Incremental minimization in spaces of nonpositive curvature

Ariel Goodwin

Cornell University

Joint work with A.S. Lewis, G. López-Acedo, and A. Nicolae

ICCOPT, July 2025

Questions

Question 1:
What is the "right" class of nonlinear spaces for developing theory and algorithms for convex optimization?

Strong candidate: Hadamard spaces

Question 2:
Do classical subgradient methods extend to Hadamard spaces?

Answer: The rest of this talk

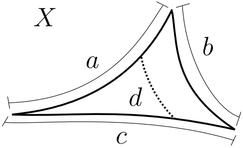
Hadamard Space

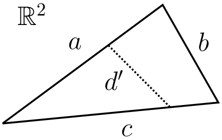
$\gamma\colon [a,b] \to X$ is a **geodesic** if $d(\gamma(t),\gamma(t')) = |t - t'|$ $(t, t' \in [a,b])$

A **complete** **geodesic space** (X,d) that is **CAT(0)**, meaning $d(\cdot, y)^2$ is strongly convex for all $y \in X$, i.e.

$$t \mapsto d(\gamma(t), y)^2 - t^2 \text{ is convex for any geodesic } \gamma$$

- $f\colon X \to \mathbb{R}$ is **convex** $\equiv$ $f \circ \gamma$ is convex for any geodesic γ
- Basic examples: Euclidean and Hilbert spaces
- Triangles in CAT(0) spaces are (at least as) **thin** as in Euclidean space

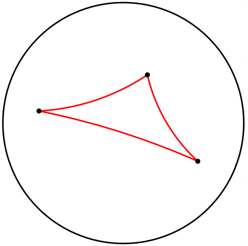




Hadamard manifolds

Hyperbolic space: open ball with metric

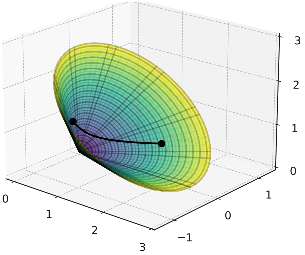
$$d(x, y) = \cosh^{-1}\left(1 + \frac{2\|x - y\|^2}{(1 - \|x\|^2)(1 - \|y\|^2)}\right)$$



Applications: Hierarchical data embedding (De Sa...'18), large-margin classification (Cho...'18)

Positive definite cone with metric

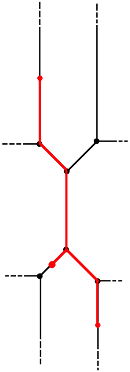
$$d(X, Y) = \|\log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}})\|_{\mathrm{Frob}}$$

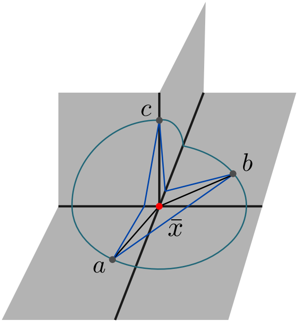


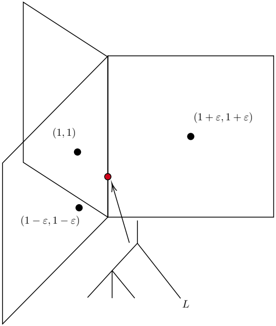
Applications: Covariance estimation (Auderset '05), matrix means (Bhatia...'12)

CAT(0) Cubical Complexes

A **complex** X of Euclidean cubes and their faces, each pair of cubes sharing at most one face, is **CAT(0)** $\iff$ simply connected and **Gromov's link condition** holds







Applications: Facility location (Hansen. . . '87), phylogenetics (Billera-Holmes-Vogtmann '01), robot configurations (Ardila-Mantilla '20)

Manifolds and beyond

In Euclidean space $\mathbb{E}$, subgradients are dual objects: $v \in \partial f(x)$ means

$$f(y) \geq f(x) + \langle v, y - x \rangle \text{ for all } y \in \mathbb{E}$$

Hadamard **manifolds**: subgradient methods (Zhang-Sra '16) but ...

- Complexity analysis is **curvature-dependent**
- **Acceleration** is obstructed (Criscitiello-Boumal '22)
- Non-manifolds exist — can we define subgradients in a primal way?

