arXiv:2403.15749v2 [math.OC] 2 Apr 2024

Horoballs and the subgradient method

A.S. Lewis * G. López-Acedo † A. Nicolae ‡

April 4, 2024

Abstract

To explore convex optimization on Hadamard spaces, we consider an iteration in the style of a subgradient algorithm. Traditionally, such methods assume that the underlying spaces are manifolds and that the objectives are geodesically convex: the methods are described using tangent spaces and exponential maps. By contrast, our iteration applies in a general Hadamard space, is framed in the underlying space itself, and relies instead on horospherical convexity of the objective level sets. For this restricted class of objectives, we prove a complexity result of the usual form. Notably, the complexity does not depend on a lower bound on the space curvature. We illustrate our subgradient algorithm on the minimal enclosing ball problem in Hadamard spaces.

Key words: convex, subgradient method, horoball, Hadamard space, complexity, minimal enclosing ball

AMS Subject Classification: 90C48, 65Y20, 49M29

1 Introduction

We consider optimization problems posed in a geodesic metric space $\mathcal{M}$. Specifically, we consider first-order algorithms for minimizing an objective function $f\colon \mathcal{M} \to \mathbf{R}$ over a set $X \subset \mathcal{M}$. In non-Euclidean spaces $\mathcal{M}$, such algorithms are more subtle than their Euclidean counterparts. In particular, the literature on subgradient

*ORIE, Cornell University, Ithaca, NY. people.orie.cornell.edu/aslewis
Research supported in part by National Science Foundation Grant DMS-2006990.
†Department of Mathematical Analysis – IMUS, University of Seville, 41012 Seville, Spain glopez@us.es

‡Department of Mathematics, Babeş-Bolyai University, 400084, Cluj-Napoca, Romania anicolae@math.ubbcluj.ro

methods in Riemannian manifolds $\mathcal{M}$ generally assumes that f is geodesically convex, and in contrast with the classical Euclidean case, frames algorithms using local linearization, relying on the tangent space to $\mathcal{M}$ at each iterate. For a thorough exposition, see [10].

This standard approach to subgradient methods on manifolds, originating with [15], thus has two related disadvantages: the technicalities inherent in local linearization, and the resulting intuitive and formal challenge in extensions to more general spaces. Optimization in general Hadamard spaces has fascinating applications, including phylogenetic trees [8] and robotics [1]. Subgradient methods are an inviting possibility in such settings: existing algorithms all rely on proximal techniques [4], which despite their elegant theory, are only rarely implementable.

Motivated both by conceptual simplicity and by generality, therefore, our current development explores an alternative approach to understanding subgradient methods. The iteration we propose makes no reference to tangent spaces or exponential maps, relying instead on a notion of convexity in Hadamard spaces distinct from geodesic convexity. This idea — horospherical convexity — has its roots in hyperbolic geometry [9,24].

We first diverge from the standard approach to Riemannian subgradient methods in considering objectives that are quasiconvex rather than convex: in other words, we consider objectives that may not be convex, but have convex lower level sets. In the Euclidean case, this generalization originated with [20]. The generalization aside, our principal aim here is to focus on the geometry of level sets. Our objectives are continuous, and typically coercive, so the level sets we consider are usually compact, with nonempty interior.

In Euclidean space, the fact that geodesic and horospherical convexity coincide is the central tool of convex analysis: the supporting hyperplane theorem. By contrast, in the case of hyperbolic space, for compact sets with nonempty interior, horospherical convexity entails a certain level of curvature in the boundary, and is a strictly stronger property than geodesic convexity. The forward implication follows from [18, Theorem 4.2], for example. Conversely, in the Poincaré disk, where geodesics are segments of circles intersecting the unit circle orthogonally and horospheres are circles tangent to the unit circle, the upper half-disk of radius $\frac{1}{2}$ is geodesically convex but not horospherically convex, not being contained in any horosphere passing through zero. Nonetheless, applications in Hadamard spaces often involve sets that are both geodesically and horospherically convex, such as intersections of metric balls. Several recent works have noted the potential applications of horospherical ideas for non-Euclidean convex optimization, both in theory [6,18] and practice [14]. However, the subgradient method that we present here appears new.

Complexity analysis for standard projected subgradient methods on Hadamard manifolds originated with the seminal work [26], which shows in particular that the mean excess objective value over n iterations behaves like $O(\frac{1}{\sqrt{n}})$, just as in the

Euclidean version. Interestingly, the hidden constant deteriorates as the curvature becomes more negative, as subsequent analysis [13, Theorem 8] shows that it must, at least for general geodesically convex objectives.

Our second and principal point of divergence from [26] and other standard non-Euclidean subgradient developments lies in our avoidance of any reference to tangent spaces and exponential maps. Rather than situating the algorithm in a manifold, our subgradient-style method makes sense in any Hadamard space. Our complexity analysis is geometrically appealing: like the algorithm itself, it makes no use of tangent spaces, and as we have noted, applies to objectives that may be nonconvex but have horospherically convex level sets. Strikingly, unlike [26], our complexity result and proof involves no lower curvature bound: the lower bound on complexity in [13, Theorem 8] does not apply, because we only consider objectives with horospherically convex level sets.

