1}{2t} \mathrm{dist}^2(x, y) \right\}$$

A few properties of horo-convexity

When $\mathcal{M} = \mathbb{R}^d$, is just Euclidean convexity.

All horo-convex functions are g-convex (but not vice versa!)

Distances $x \mapsto \mathrm{dist}^p(x, y)$ are horo-convex, $p \geq 1$.

Moreau envelope of horo-convex function gives a smooth horoconvex function (with all the usual properties)

$$f_t(x) = \min_{y \in \mathcal{M}} \left\{ f(y) + \frac{1}{2t} \mathrm{dist}^2(x, y) \right\}$$

Big caveat: not closed under addition! (unlike g-convex functions)

Applications of g-convexity

Minimal enclosing ball (MEB) of set $C \subseteq \mathcal{M}$ (e.g., Arnaudon + Nielsen)

$$\min_{x \in \mathcal{M}} \max_{y \in C} \operatorname{dist}(x, y)$$

Horoconvex!

Fréchet means and medians (e.g., Karcher): $\min_{x \in \mathcal{M}} \sum_i \operatorname{dist}(x, x_i)^p$

Sum of horoconvex!

Tyler's M-estimator for robust covariance estimation (Weisel + Zhang, Franks + Moitra):

$$\min_{\Sigma \in \mathrm{PD}(d)} \sum_i \log\det\Sigma + d\log\left(x_i^\top \Sigma^{-1} x_i\right)$$

Sum of horoconvex!

Operator scaling, tensor scaling, Horn's problem, Brascamp-Lieb constants, …

- "Towards a theory of non-commutative optimization" -- Bürgisser, Franks, Garg, Oliveira, Walter, Wigderson

Complexity of horo-convex optimization

$\min_{x \in \mathcal{M}} f(x) = \sum_{i \in [1,m]} f_i(x)$ where each f_i is horo-convex

Complexity of horo-convex optimization

$$\min_{x \in \mathcal{M}} f(x) = \sum_{i \in [1,m]} f_i(x) \qquad \text{where each } f_i \text{ is horo-convex}$$

- Oracle access to $(f_1(x), \nabla f_1(x)), \ldots, (f_m(x), \nabla f_m(x))$

Complexity of horo-convex optimization

$$\min_{x \in \mathcal{M}} f(x) = \sum_{i \in [1,m]} f_i(x) \qquad \text{where each } f_i \text{ is horo-convex}$$

- Oracle access to $(f_1(x), \nabla f_1(x)), \dots, (f_m(x), \nabla f_m(x))$

All convex algorithms generalize easily to the horo-convex case,
with exactly the same rates, and same step sizes!

Subgradient descent $x_{k+1} = \exp_{x_k}(-\eta_k g_k)$: $O(1/\epsilon^2)$
Nesterov accelerated gradient descent: $O\left(\sqrt{1/\epsilon}\right)$

No dependence on curvature ζ!

Complexity of horo-convex optimization

$$\min_{x \in \mathcal{M}} f(x) = \sum_{i \in [1,m]} f_i(x) \qquad \text{where each } f_i \text{ is horo-convex}$$

- Oracle access to $(f_1(x), \nabla f_1(x)), \ldots, (f_m(x), \nabla f_m(x))$

All convex algorithms generalize easily to the horo-convex case,
with exactly the same rates, and same step sizes!

Subgradient descent $x_{k+1} = \exp_{x_k}(-\eta_k g_k)$: $O(1/\epsilon^2)$
Nesterov accelerated gradient descent: $O\left(\sqrt{1/\epsilon}\right)$

No dependence on curvature ζ!

For **MEB**: get the best known rate of $\min\left\{\frac{n}{\epsilon^2}, \frac{n\sqrt{\log(n)}\cdot poly(\zeta)}{\epsilon}\right\}$ to reach relative accuracy ϵ.

Interpolation: g-convex

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.

Interpolation: g-convex

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

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

Interpolation: g-convex

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.

Interpolation: horo-convex

Nec&suff interpolation conditions for horo-convex functions:

$$F_j \geq F_i + B_{x_i, g_i}(x_j) \text{ for all } i, j$$

Interpolation: horo-convex

Nec&suff interpolation conditions for horo-convex functions:

$$F_j \geq F_i + B_{x_i, g_i}(x_j) \text{ for all } i, j$$

Nec&suff interpolation conditions for L-smooth horo-convex function:

$$F_j \geq F_i + \frac{\|g_i\|^2}{2L} + \frac{\|g_j\|^2}{2L} + B_{x_i, g_i}\left(\exp_{x_j}\left(-\frac{g_j}{L}\right)\right) \quad \text{for all } i, j.$$

Interpolation: horo-convex

Nec&suff interpolation conditions for horo-convex functions:

$$F_j \geq F_i + B_{x_i,g_i}(x_j) \text{ for all } i, j$$

Nec&suff interpolation conditions for L-smooth horo-convex function:

$$F_j \geq F_i + \frac{\|g_i\|^2}{2L} + \frac{\|g_j\|^2}{2L} + B_{x_i,g_i}\left(\exp_{x_j}\left(-\frac{g_j}{L}\right)\right) \quad \text{for all } i, j.$$

Surprising because no known notion of Fenchel conjugate!

