Incremental Minimization in Spaces of Nonpositive Curvature

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

The Problem: Minimizing a Sum of Functions

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

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

Existing algorithms converge at unknown rate and rely on proximal steps (difficult)

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

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

The special case $p_i = p \geq 1$ for $1 \leq i \leq m$ is the p-**mean problem**

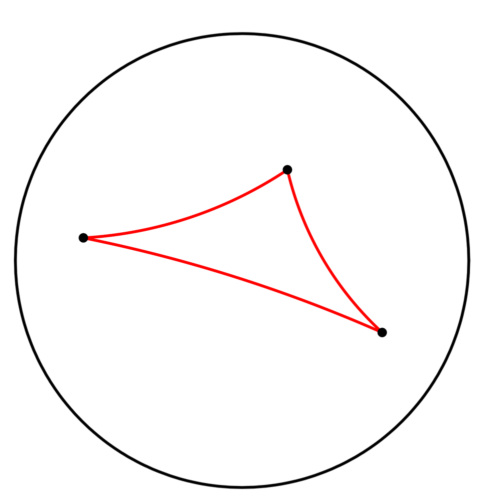
Hadamard Spaces

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

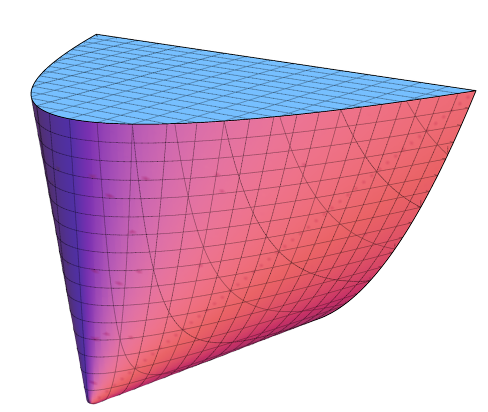
X has **curvature** ≤ 0 **(CAT(0))** if $t \mapsto d(\gamma(t), y)^2 - t^2$ is convex $\forall y \in X$, γ geodesic

Hadamard Space: Complete geodesic space of curvature ≤ 0

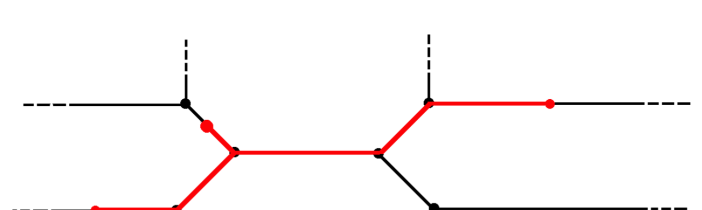
Includes Euclidean and Hilbert space (classical optimization), but also:



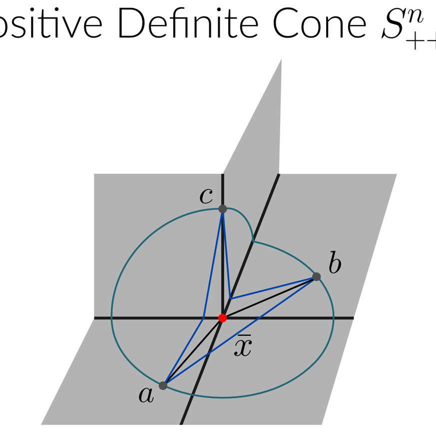
Hyperbolic Space $\mathbb{H}^n$



Positive Definite Cone S^n_{++}



Metric Trees



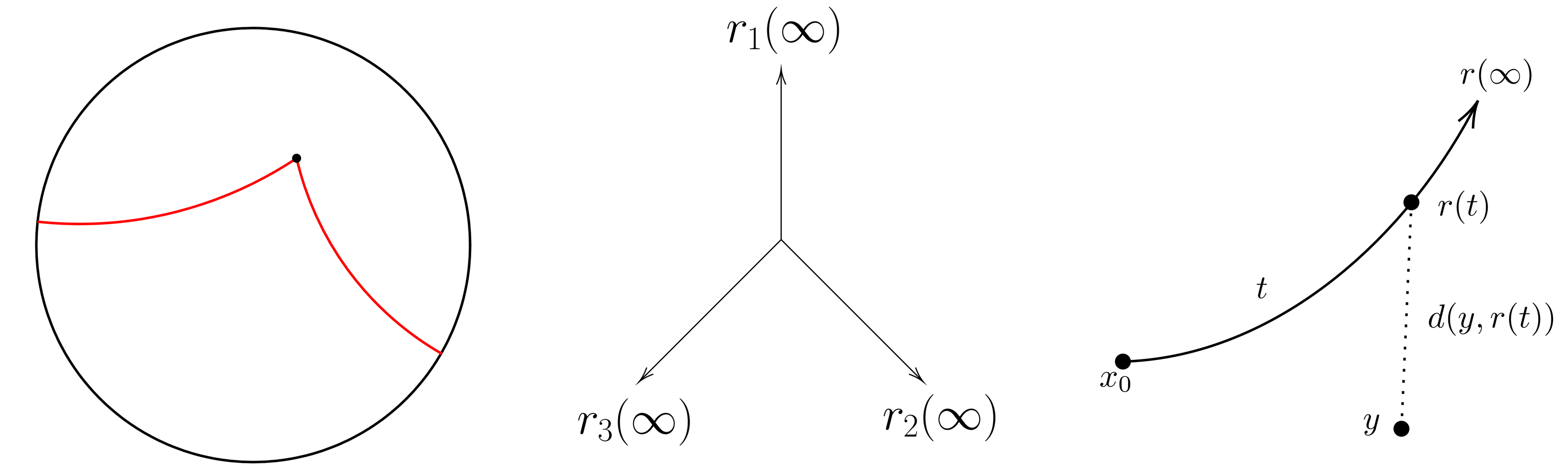
CAT(0) Cubical Complexes

Applications modeled in such spaces include hierarchical classification, matrix means, phylogenetics, facility location, and robotic motion

Any two points in a Hadamard space are joined by a unique geodesic

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

The Euclidean case shows Busemann functions generalize affine functions

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

$$f(y) - s b_\xi(y) \geq f(x) - s b_\xi(x) \quad \forall y \in C$$

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

- Stronger than geodesic convexity in general (equivalent in $\mathbb{R}^n$)
- Simple calculus: max rule, chain rule, but no sum rule... (splitting is key)

Examples: Busemann functions, distances to points/balls/**horoballs** (sublevel sets of Busemann functions)

An Incremental Subgradient Algorithm

Simple algorithm for solving (SUM) in $X = \mathbb{R}^n$ due to Bertsekas and Nedić (2001):

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

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

$x^{k+1} = x^{k,m}$

Generalizing, use Busemann subgradient $(\xi^{k,i}, s_{k,i})$ for f_{i+1} at $x^{k,i}$ to update iterate:

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

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

The resulting incremental subgradient step is $x^{k,i+1} = \mathrm{proj}_C(r_i(t_k w_i))$

At step i in each internal loop, the iterate moves towards a_i proportionally to w_i

Theorem (Median Complexity)

If $C = B(x^0, f(x^0)/w_1)$, $t_k = 2/(w_1 m\sqrt{k+1})$ then f has a minimizer in C and

$$\min_{i=1,\ldots,k} f(x^i) - f_{\text{opt}} = O(1/\sqrt{k})$$

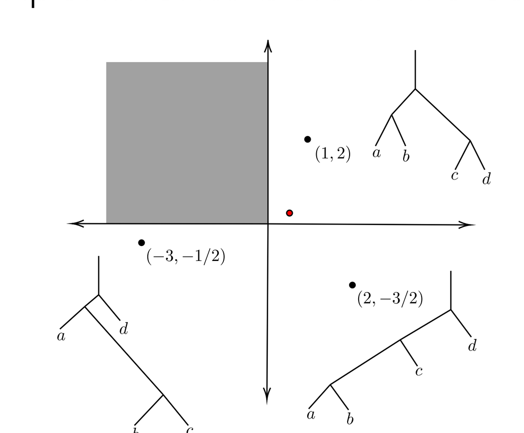
Application: Computing the Median of Phylogenetic Trees