To summarize our contributions, we present a new horospherical approach to subgradient methods that we believe is geometrically illuminating. Our algorithm is framed in a general Hadamard space, and relies on no tangent constructions. The algorithm does not apply to every geodesically convex objective, but its range of application is still rich. Furthermore, our complexity result involves no lower curvature bound, so even applies in spaces with infinite negative curvature and bifurcating geodesics, such as the tree spaces and cubical complexes of [1, 8]. We apply our horospherical subgradient algorithm in particular to the minimal enclosing ball problem. In Euclidean space, and more generally in Hadamard manifolds, our algorithm amounts to the approach of [2, 12], but algorithms for the minimal enclosing ball problem in a general Hadamard space were not previously available. We present a small computational illustration in a CAT(0) complex.

2 Examples

In a Hadamard space $(\mathcal{M}, d)$, many interesting optimization problems involve simple combinations of distance functions. To ground our discussion, we consider two representative examples.

Example 2.1 (Circumcenters) Any nonempty finite set $A \subset \mathcal{M}$ has a unique *circumcenter* [11, Proposition II.2.7], which is the unique minimizer $\bar{x}$ of the function $f\colon \mathcal{M} \to \mathbf{R}$ defined by

$$f(x) \;=\; \max_{a \in A} d(x, a) \qquad (x \in \mathcal{M}).$$

The value $f(\bar{x})$ is called the *circumradius*. Other names for this optimization problem include the *smallest* or *minimal enclosing ball problem*, and the 1-*center problem*. For a survey, see [2].

Example 2.2 (Intersecting balls) Consider any nonempty finite set $A \subset \mathcal{M}$. Given a radius $\rho_a \geq 0$ for each point $a \in A$, the corresponding balls $B_{\rho_a}(a)$ have nonempty intersection if and only if the function $f\colon \mathcal{M} \to \mathbf{R}$ defined by

$$f(x) \;=\; \max_{a \in A}\{d(x,a) - \rho_a\} \qquad (x \in \mathcal{M}).$$

has nonpositive minimum value.

Both examples involve minimizing an objective function belonging to the following broad class.

Definition 2.3 A function $f\colon \mathcal{M} \to \mathbf{R}$ is a *distance envelope* if there exists a compact topological parameter space Ψ, and continuous functions

$$a\colon \Psi \to \mathcal{M}, \qquad \beta\colon \Psi \to \mathbf{R}_+, \qquad \gamma\colon \Psi \to \mathbf{R},$$

such that

$$f(x) \;=\; \max_{\psi \in \Psi}\{\beta(\psi) d\big(x, a(\psi)\big) + \gamma(\psi)\} \qquad (x \in \mathcal{M}). \tag{2.4}$$

The points $a(\psi)$ are the *centers*.

3 Horoballs

The subgradient-style method we develop, rather than using the local idea of the tangent space, instead relies on global constructions fundamental in CAT(0) geometry, namely horoballs. We consider a Hadamard space $(\mathcal{M}, d)$ with the *geodesic extension property*: in other words, for any geodesic segment $[x, y]$ in $\mathbf{X}$, there exists a *ray* $r\colon \mathbf{R}_+ \to \mathcal{M}$ (by which we always mean a unit-speed geodesic ray) such that $r(0) = x$ and $r(t) = y$ for some $t \geq 0$. Following standard terminology [11], to any ray r we associate the *Busemann function* $b_r\colon \mathcal{M} \to \mathbf{R}$ defined by

$$b_r(x) \;=\; \lim_{t \to \infty}(d\big(x, r(t)\big) - t) \qquad (x \in \mathcal{M}).$$

Busemann functions are 1-Lipschitz and convex, and their lower level sets are called *horoballs*. Following standard terminology in hyperbolic geometry [16, Definition 4.5], we call a closed set $F \subset \mathcal{M}$ *horospherically convex* if every boundary point x of F has a *supporting ray*, by which we mean a ray r satisfying $r(0) = x$ and

$$F \;\subset\; \{z \in \mathcal{M} : b_r(z) \leq 0\}.$$

In that case we describe the right-hand side as a *supporting horoball* for F at x.

Example 3.1 (The Euclidean case) Consider the space $\mathcal{M} = \mathbf{R}^n$ with Euclidean distance. Given any ray $r(t) = x - tu$ (for $t \geq 0$), for a point $x \in \mathcal{M}$ and a unit vector $u \in \mathbf{R}^n$, the corresponding Busemann function is given by

$$b_r(z) \;=\; \langle u, z - x \rangle \qquad (z \in \mathcal{M}).$$

Thus horoballs in $\mathbf{R}^n$ are halfspaces. The ray r supports a closed convex set $F \subset \mathbf{R}^n$ at x if and only if u lies in the normal cone $N_F(x)$, in which case the halfspace $\{z : \langle u, z - x \rangle \leq 0\}$ is a supporting horoball for F at x. In particular, all closed convex sets F are horospherically convex.