Obtained through the Moreau envelope.

Conclusions and future directions

Summary:

- Proposed and studied a generalization of convex functions to Hadamard manifolds
- Geometrically very natural
- Curvature independent rates, and faster algos for MEB
- Nec&suff interpolation conditions

Conclusions and future directions

Summary:

- Proposed and studied a generalization of convex functions to Hadamard manifolds
- Geometrically very natural
- Curvature independent rates, and faster algos for MEB
- Nec&suff interpolation conditions

Questions:

- Duality? (boundary at infinity …)
- PEP for horo-convex functions? (to help answer lower bound questions?)
- Inner characterization of horo-convexity?
- Original motivation: Operator scaling, tensor scaling, Horn's problem, Brascamp-Lieb constants, …

Appendix

The oracle

5.2 Gradient Methods

In this subsection, we study a generalization of gradient descent for solving the h-convex optimization problem (7). On the Euclidean space $M = \mathbb{R}^n$, given step sizes $s_k > 0$, the gradient descent updates iterates as $x_{k+1} = x_k - s_k \nabla f(x_k)$, or equivalently:[8]

$$x_{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ \frac{1}{2s_k} \|x_k - x\|^2 + \langle \nabla f(x_k), x - x_k \rangle \right\}.$$

We generalize this algorithm to arbitrary Hadamard manifolds M by replacing the squared norm with the squared distance on manifolds, and replacing the linear function $\langle \nabla f_i(x_k), x - x_k \rangle$ with the scaled Busemann function $B_{\nabla f_i(x_k)}(\overrightarrow{x_k x})$:

$$x_{k+1} = \operatorname{argmin}_{x \in M} \left\{ \frac{1}{2s_k} d(x_k, x)^2 + \frac{1}{m} \sum_{i=1}^{m} B_{\nabla f_i(x_k)}(\overrightarrow{x_k x}) \right\}. \tag{8}$$

When the objective function is a single h-convex function ($m = 1$), this algorithm simplifies to $x_{k+1} = \exp_{x_k}(-s_k \nabla f(x_k))$, a well-known form of the Riemannian gradient descent. While the oracle complexity of gradient descent for minimizing g-convex functions was studied to depend on the lower bound of sectional curvature (Zhang and Sra, 2016), we show here that curvature-free convergence rates are achievable when the objective function is the sum of h-convex functions.

Nesterov Accelerated Gradient Method

Non-strongly convex case. When $\mu = 0$, the accelerated gradient method updates the iterates as follows:

$$
\begin{aligned}
y_k &= x_k + \frac{2}{k+1}(z_k - x_k) \\
x_{k+1} &= y_k - \frac{1}{L}\nabla f(y_k) \\
z_{k+1} &= z_k - \frac{k+1}{2L}\nabla f(y_k),
\end{aligned} \tag{11}
$$

starting at an initial point $x_0 = z_0$. To solve the h-convex optimization problem (7), we propose the following generalization of this algorithm:

$$
\begin{aligned}
y_k &= \exp_{x_k}\left(\frac{2}{k+1}\log_{x_k}(z_k)\right) \\
x_{k+1} &= \exp_{y_k}\left(-\frac{1}{L}\nabla f(y_k)\right) \\
z_{k+1} &= \operatorname{argmin}_{z\in M}\left\{\frac{1}{2}d(z, z_k)^2 + \frac{k+1}{2L}\frac{1}{m}\sum_{i=1}^{m} B_{y_k,\nabla f_i(y_k)}(z)\right\}
\end{aligned} \tag{12}
$$

In particular, when $m = 1$, the updating rule for z_k simplifies to

$$
z_{k+1} = \exp_{z_k}\left(-\frac{k+1}{2L}\nabla B_{y_k,\nabla f(y_k)}(z_k)\right).
$$

Theorem 18 *The algorithm* (12) *satisfies*

$$
f(x_N) - f(x^*) \leq \frac{2L}{N^2}d(x_0, x^*)^2.
$$

Smooth horo-convex interpolation

Claim 1: Consider the data $D = \{(x_i, g_i, f_i)\}_{i=1,\ldots,n}$, and define the data

$$\tilde{D} = \left\{(\tilde{x}_i, \tilde{g}_i, \tilde{f}_i) := \big(\exp_{x_i}(-\frac{g_i}{L}), \Gamma_{x_i, -\frac{g_i}{L}} g_i, f_i - \frac{\|g_i\|^2}{2L}\big)\right\}_{i=1,\ldots,n}.$$

If $\tilde{D}$ is interpolated by a g-convex function $\tilde{f}$ then D is interpolated by the g-convex function $f(x) = \min_{y\in M}\{\tilde{f}(y) + \frac{L}{2}d(x,y)^2\}$ [a Moreau envelope!]. f is L-B-smooth. If $\tilde{f}$ is B-convex then f is B-convex.