Subgradient Methods in Geodesic Spaces

Ariel Goodwin [1] Adrian S. Lewis [1] Genaro López-Acedo [2] Adriana Nicolae [3]

[1]Cornell University [2]University of Seville [3]Babeş-Bolyai University

Cornell University

Minimizing a Sum

$$\min \left\{ f(x) = \sum_{i=1}^{m} f_i(x) \mid x \in C \right\} \qquad \text{(SUM)}$$

$f_i \colon C \subseteq X \to \mathbb{R}$ are functions, (X,d) is a **complete geodesic metric space**

Existing algorithms rely on proximal steps (difficult)

Example: The **Weber problem** (optimal facility location) is

$$\min_{x \in X} \sum_{i=1}^{m} w_i d(x, a_i)^p$$

Hadamard Spaces

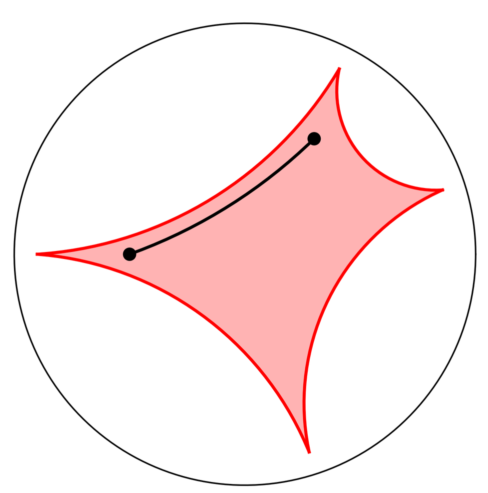
Geodesics are paths γ in X with $d(\gamma(t), \gamma(t')) = |t - t'|$

X has **curvature** ≤ 0 if $t \mapsto d(\gamma(t), y)^2 - t^2$ is convex for each $y \in X$, geodesic γ

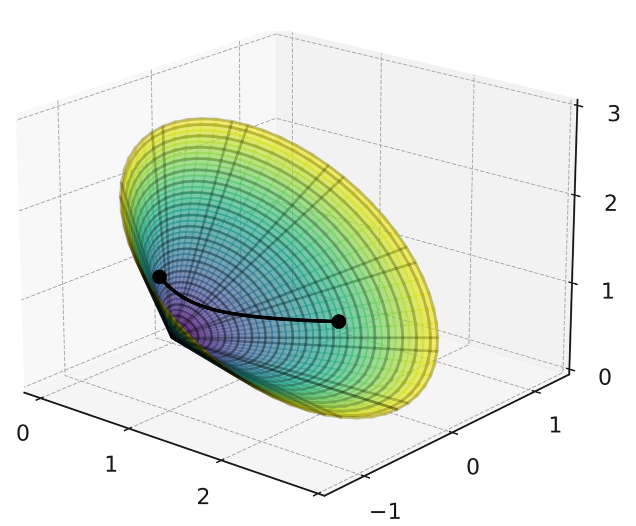
Hadamard Space: Complete geodesic space of curvature ≤ 0

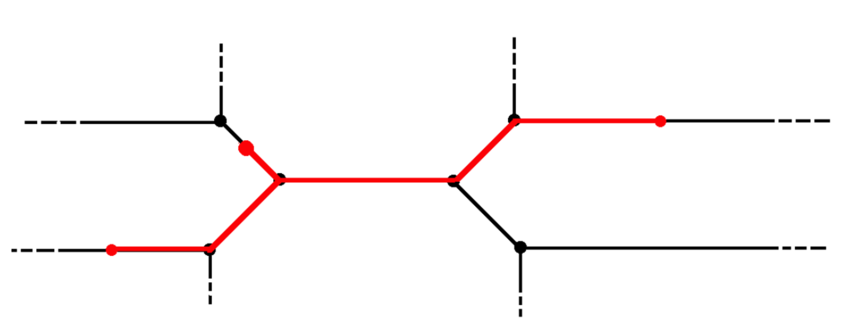
Includes Euclidean and Hilbert space, but also:

Hyperbolic Space $\mathbb{H}^n$

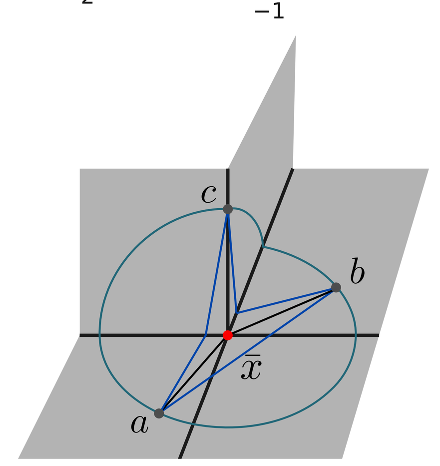


Positive Definite Cone S^n_{++}





Metric Trees

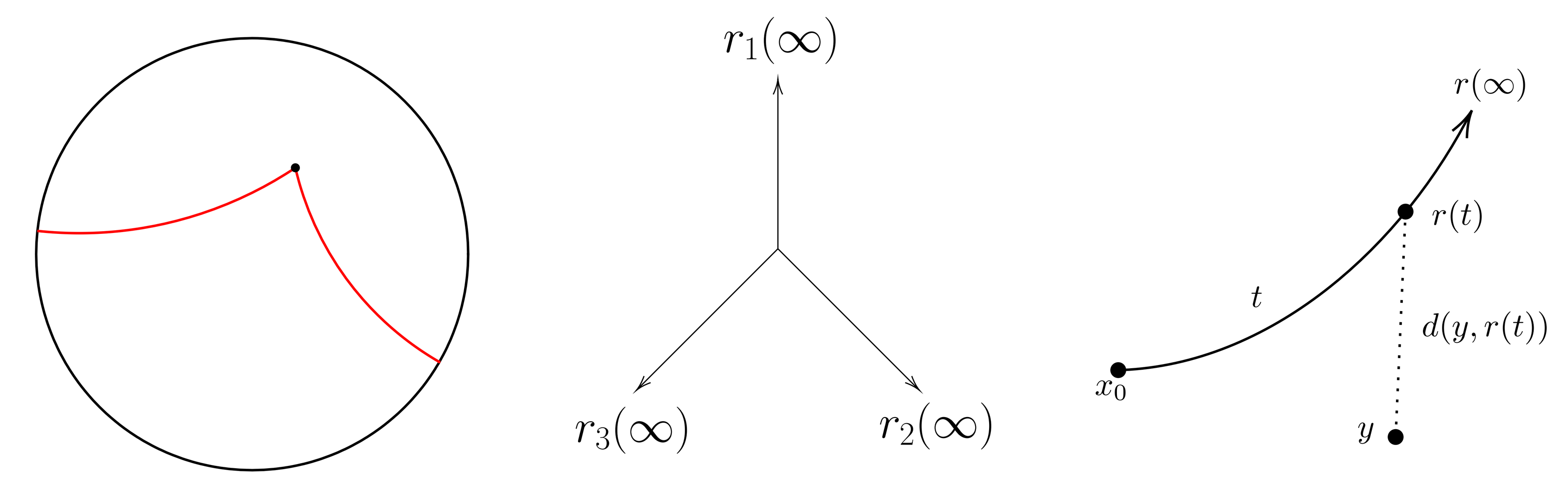


CAT(0) Cubical Complexes

Applications: Hierarchical classification, matrix means, covariance estimation, operator scaling, phylogenetics, facility location, robotic motion...