Definition 3.2 A function $f \colon \mathcal{M} \to \mathbf{R}$ is *horospherically polyhedral* if there exist rays $r_i \colon \mathbf{R}_+ \to \mathcal{M}$, for $i = 1, 2, \ldots, m$, such that

$$f(x) \;=\; \max_{i=1,\ldots,m} \{\beta_i b_{r_i}(x) + \gamma_i\} \qquad (x \in \mathcal{M}). \tag{3.3}$$

4 The oracles

The iteration that we propose relies on two computational resources, described next. The first is standard.

Assumption 4.1 (Projection oracle) *The feasible region $X \subset \mathcal{M}$ is nonempty, closed, bounded, and geodesically convex. For any input $x \in \mathcal{M}$, the projection oracle outputs the nearest point in X, denoted* $\mathrm{Proj}(x)$.

The nearest point always exists and is unique [11, Proposition II.2.4]. The simplest example of the projection oracle is when the feasible region X is just a metric ball $B_\rho(a)$ for some radius $\rho \geq 0$ and some center $a \in \mathcal{M}$. In that case, the projection $\mathrm{Proj}(x)$ equals x whenever $d(x, a) \leq \rho$, and otherwise is the unique point on the geodesic segment $[a, x]$ at a distance ρ from a.

The second assumption is less standard.

Assumption 4.2 (Support oracle) *The objective function $f \colon \mathcal{M} \to \mathbf{R}$ is Lipschitz, with horospherically convex lower level sets*

$$f_x \;=\; \{z \in \mathcal{M} : f(z) \leq f(x)\} \qquad (x \in \mathcal{M}).$$

For any input $x \in \mathcal{M}$ and any step length $\epsilon > 0$, the support oracle either responds that x minimizes f, or it outputs a point x_ϵ on a supporting ray for f_x at x such that $d(x, x_\epsilon) = \epsilon$.

We illustrate the support oracle with some examples. The first shows that, in the classical case of convex optimization in Euclidean space, the support oracle coincides with the standard version.

Example 4.3 (The Euclidean case) Consider a continuous convex function $f\colon \mathbf{R}^n \to \mathbf{R}$, and any point $x \in \mathcal{M}$ that does not minimize f. Then valid outputs of the support oracle are those points of the form

$$x_\epsilon = x - \frac{\epsilon}{|g|} g \qquad \text{for } g \in \partial f(x).$$

To see this, we recall Example 3.1, and use the standard relationship between the normal cone to the level set and the subdifferential, $N_{f_x}(x) = \mathbf{R}_+ \partial f(x)$.

Example 4.4 (Distance envelopes) In Definition 2.3, consider the distance envelope f given by equation (2.4). Given any point $x \in \mathcal{M}$, fix any parameter value ψ attaining this maximum. If $\beta(\psi) = 0$, then x minimizes f. On the other hand, if $\beta(\psi) > 0$, then any ray $r\colon \mathbf{R}_+ \to \mathcal{M}$ satisfying $r(0) = x$ and passing through the center $a(\psi)$ supports the level set f_x, because any point $z \in f_x$ satisfies

$$\beta(\psi) d\big(z, a(\psi)\big) + \gamma(\psi) \;\le\; f(z) \;\le\; f(x) \;=\; \beta(\psi) d\big(x, a(\psi)\big) + \gamma(\psi)$$

and hence

$$d\big(z, a(\psi)\big) \;\le\; d\big(x, a(\psi)\big),$$

so

$$b_r(z) \;=\; \inf_{t \ge 0} \{ d\big(z, r(t)\big) - t \} \;\le\; d\big(z, a(\psi)\big) - d\big(x, a(\psi)\big) \;\le\; 0.$$

The support oracle can therefore output the point $x_\epsilon = r(\epsilon)$. In summary, for a distance envelope, to implement the support oracle we simply move a distance ϵ along a ray towards a center attaining the envelope's value.

Example 4.5 (Horospherically polyhedral functions) In Definition 3.2, consider the distance envelope f given by equation (3.3). Given any point $x \in \mathcal{M}$, fix any index i attaining this maximum. If $\beta_i = 0$, then x minimizes f. On the other hand, suppose $\beta_i > 0$. There exists a unique ray $r\colon \mathbf{R}_+ \to \mathcal{M}$ asymptotic to r_i and satisfying $r(0) = x$, by [11, Proposition II.8.2]. We claim that this ray supports the level set f_x at x. Note that the corresponding Busemann function b_r differs from the Busemann function b_{r_i} only by a constant [11, Exercises II.8.23]. Any point $z \in f_x$ satisfies

$$\beta_i b_{r_i}(z) + \gamma_i \;\le\; f(z) \;\le\; f(x) \;=\; \beta_i b_{r_i}(x) + \gamma_i,$$

so $b_{r_i}(z) \le b_{r_i}(x)$, and hence $b_r(z) \le b_r(x) = 0$. The support oracle can therefore output the point $x_\epsilon = r(\epsilon)$.