Main tool: **tractable geodesics**

Model problems

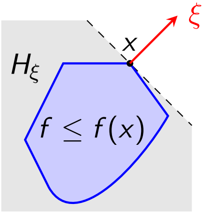
$$\min_{x \in X} \sum_{i=1}^{k} d^p(x, a_i) \quad (p\text{-}\mathbf{mean})$$

Proximal splitting (Bačák '14): $n \in \mathbb{N}$, $x \leftarrow \mathrm{prox}_{t_n f_i}(x)$ $(i = 1, \ldots, k)$

Simple for $p \in \{1, 2\}$ but prox typically intractable and can't solve

$$\min_{x \in X} \max_{i=1,\ldots,k} \{d(x, a_i)\} \quad (\mathbf{circumcenter})$$

Subgradient-like method (Lewis... '24): $n \in \mathbb{N}$, move t_n in **direction** $-\xi_n$



"Supporting" direction creates "halfspace"

Curvature-independent

Based on level sets ($\simeq$ quasiconvexity)

Cannot handle sums (e.g. p-means)

$\min_{n=1,\ldots,m} f(x_n) - \min f = O(m^{-1/2})$

Unified algorithm?

Busemann functions

Given a ray $r\colon \mathbb{R}_+ \to X$, the **Busemann function** (1955) is $b_r\colon X \to \mathbb{R}$,

$$b_r(x) := \lim_{t\to\infty} (d(x, r(t)) - t)$$

Busemann functions are **convex**

Examples: $(\xi \in \mathbb{S}^{n-1})$

$\mathbb{R}^n, r(t) = x - t\xi$:

$$b_r(y) = \langle \xi, y - x \rangle$$

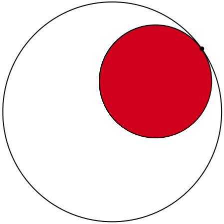
$\mathbb{H}^n, r(t) = \tanh(t/2)\xi$:

$$b_r(y) = -\log\left(\frac{1 - \|y\|^2}{\|\xi - y\|^2}\right)$$

Sublevel sets of b_r are **horoballs**:

$$H = \{x \in X \mid b_r(x) \le 0\}$$

$\mathbb{R}^n$: **Halfspaces**



$\mathbb{H}^n$: Euclidean balls tangent to $\mathbb{S}^{n-1}$

Busemann subgradients

Characterization of (classical) subgradients: $v \in \partial f(x)$ iff

$$x \text{ minimizes } f(\cdot) - \langle v, \cdot \rangle$$

i.e. x minimizes f with an affine perturbation

Affine functions $\rightsquigarrow$ Busemann functions in Hadamard space, so if

- r is a ray issuing from x (**direction**)
- $s \in \mathbb{R}_+$ (**magnitude** or speed)

we say (r, s) is a **Busemann subgradient** of $f : X \to \mathbb{R}$ at x if

$$x \text{ minimizes } f - sb_r$$

- **Busemann subdifferentiability** $\implies$ convexity
- Stronger than convexity in general: $x \mapsto \mathrm{dist}_C(x)$
- $\{f \leq \alpha\}$ is **supported** by a horoball at each boundary point

Examples

- Euclidean subgradient $0 \neq v \in \partial f(x)$ induces subgradient

$$(r, \|v\|) \text{ where } r(t) = x - t\frac{v}{\|v\|}$$

- **(Distance to point)** $f(x) = d^p(x, a)$ $(p \geq 1)$ has subgradient at $x \in X$ given by

$$(r, pd(x, a)^{p-1}) \text{ where } r(0) = x, r(d(x, a)) = a$$

- **(Busemann functions)** $f(x) = b_r(x)$ has subgradient $(r_x, 1)$ at $x \in X$, where

$$r_x(0) = x, \text{“}r_x(\infty) = r(\infty)\text{”} \quad \text{(exists unique such } r_x\text{)}$$

Further examples: $x \mapsto \mathrm{dist}_B(x)$ where $B \subseteq X$ is a (horo)ball

Other developments

Lewis... '23: **Subgradients** for convex functions on CAT(κ) spaces

- Much easier to satisfy **(local vs. global)**: (r, s) is a **subgradient** if

$$f(y) \geq f(x) + sd(x,y)\cos \underbrace{\angle(\gamma_{xy}, r)}_{\text{local geometry}} \quad \text{for all } y \in X$$

- Busemann subgradient induces **subgradient** in "opposite" direction:

$$\underbrace{b_r(y)}_{\text{global geometry}} \geq d(x,y)\cos\angle(\gamma_{xy}, r_-)$$

Criscitiello-Kim '25: **H-convex** optimization on Hadamard manifolds

- Similar idea to Busemann subdifferentiability, only for manifolds
- Achieves **acceleration**
- To minimize $\sum_i f_i$, requires 2-mean oracle