Busemann Convexity

Geodesic rays $r \colon \mathbb{R}_+ \to X$ induce a notion of **direction** $r(\infty)$:



For n-manifolds the **space of directions** X^∞ is $\mathbb{S}^{n-1}$, but for the **tripod** it is discrete

To a direction $\xi \in X^\infty$ we associate the **Busemann function** $b_\xi \colon X \to \mathbb{R}$:

$$b_\xi(y) := \lim_{t \to \infty} (d(y, r(t)) - t) \qquad (r(0) = x_0, r(\infty) = \xi)$$

(i) $\mathbb{R}^n$: $b_\xi(y) = \langle y, -\xi \rangle$

(ii) $\mathbb{H}^n$: $b_\xi(y) = -\log \left(\frac{1 - \|y\|^2}{\|\xi - y\|^2} \right)$

(iii) Tripod: $b_{\xi_i}((y,j)) = (-1)^{\delta_{ij}} y$

Affine functions

$\Downarrow$

Busemann functions

Definition: $f \colon C \to \mathbb{R}$ has a **Busemann subgradient** $(\xi, s) \in X^\infty \times \mathbb{R}_+$ at x if

$$f(y) - s b_\xi(y) \geq f(x) - s b_\xi(x) \quad \text{for all } y \in C$$

Then f is **Busemann convex** if it has a Busemann subgradient at each $x \in C$

- Stronger than geodesic convexity in general
- Simple calculus: max rule, chain rule, but no sum rule...

Examples: Busemann functions, distances to points/balls/**horoballs** $b_\xi^{-1}((-\infty, 0])$

A Stochastic Subgradient Algorithm

Popular algorithm for solving (SUM) in $X = \mathbb{R}^n$

For $k = 0, 1, 2, \ldots$ do

$$x^{k+1} = \mathrm{proj}_C(x^k - t_k v^k) \text{ where } v^k \in \partial f_{i(k)}(x^k),\ i(k) \sim \text{Uniform}\{1, \ldots, m\}$$

Generalizing, use Busemann subgradient (ξ^k, s_k) for $f_{i(k)}$ at x^k to update iterate:

$$x^{k+1} = \mathrm{proj}_C(r(t_k s_k)) \text{ where } r(0) = x^k, r(\infty) = \xi^k$$

A derandomized incremental version is also available

Computing Medians

A **median** of $A = \{a_1, \ldots, a_m\} \subseteq X$ is a solution to (SUM) with $f_i = w_i d(\cdot, a_i)$

f_i has subgradient $(r_i(\infty), w_i)$ at $x \neq a_i$, where $r_i(d(x, a_i)) = a_i$

Stochastic subgradient step: $x^{k+1} = \mathrm{proj}_C(r_{i(k)}(t_k w_{i(k)}))$

Interpretation: Move toward $a_{i(k)}$ proportionally to $w_{i(k)}$

Theorem (Median Complexity)

Suppose $w_1 = \max_i w_i$. If $C = B(a_1, f(x^0)/w_1)$, $t_k = 2/(w_1 m \sqrt{k+1})$ then

$$\min_{i=1,\ldots,k} f(x^i) - \inf_X f = O\left(\frac{1}{\sqrt{k}}\right)$$

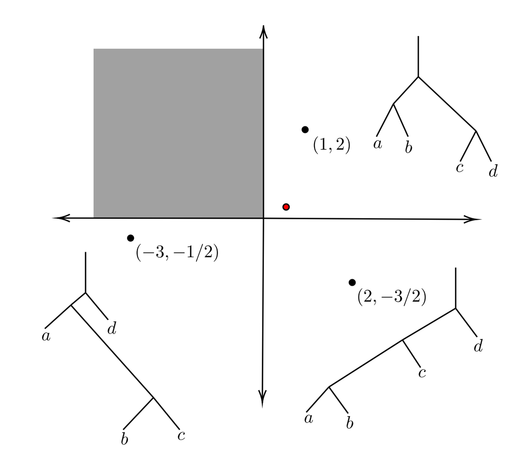
Application: Computing the Median of Phylogenetic Trees

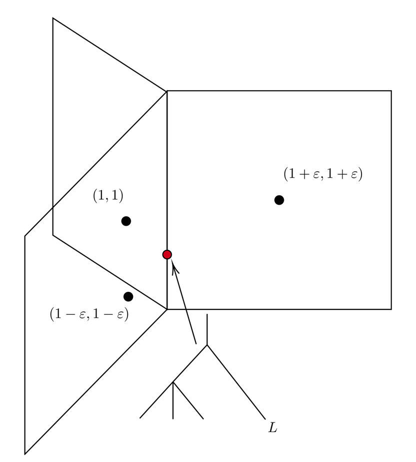
Given a dataset of phylogenetic trees, how can we find a representative point, like a mean or median?