Discussion

The availability of the support oracle is, of course, a strong assumption. As we discussed in the introduction, geodesic and horospherical convexity are distinct notions. Thus the level sets of a geodesically convex objective function f may not be

horospherically convex, and even if they are, traditional subgradients may not correspond to the requisite supporting horoballs. Using standard Riemannian notation, the traditional oracle returns a vector g in the tangent space $T_x(\mathcal{M})$ satisfying the inequality

$$f(z) \;\geq\; f(x) + \langle g, \mathrm{Exp}_x^{-1}(z)\rangle_x \qquad (z \in \mathcal{M}),$$

which ensures that the level set L_x is contained in the set

$$\{z \in \mathcal{M} : \langle g, \mathrm{Exp}_x^{-1}(z)\rangle_x \leq 0\}.$$

This set contains the point x, but it may not correspond to any supporting horoball.

On the other hand, the support oracle is appealing geometrically, since it avoids any reference to tangent spaces and exponential maps. Furthermore, it relies only on convexity properties of the level sets of the objective f, rather than on convexity of f itself. The applicability of the subgradient method to Euclidean quasiconvex minimization was developed by [20].

5 A subgradient-style method

In a Hadamard space $(\mathcal{M}, d)$, suppose that Assumptions 4.1 and 4.2 hold, and consider the optimization problem

$$\inf_{x \in X} f(x).$$

Algorithm 5.1 (Projected subgradient iteration)
Choose a fixed step length $\epsilon > 0$.
Repeatedly update the current iterate $x \in X$ as follows.

- Call the support oracle in Assumption 4.2 at the point x.
- If the support oracle recognizes that x attains $\min_{\mathcal{M}} f$, stop.
- Else, update $x \leftarrow x_\epsilon$.
- Call the projection oracle in Assumption 4.1 at the point x.
- Update $x \leftarrow \mathrm{Proj}(x)$.

Example 5.2 (Circumcenters and intersecting balls) We return to Examples 2.1 and 2.2. Given an arbitrary point $a_0 \in \mathcal{M}$, if the radius $\rho > 0$ is large enough, then both the circumcenter and intersecting problems are equivalent to minimizing their respective objectives f over the ball $X = B_\rho(a_0)$. Projecting onto the ball X is easy, so we can implement the subgradient iteration simply: we simply need to compute (and possibly extend) geodesics between iterates and points in A or a_0.

6 Complexity analysis

The projected subgradient iteration that we have described is applicable in any Hadamard space with the geodesic extension property. We make no use of tangent spaces, and the curvature of the space may be unbounded below. In addition to Hadamard manifolds, the result applies in spaces like CAT(0) cubical complexes, which are not manifolds, and have infinite negative curvature.

In addition to Assumptions 4.1 and 4.2, we collect up the remaining assumptions.

Assumption 6.1

- *The Hadamard space* $(\mathcal{M}, d)$ *has the geodesic extension property.*
- *The diameter of the feasible region* $X \subset \mathcal{M}$ *is bounded above by* $D > 0$.
- *The objective function* $f \colon \mathcal{M} \to \mathbf{R}$ *is globally Lipschitz, with constant* $L > 0$.
- *The optimal value* $f^* = \inf_X f$ *is attained at some point* $x^* \in X$.

Under all of our assumptions, we prove that the iterates $x^1, x^2, \ldots$ generated by the subgradient method with a suitable step length, ensures a complexity estimate of the usual kind:

$$\min_{i=1,\ldots,n} f(x_i) \;=\; \min_X f \;+\; O\Big(\frac{1}{\sqrt{n}}\Big).$$

The complexity estimate depends only on the constants L and D, and not on any lower bound on the curvature.

Theorem 6.2 *Consider the optimization problem* $\inf_X f$*, and suppose that Assumptions 4.1, 4.2, and 6.1 hold. For any integer* $n > 0$*, consider Algorithm 5.1 with step length* $\epsilon = \frac{D}{\sqrt{n}}$*. After* n *iterations, the average of the function values at the iterates exceeds* $\min_X f$ *by a quantity no larger than*

$$\frac{LD}{\sqrt{n}}.$$

Proof Beginning with any current iterate $x \in X$, one iteration of Algorithm 5.1, outputs a point x_ϵ on a supporting ray $r \colon \mathbf{R}_+ \to \mathcal{M}$ for f_x at x such that $d(x, x_\epsilon) = \epsilon$. We denote the corresponding Busemann function b_r simply by b. Since x^* minimizes f, we know $f(x^*) \le f(x)$, and hence $b(x^*) \le 0$. We have $x_\epsilon = r(\epsilon)$, and more generally, for $t \ge 0$, we denote the point $r(t)$ by x_t.

