Discrete-time gradient flows for unbounded convex functions on Gromov hyperbolic spaces

Shin-ichi OHTA*

March 25, 2025

Abstract

In proper, geodesic Gromov hyperbolic spaces, we investigate discrete-time gradient flows via the proximal point algorithm for unbounded Lipschitz convex functions. Assuming that the target convex function has negative asymptotic slope along some ray (thus unbounded below), we first prove the uniqueness of such a negative direction in the boundary at infinity. Then, using a contraction estimate for the proximal (resolvent) operator established in our previous work, we show that the discrete-time gradient flow from an arbitrary initial point diverges to that unique direction of negative asymptotic slope. This is inspired by and generalizes results of Karlsson–Margulis and Hirai–Sakabe on nonpositively curved spaces and a result of Karlsson concerning semi-contractions on Gromov hyperbolic spaces.

1 Introduction

This article is a continuation of [30] concerning convex optimization on Gromov hyperbolic spaces. We shall study the *discrete-time gradient flow* for a convex function f on a metric space (X,d) built of the *proximal* (or *resolvent*) *operator*

$$\mathsf{J}^f_\tau(x) := \arg\min_{y\in X}\left\{f(y)+\frac{d^2(x,y)}{2\tau}\right\}, \tag{1.1}$$

where $\tau>0$ is the step size. Iterating J^f_τ is a well known scheme to construct a continuous-time gradient flow for f in the limit as $\tau\to 0$. Generalizations of the theory of gradient flows to convex functions on metric spaces have been making impressive progress since 1990s, including those on CAT(0)-spaces [1, 3, 17, 25] (see [12, 14, 16] for some applications to optimization theory), CAT(1)-spaces [32, 33], Alexandrov spaces and the Wasserstein spaces over them [23, 27, 32, 35, 36], and RCD(K,∞)-spaces [37]. These spaces are, however, all Riemannian in the sense that they exclude non-Riemannian Finsler manifolds (in particular, non-inner product normed spaces). In fact, despite great success for Riemannian spaces, much less is known for non-Riemannian spaces (even for normed spaces); we refer to [31, 34]

*Department of Mathematics, Osaka University, Osaka 560-0043, Japan (s.ohta@math.sci.osaka-u.ac.jp); RIKEN Center for Advanced Intelligence Project (AIP), 1-4-1 Nihonbashi, Tokyo 103-0027, Japan

for the failure of the contraction property. Motivated by this large gap and a fact that some non-Riemannian Finsler manifolds can be Gromov hyperbolic (Example 2.1(b)), we initiated investigation of convex optimization on Gromov hyperbolic spaces in [30] (see also [29] for a related study on barycenters in Gromov hyperbolic spaces).

The *Gromov hyperbolicity*, introduced in a seminal work [11] of Gromov, is a notion of negative curvature of large scale. A metric space (X, d) is said to be Gromov hyperbolic if it is δ*-hyperbolic* for some $\delta \geq 0$ in the sense that

$$(x|z)_p \geq \min\{(x|y)_p, (y|z)_p\} - \delta \tag{1.2}$$

holds for all $p, x, y, z \in X$, where

$$(x|y)_p := \frac{1}{2}\big\{d(p,x) + d(p,y) - d(x,y)\big\}$$

is the *Gromov product*. If (1.2) holds with $\delta = 0$, then the quadruple p, x, y, z is isometrically embedded into a tree. Therefore, the δ-hyperbolicity means that (X, d) is close to a tree up to local perturbations of size δ (cf. Example 2.1(e)). Admitting such local perturbations is a characteristic feature of the Gromov hyperbolicity.