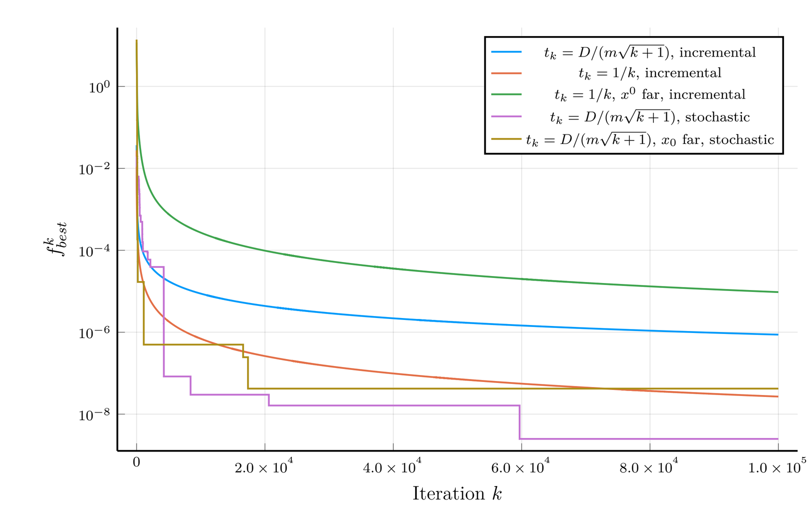
BHV tree space $\mathcal{T}_n$ is a Hadamard space whose points are binary trees

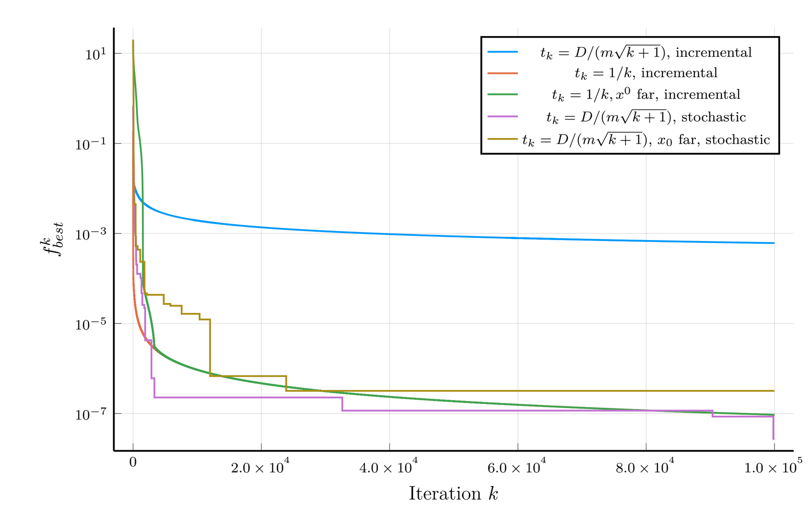
Geodesics in $\mathcal{T}_n$ are computable in polynomial time (Owen-Provan '11)

On two examples below we compute the median of three trees in $\mathcal{T}_4$:









References and Acknowledgements:

[1] M. Bačák. *Convex Analysis and Optimization in Hadamard Spaces*. De Gruyter, Berlin, 2014.

[2] M.R. Bridson and A. Haefliger. *Metric Spaces of Non-Positive Curvature*. Springer-Verlag Berlin, 1999.

[3] A. Goodwin, A.S. Lewis, G. López-Acedo, and A. Nicolae. A subgradient splitting algorithm for optimization on nonpositively curved metric spaces. arXiv:2412.06730, 2024.

Ariel Goodwin is supported by the NSERC postgraduate fellowship PGSD-587671-2024