For any value $t \ge \epsilon$ we have

$$x_\epsilon \;=\; \Big(1 - \frac{\epsilon}{t}\Big)x + \frac{\epsilon}{t}x_t.$$

Since $\mathcal{M}$ is a Hadamard space, nonpositive curvature implies

$$d^2(x^*, x_\epsilon) \;\le\; \Big(1 - \frac{\epsilon}{t}\Big) d^2(x^*, x) + \frac{\epsilon}{t} d^2(x^*, x_t) - \frac{\epsilon}{t}\Big(1 - \frac{\epsilon}{t}\Big) d^2(x, x_t)$$

so

$$\begin{aligned} d^2(x^*, x_\epsilon) - \Big(1 - \frac{\epsilon}{t}\Big) d^2(x^*, x) - \epsilon^2 \;&\le\; \frac{\epsilon}{t}(d^2(x^*, x_t) - t^2) \\ &=\; \epsilon(d(x^*, x_t) - t)\Big(\frac{1}{t} d(x^*, x_t) + 1\Big). \end{aligned}$$

As $t \to +\infty$, we know $d(x^*, x_t) - t \to b(x^*)$ by definition, and hence $\frac{1}{t} d(x^*, x_t) \to 1$, from which we deduce

$$d^2(x^*, x_\epsilon) - d^2(x^*, x) - \epsilon^2 \;\le\; 2\epsilon b(x^*).$$

There exists a unique ray originating at x^* and asymptotic to the ray r. Extend this ray to a point $\tilde{x} \in \mathcal{M}$ satisfying $d(x^*, \tilde{x}) = 1 - b(x^*)$, giving a ray $\tilde{r} \colon \mathbf{R}_+ \to \mathcal{M}$ asymptotic to r satisfying $\tilde{r}(0) = \tilde{x}$ and $\tilde{r}\big(1 - b(x^*)\big) = x^*$. The corresponding Busemann function $\tilde{b}$ satisfies

$$\tilde{b}(x^*) \;=\; \tilde{b}(\tilde{r}\big(1 - b(x^*)\big)) \;=\; b(x^*) - 1,$$

and it differs from the Busemann function b only by a constant, so $\tilde{b} = b - 1$. For all $t < 1$ we have

$$b\big(\tilde{r}(t)\big) \;=\; \tilde{b}\big(\tilde{r}(t)\big) + 1 \;=\; 1 - t \;>\; 0.$$

Since the ray r supports the level set f_x at x, we know $b(z) \le 0$ for all points $z \in f_x$. Hence we have $\tilde{r}(t) \notin f_x$, or equivalently, $f\big(\tilde{r}(t)\big) > f(x)$. By continuity, the point $\hat{x} = \tilde{r}(1)$ satisfies $f(\hat{x}) \ge f(x)$. Since $d(x^*, \hat{x}) = -b(x^*)$, the Lipschitz condition ensures

$$f(x) - f(x^*) \;\le\; f(\hat{x}) - f(x^*) \;\le\; -Lb(x^*).$$

To summarize, we have proved

$$d^2(x_\epsilon, x^*) - d^2(x, x^*) \;\le\; \epsilon^2 - \frac{2\epsilon}{L}(f(x) - f^*).$$

Since $x^* \in X$ and the projection Proj_X is nonexpansive, we deduce

$$\frac{2}{L}(f(x) - f^*) \;\le\; \frac{1}{\epsilon}\big(d^2(x, x^*) - d^2(x_{\mathrm{new}}, x^*)\big) \;+\; \epsilon.$$

Now suppose we initialize the algorithm at a point $x^1 \in X$, and apply the update n times, generating the points $x^1, x^2, x^3, \ldots, x^n, x^{n+1} \in X$. Summing the corresponding inequalities shows

$$\frac{2}{L}\Big(\frac{1}{n}\sum_{i=1}^{n} f(x^i) - f^*\Big) \;\le\; \frac{D^2}{n\epsilon} + \epsilon.$$

We deduce

$$\frac{1}{n}\sum_{i=1}^{n} f(x^i) \;\leq\; f^* + \frac{LD}{\sqrt{n}},$$

as required. □

7 Computing circumcenters in Hadamard space

We conclude by returning to the circumcenter problem, Example 2.1). The problem has a long history dating back to Sylvester (1857) [25], and remains fundamental in computational geometry. A simple Euclidean algorithm, with complexity $O(\frac{1}{\sqrt{n}})$ appeared in [12], and can be interpreted as the subgradient method [21]. The Euclidean algorithm was extended to Riemannian manifolds in [2], and we shall see that a constant-step-length version in fact extends further to arbitrary Hadamard spaces, where it can be interpreted as the horospherical subgradient method, Algorithm 5.1.

This extension is motivated in part by an interesting class of optimization problems in Hadamard space that arises from averaging finite sets in the BHV space of phylogenetic trees introduced in [8]. CAT(0) cubical complexes, of which the BHV space is an example, are Hadamard spaces with curvature unbounded below: in particular, they are not manifolds. In addition to BHV space, CAT(0) cubical complexes have also proved useful models in robotics, as surveyed recently in [1]. Crucially from an algorithmic perspective, geodesics in CAT(0) cubical complexes are efficiently computable [17, 23].

Consider the problem of averaging a given nonempty finite set A in a Hadamard space $\mathcal{M}$. Standard notions of average include the *Fréchet mean* and *medians*, which are minimizers of the functions

$$\sum_{a\in A} d^2(x,a) \qquad \text{and} \qquad \sum_{a\in A} d(x,a) \qquad (x\in\mathcal{M})$$

respectively. Such points can be approximated via a sequence of geodesic computations in a splitting proximal point algorithm. The convergence of the algorithm is proved in [5], and illustrated on real data in BHV space in [3]. While [5] states no complexity result, the algorithm and its convergence proof are based on a Euclidean version [7], whose complexity, like that of Algorithm 5.1, is $O(\frac{1}{\sqrt{n}})$.