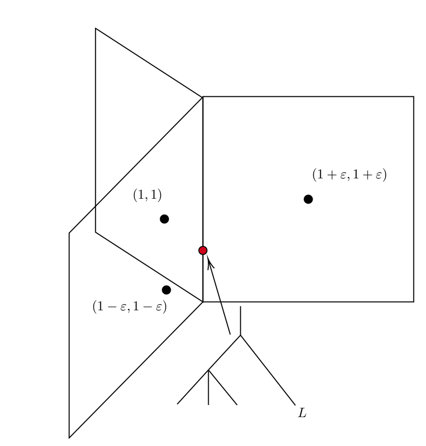
Several candidate phylogenetic trees may be generated to model an evolutionary history; means and medians condense this data into one representative tree

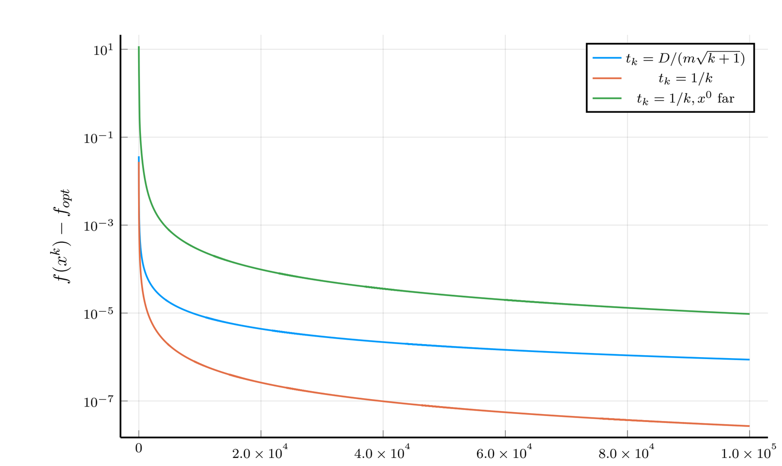
The **BHV tree space** $\mathcal{T}_n$ models the set of all binary trees on n labelled leaves, each with $n-2$ nonnegative internal edge lengths (viewed as a point in $[0,\infty)^{n-2}$)

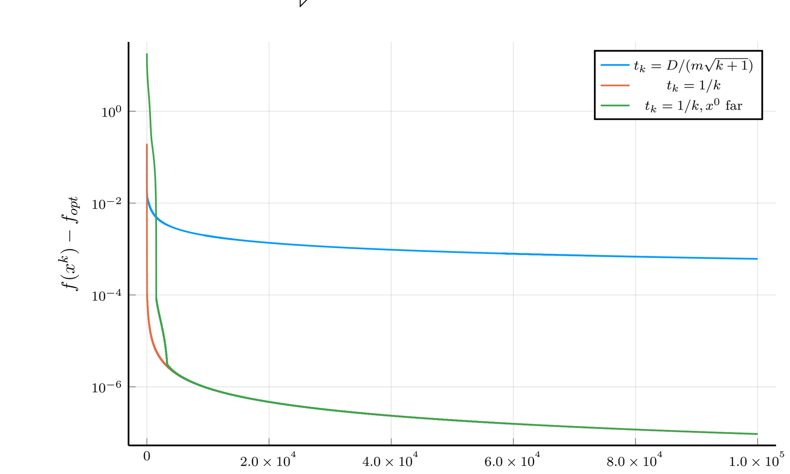
Geodesics in $\mathcal{T}_n$ are computable in polynomial time (Owen and Provan, 2011)

In both experiments 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.S. Lewis, G. López-Acedo, and A. Nicolae. Horoballs and the subgradient method. arXiv:2403.15749, 2024.

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