In the investigation of gradient flows in Gromov hyperbolic spaces, we should employ discrete-time gradient flows of large time step τ ("giant steps"), because of the inevitable local perturbations. Precisely, for a convex function $f\colon X \longrightarrow \mathbb{R}$ and an arbitrary initial point $x_0 \in X$, we study the behavior of recursive applications of the proximal operator (1.1):

$$x_k \in \mathsf{J}^f_\tau(x_{k-1}), \quad k \in \mathbb{N}. \tag{1.3}$$

The resulting sequence $x_0, x_1, x_2, \ldots$ can be regarded as a discrete approximation of a (continuous-time) gradient curve for f starting from x_0. If f is bounded below, then it is natural to expect that x_k converges to a minimizer of f; we refer to, e.g., [2, 32] for such convergence results in metric spaces with upper or lower curvature bounds. In [30, Theorem 1.1], we established that x_k is closer to a minimizer than x_{k-1}, up to an additional term depending on the hyperbolicity constant δ. The estimates in [30, Theorems 1.1, 1.3] can be thought of as contraction properties akin to tress, and were the first contraction estimates concerning gradient flows for convex functions on non-Riemannian spaces.

In this article, inspired by Hirai–Sakabe's recent work [16], we consider the case where the target convex function is unbounded below. Then, instead of convergence to a minimizer, we study divergence to the steepest direction in the boundary at infinity. Note that the structure of boundary at infinity ∂X, constructed as equivalence classes of rays in X, is well investigated under both nonpositive curvature and Gromov hyperbolicity. This kind of divergent phenomenon has been established by Karlsson–Margulis [20] for *semi-contractions* $\phi\colon X \longrightarrow X$ (namely $d(\phi(x), \phi(y)) \leq d(x, y)$) on nonpositively curved metric spaces (X, d). Precisely, their *multiplicative ergodic theorem* [20, Theorem 2.1] asserts that, for a complete, uniformly convex metric space (X, d) of nonpositive curvature in the sense of Busemann and almost every $x_0 \in X$, there exists a ray $\xi\colon [0, \infty) \longrightarrow X$ such that

$$\lim_{k\to\infty} \frac{d(\xi(\alpha k), x_k)}{k} = 0,$$

where $x_k := \phi(x_{k-1})$, provided $\alpha := \lim_{k\to\infty} d(x_0, x_k)/k > 0$. We refer to [8, Proposition 4.2] for an application to continuous-time gradient flows for Lipschitz convex functions on CAT(0)-spaces, and to [19, Proposition 5.1] for a generalization to semi-contractions on Gromov hyperbolic spaces.

The multiplicative ergodic theorem in [20] is applicable to the proximal operator on CAT(0)-spaces (thanks to the semi-contractivity of J^f_τ; see [3, Theorem 2.2.22]), while in [16, §3.2] they also studied gradient descent in Hadamard manifolds. The results in [16] extract and discretize convex optimization ingredients of the *moment-weight inequality* for reductive group actions by Georgoulas–Robbin–Salamon [10] in geometric invariant theory, and have applications to operator scaling problems (see also more recent [15]). In Gromov hyperbolic spaces, we employ the proximal operator with large step size τ (compared with the hyperbolicity constant δ) as we discussed above; however, it is not a semi-contraction (see Remark 4.3). Nonetheless, using the contraction estimate [30, Theorem 1.1] and analyzing the behavior of divergence to the boundary at infinity, we establish the following.

Theorem 1.1 (Divergence to the steepest direction)**.** *Let (X,d) be a proper, geodesic δ-hyperbolic space with $\partial X \neq \emptyset$, and $f\colon X \longrightarrow \mathbb{R}$ be an L-Lipschitz convex function for some $L > 0$. Assume that $\alpha := -\inf_{v\in\partial X} \partial_\infty f(v) > 0$. Then, we have the following.*

(i) *There exists a unique element $v_* \in \partial X$ satisfying $\partial_\infty f(v_*) < 0$.*

(ii) *For any $x_0 \in X$ and*

$$\tau > \frac{2^{11}L^7}{(4L^2+\alpha^2)^2\alpha^4}\delta, \tag{1.4}$$

the discrete-time gradient curve $(x_k)_{k\in\mathbb{N}}$ as in (1.3) converges to $v_* \in \partial X$.