The horospherical subgradient method that we have described here opens an alternative possibility for averaging a finite set A. Instead of medians or the Fréchet mean, we might instead approximate the circumcenter, by minimizing the objective function $f\colon \mathcal{M}\to\mathbf{R}$ defined, as in Example 2.1:

$$f(x) \;=\; \max_{a\in A} d(x,a) \qquad (x\in\mathcal{M}). \tag{7.1}$$

As we shall see, implementing the horospherical subgradient method to minimize the objective function f is straightforward.

One curiosity suggesting the circumcenter as an alternative averaging technique in CAT(0) cubical complexes is the *stickiness* to which the Fréchet mean is prone: for a short exposition, see [22]. Consider for example the simplest BHV space, the *tripod* consisting of three *legs*: copies of $\mathbf{R}_+$ glued together at zero. Three points, one on each leg at distances $a, b, c > 0$ from zero, have mean at zero if and only if

$$a + b > c, \qquad b + c > a, \qquad c + a > b. \tag{7.2}$$

The point zero is *sticky* because the corresponding set of vectors (a, b, c) has nonempty interior. More precisely, we can identify each vector $(a, b, c) \in \mathbf{R}^3_{++}$ with a probability measure on the tripod supported on the three corresponding points, with equal weights: this set of measures inherits the topology of $\mathbf{R}^3$, and any measure in the resulting space that satisfies the inequalities (7.2) *sticks* to zero, in the terminology of [19]. The median in this example behaves even more simply: the unique median is zero for all values a, b, c. Notice, in contrast, that zero is a circumcenter if and only if the maximum component of the vector (a, b, c) is not unique. Since the set of such vectors has measure zero, the point zero is not sticky as a circumcenter.

To compute the circumcenter using the horospherical subgradient method, we choose an arbitrary point $\bar{a} \in A$, define the radius $\rho = f(\bar{a})$, and consider the feasible region $X = B_\rho(\bar{a})$. Notice $A \subset X$. Since all points $x \notin X$ satisfy $f(x) \geq d(x, \bar{a}) > \rho$, minimizing f over the whole space $\mathcal{M}$ is equivalent to minimizing f over X. The set X satisfies Assumption 4.1, and the projection oracle is simple:

$$\mathrm{Proj}(x) \;=\; \begin{cases} x & (x \in X) \\ \frac{d(x,a)}{\rho} x + (1 - \frac{d(x,a)}{\rho})\bar{a}) & (x \notin X). \end{cases}$$

As we shall see, in fact the subgradient algorithm only calls the projection oracle for inputs $x \in X$, so the projection plays no role.

The objective function f is 1-Lipschitz. By Example 4.4, its level sets are horospherically convex, and given any input point $x \in \mathcal{M}$ and step length $\epsilon > 0$, the support oracle chooses a point $a \in A$ attaining $\max_A d(x, \cdot) = f(x)$, and then outputs the point x_ϵ on the ray originating at x and passing through a, at a distance ϵ from x. Thus Assumption 4.2 holds. Furthermore, providing $\epsilon \leq f(x)$, the point x_ϵ is a convex combination of x and a.

We assume that the Hadamard space $\mathcal{M}$ has the geodesic extension property, as holds in particular for BHV space. We define the constant $D = 2\rho$. Assumption 6.1 then holds.

Now consider Algorithm 5.1 with step length $\epsilon = \frac{D}{\sqrt{n}}$ and number of iterations $n \geq 16$. Notice $\epsilon \leq \frac{\rho}{2}$. If any point $x \in \mathcal{M}$ satisfied $f(x) < \epsilon$, then we deduce the contradiction

$$d(\bar{a}, a) \;\leq\; d(\bar{a}, x) + d(x, a) \;<\; \epsilon + \epsilon \;\leq\; \rho \;=\; f(\bar{a}) \qquad \text{for all } a \in A.$$

We therefore have $f(x) \geq \epsilon$ for all $x \in \mathcal{M}$. Consequently, for any iterate x in the feasible region X, the output x_ϵ of the support oracle also lies in X. By induction, the projection oracle thus plays no role, as we claimed.

We can summarize the algorithm, in terms of the function (7.1), as follows. While the approach is similar to the Euclidean algorithm of [12] and its Riemannian extension [2], we do not know of any previous algorithm for circumcenters in arbitrary Hadamard spaces.