Subgradient calculus

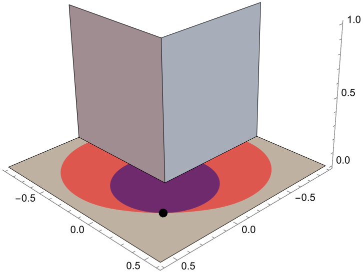
Basic **calculus**:

Goal: $\min_{x \in C} \sum_{i=1}^{k} f_i(x)$

- Maximum of $f_1, \dots, f_k$
- Chain rule: $\phi \circ f$ with $\phi \colon \mathbb{R} \to \mathbb{R}$ convex and increasing

For $f(x) = \sum_{i=1}^k f_i(x)$ in $\mathbb{R}^n$ we have **sum rule**: $\partial f(x) = \sum_{i=1}^k \partial f_i(x)$

No analogue in Hadamard space: $X =$ union of 5 quadrants in $\mathbb{R}^3$



$$f(x) = d^2(x, e_1) + d^2(x, e_2)$$

$\{\mathbf{f} \leq \mathbf{f}(\bar{\mathbf{x}})\}$ not supported by **horoball**

$\implies$ f cannot be Busemann subdiff.

A splitting algorithm

Bertsekas-Nedic '01 propose a **splitting** algorithm for $\min_C \sum_{i=1}^k f_i$:

For $j = 0, 1, 2, \ldots$ **do**
  For $i = 0, 1, \ldots, k-1$ **do**

$$x^{i+1} \leftarrow \mathrm{proj}_C(x^i - t_j v^i) \;\text{ where }\; v^i \in \partial f_{i+1}(x^i)$$

  $x^0 \leftarrow x^k$

$x^i - t_j v^i = r^i(\|v^i\| t_j)$ where $r^i(t) = x^i - t \frac{v^i}{\|v^i\|}$

Generalizing, with Busemann subgradient (r^i, s_i) for f_{i+1} at x^i:

$$x^{i+1} \leftarrow \mathrm{proj}_C(r^i(t_j s_i))$$

Complexity results

Theorem: [G.-Lewis-López-Nicolae, '24]
For Lipschitz and Busemann subdiff. $f_i \colon C \subseteq X \to \mathbb{R}$, C closed convex bounded, if $\min\limits_C f = \sum_{i=1}^k f_i$ is attained then the incremental subgradient method with $t_n \sim \frac{1}{\sqrt{n+1}}$ achieves

$$\min_{j=1,\ldots,n} f(x^j) - \min_C f = O\left(\frac{1}{\sqrt{n}}\right)$$

Aside: If f is 1-strongly convex, the same method with $t_n \sim \frac{1}{n+1}$ achieves

$$d(x^n, \operatorname*{argmin}_C f) = \tilde{O}\left(n^{-1/2}\right)$$

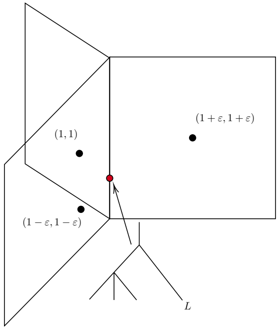
Application: Medians in tree space

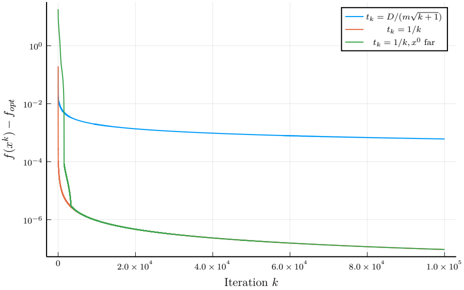
Billera-Holmes-Vogtmann '01: **Tree space** $\mathcal{T}_n$ whose points are binary trees (on n leaves) with non-negative internal edge lengths

$\mathcal{T}_n$ is a CAT(0) cubical complex

Polynomial time geodesics (Owen-Provan '11)

Medians (1-means) of $A = \{a_1, \ldots, a_k\}$ minimize $f(x) = \sum_{i=1}^{k} d(x, a_i)$





Summary

- **Hadamard spaces** subsume a variety of nonlinear spaces, ripe with **applications and optimization problems**
- **Incremental subgradient** methods with **curvature-independent** complexity succeed in Hadamard space for broad class of functions
- A subgradient splitting algorithm for optimization on nonpositively curved metric spaces (G.-Lewis-López-Nicolae) `arXiv:2412.06730`