Algorithm 7.3 (Subgradient iteration for circumcenter)

```
input: nonempty finite A ⊂ M, number of iterations n ≥ 16
choose x ∈ A                                        % current center
x_best = x                                          % best center so far
ρ = 0
for a ∈ A \ {x} do
   ρ = max{ρ, d(a, x)}                              % radius of current ball enclosing A
end for
f_best = ρ                                          % best radius so far
ε = 2ρ/√n                                           % step length
for iteration = 1, 2, ..., n do
   γ = 0
   for a ∈ A do
      δ = d(a, x)
      if δ > γ then
         γ = δ                                      % radius of current enclosing ball...
         â = a                                      % ...and furthest point in A
      end if
   end for
   if γ < f_best then
      f_best = γ                                    % update best radius ...
      x_best = x                                    % ...and best center
   end if
   λ = ε/γ
   x = (1 − λ)x + λâ                                % move towards furthest point
end for
return x_best
```

If the set A has cardinality k, then the initial `for` loop requires $k-1$ distance calculations, and then each of the n iterations requires k more distance computations, one of which results in the calculation of a convex combination of two points. The diameter μ of the set A satisfies $\mu \geq \rho$, so the feasible region X has diameter no larger than 2μ. Theorem 6.2 now implies the following complexity result.

Corollary 7.4 (Complexity of circumcenters) *Consider any Hadamard space $\mathcal{M}$ with the geodesic extension property, and any subset $A \subset \mathcal{M}$ of cardinality $k > 0$ and diameter μ. Fix any integer $n \geq 16$. Then Algorithm 7.3 terminates after computing $k(n+1) - 1$ distances and n convex combinations in $\mathcal{M}$, and returns a point whose distance from every point in A exceeds the circumradius of A by no more than $\frac{2\mu}{\sqrt{n}}$.*

Example: circumcenter of three points in an orthant space

To illustrate Algorithm 5 on a polyhedral complex analogous to the tree spaces of [8], we consider the geodesic space consisting of five Euclidean quadrants glued together along their edges to form a cycle. This orthant space $\mathcal{M}$ is a well-known example of a Hadamard space: it can be modeled isometrically as the subset $\mathcal{M}'$ of the space $\mathbf{R}^3$ consisting of the union of the quadrants

$$\mathbf{R}_+ \times \mathbf{R}_- \times \{0\}, \quad \mathbf{R}_- \times \mathbf{R}_- \times \{0\}, \quad \mathbf{R}_- \times \mathbf{R}_+ \times \{0\}, \quad \{0\} \times \mathbf{R}_+ \times \mathbf{R}_+, \quad \mathbf{R}_+ \times \{0\} \times \mathbf{R}_+,$$

with the intrinsic metric induced by Euclidean distance [1, Figure 9].

Computing geodesics in this orthant space $\mathcal{M}$ is easy. Given any two points $x, y \in \mathcal{M}$, after a cyclic permutation of the quadrants, we can suppose that the corresponding points x' and y' in the model both lie in the plane $\mathbf{R} \times \mathbf{R} \times \{0\}$. In the model, the geodesic $[x', y']$ is just the Euclidean line segment $[x', y']$ in the plane $\mathbf{R} \times \mathbf{R} \times \{0\}$, unless that line segment intersects the open quadrant $\mathbf{R}_{++} \times \mathbf{R}_{++} \times \{0\}$, in which case the geodesic consists of the union of the two geodesics $[x', 0]$ and $[0, y']$.

In the model $\mathcal{M}'$, consider the set A consisting of the three points

$$(0, 0, \sqrt{5}) \quad , \quad (1, -2, 0) \quad , \quad (-2, 1, 0).$$

The circumcenter of A is the point $(0, 0, 0)$. To see this, consider the objective function $f(x) = \max_A d(x, \cdot)$ (for $x \in \mathcal{M}'$) whose unique minimizer is the circumcenter (p, q, r) of A. By symmetry, the point (q, p, r) also lies in $\mathcal{M}'$, and the value of f there is identical, so in fact $(q, p, r) = (p, q, r)$, and hence $p = q$. Consequently the circumcenter of A must lie on the set

$$\{(0, 0, t) : t \geq 0\} \ \cup \ \{(t, t, 0) : t \leq 0\}.$$

A quick calculation shows that f is minimized on this set at the point $(0, 0, 0)$.

The performance of Algorithm 7.3 is shown in Figure 1. The plot clearly shows the $O(\frac{1}{\sqrt{n}})$ behavior relative to the number of iterations n. For comparison, in this example the set A has diameter $\sqrt{10 + 4\sqrt{5}} < 5$, so an upper bound on the error in approximating the circumradius, by Corollary 7.4, is $\frac{10}{\sqrt{n}}$.

Acknowledgements

The implementation of the horospherical subgradient algorithm and the results in Figure 1 are due to Ariel Goodwin.

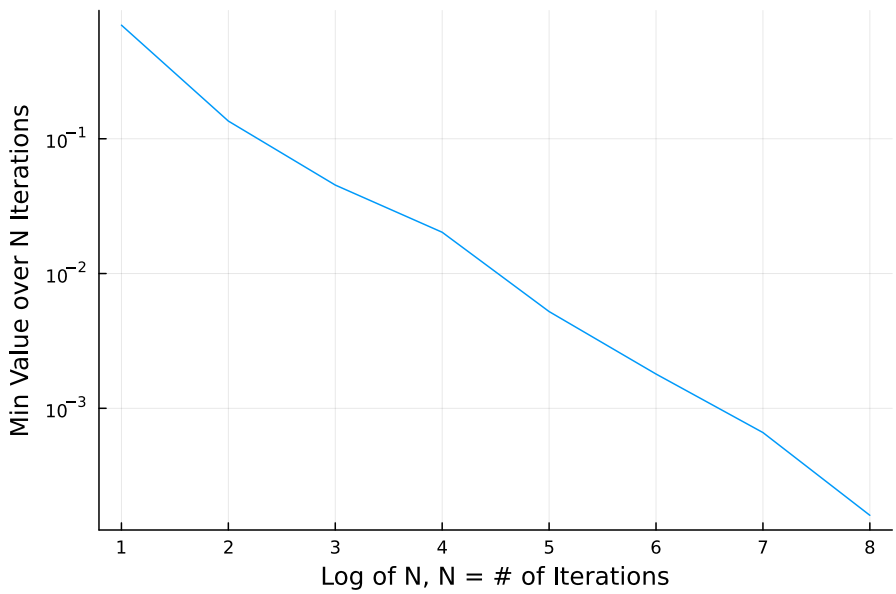

Figure 1: Algorithm 7.3 applied to the circumcenter and circumradius σ of three points in a Hadamard space of five orthants. For number of iterations $N = 10^1, 10^2, \ldots, 10^8$, the plot shows the minimum value of $\log_{10}(f(\cdot) - \sigma)$ over the N iterates.

References

[1] F. Ardila-Mantilla. CAT(0) geometry, robots, and society. *Notices Amer. Math. Soc.*, 67:977–987, 2020.

[2] M. Arnaudon and F. Nielsen. On approximating the Riemannian 1-center. *Computational Geometry*, 46:93–104, 2013.

[3] M. Bačák. The proximal point algorithm in metric spaces. *Isr. J. Math.*, 194:689–701, 2013.

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

[5] M. Bačák. Computing medians and means in Hadamard spaces. *SIAM Journal on Optimization*, 24:1542–1566, 2014.

[6] G.de C. Bento, J.C. Neto, and I.D. L. Melo. Fenchel conjugate via Busemann function on Hadamard manifolds. *Applied Mathematics and Optimization*, 88, 2023.

[7] D.P. Bertsekas. Incremental proximal methods for large scale convex optimization. *Math. Program.*, 129:163–195, 2011.

[8] L.J. Billera, S.P. Holmes, and K. Vogtmann. Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*, 27:733–767, 2001.

[9] A.A. Borisenko and V. Miquel. Total curvatures of convex hypersurfaces in hyperbolic space. *Illinois Journal of Mathematics*, 43:61–78, 1999.

[10] N. Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, Cambridge, 2023.

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

[12] M. Bădoiu and K.L. Clarkson. Smaller core-sets for balls. In *SODA Proceedings*, pages 801–802, Philadelphia, PA, USA, 2003. SIAM.

[13] C. Criscitiello and N. Boumal. Curvature and complexity: better lower bounds for geodesically convex optimization. *Proceedings of Machine Learning Research*, 195:1–45, 2023.

[14] Xiran Fan, Chun-Hao Yang, and B.C. Vemuri. Horospherical decision boundaries for large margin classification in hyperbolic space. In *NeurIPS*, 2023.

[15] O.P. Ferreira and P.R. Oliveira. Subgradient algorithm on Riemannian manifolds. *Journal of Optimization Theory and Applications*, 97:93–104, 1998.

[16] E. Gallego, G. Solanes, and E. Teufel. Linear combinations of hypersurfaces in hyperbolic space. *Monatsh, Math.*, 169:329–354, 2013.

[17] K. Hayashi. A polynomial time algorithm to compute geodesics in CAT(0) cubical complexes. *Discrete and Computational Geometry*, 65:636–654, 2021.

[18] H. Hirai. Convex analysis on Hadamard spaces and scaling problem. *Found. Comput. Math.*, 2023.

[19] S. Huckemann, J.C. Mattingly, E. Miller, and J. Nolen. Sticky central limit theorems at isolated hyperbolic planar singularities,. *Electronic Journal of Probability*, 20:1–34, 2015.

[20] K.C. Kiwiel. Convergence and efficiency of subgradient methods for quasiconvex minimization. *Math. Programming*, 90:1–25, 2001.

[21] A. Krivošija and A. Munteanu. Probabilistic smallest enclosing ball in high dimensions via subgradient sampling. In Gi. Barequet and Y. Wang, editors, *35th International Symposium on Computational Geometry*, pages 47:1–14. Dagstuhl Publishing, Germany, 2019.

[22] E. Miller. Fruit flies and moduli: interactions between biology and mathematics. *Notices Amer. Math. Soc.*, 62:1178–1184, 2015.

[23] M. Owen and S. Provan. A fast algorithm for computing geodesic distances in tree space. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8:2–13, 2011.

[24] A. Santaló and I. Yañez. Averages for polygons formed by random lines in Euclidean and hyperbolic planes. *Journal of Applied Probability*, 9:140–157, 1972.

[25] J.J. Sylvester. A question in the geometry of situation. *Quarterly Journal of Mathematics*, 1:79, 1857.

[26] Hongyi Zhang and S. Sra. First-order methods for geodesically convex optimization. *Journal of Machine Learning Research*, 49:1–22, 2016.