Here, $\partial_\infty f(v)$ is the *asymptotic slope* defined by

$$\partial_\infty f(v) := \lim_{t\to\infty} \frac{f(\xi(t))}{t}$$

for a ray $\xi\colon [0,\infty) \longrightarrow X$ representing $v \in \partial X$. Since $L \geq \alpha$, for example,

$$\tau > \frac{82L^7}{\alpha^8}\delta$$

is sufficient for (1.4). We remark that the uniqueness in (i) is a specific feature of the negative curvature; there is no such uniqueness for CAT(0)-spaces. Note also that [16, §3.2] includes finer analysis for gradient descent under a concavity condition on f (called the *L-smoothness*).

Remark 1.2 (On Lipschitz continuity). It follows from the convexity of f that the (closed) sublevel set $X' := f^{-1}((-\infty, f(x_0)])$ is convex, and hence (X',d) is also a proper, geodesic δ-hyperbolic space. For this reason, in Theorem 1.1, it is in fact sufficient to assume that f is L-Lipschitz on X'.

This article is organized as follows. After preliminaries in Section 2 for Gromov hyperbolic spaces, Section 3 is devoted to analysis of the asymptotic slope of unbounded convex functions, including the proof of Theorem 1.1(i). In Section 4, we study discrete-time gradient flows and prove Theorem 1.1(ii).

2 Preliminaries for Gromov hyperbolic spaces

For $a, b \in \mathbb{R}$, we set $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$. Besides Gromov's original paper [11], we refer to [5, 6, 9, 38] for the basics and various applications of the Gromov hyperbolicity.

2.1 Gromov hyperbolic spaces

Let (X, d) be a metric space. Recall from the introduction that the *δ-hyperbolicity* for $\delta \geq 0$ is defined by

$$(x|z)_p \geq (x|y)_p \wedge (y|z)_p - \delta \tag{2.1}$$

for all $p, x, y, z \in X$, where

$$(x|y)_p := \frac{1}{2}\big\{d(p, x) + d(p, y) - d(x, y)\big\}$$

is the *Gromov product.* Since the triangle inequality implies

$$0 \leq (x|y)_p \leq d(p, x) \wedge d(p, y),$$

the Gromov product does not exceed the diameter $\operatorname{diam}(X) := \sup_{x,y \in X} d(x, y)$. Hence, if $\operatorname{diam}(X) \leq \delta$, then (X, d) is δ-hyperbolic. This also means that the local structure of X (up to size δ) is not influential in the δ-hyperbolicity.

Another important fact is that trees are 0-hyperbolic; in this sense, a δ-hyperbolic space is close to a tree up to an additive constant δ. The Gromov hyperbolicity can also be regarded as a large-scale notion of negative curvature. Let us mention some typical examples.

Example 2.1. (a) Complete, simply connected Riemannian manifolds of sectional curvature ≤ -1 (or, more generally, CAT(-1)-*spaces*) are Gromov hyperbolic (see, e.g., [6, Proposition H.1.2]).

(b) An important difference between CAT(-1)-spaces and Gromov hyperbolic spaces is that the latter admits some non-Riemannian Finsler manifolds. For instance, *Hilbert geometry* on a bounded convex domain in the Euclidean space is Gromov hyperbolic under mild convexity and smoothness conditions (see [21], [28, §6.5]).

(c) For the *Teichmüller space* of a surface of genus g with p punctures, the *Weil–Petersson metric* (which is incomplete, Riemannian) is known to be Gromov hyperbolic if and only if $3g - 3 + p \leq 2$ [7], whereas the *Teichmüller metric* (which is complete, Finsler) does not satisfy the Gromov hyperbolicity [24] (see also [28, §6.6]).

(d) The definition (2.1) makes sense for discrete spaces, and the Gromov hyperbolicity has found rich applications in group theory. A discrete group whose Cayley graph satisfies the Gromov hyperbolicity is called a *hyperbolic group*; we refer to [5, 11], [6, Part III].

(e) Suppose that (X, d) admits an isometric embedding $\phi\colon T \longrightarrow X$ from a tree (T, d_T) such that the δ-neighborhood $B(\phi(T), \delta)$ of $\phi(T)$ covers X. Then, since (T, d_T) is 0-hyperbolic, we can easily see that (X, d) is 6δ-hyperbolic.

We call (X,d) a *geodesic space* if any two points $x,y \in X$ are connected by a (minimal) *geodesic* $\gamma\colon [0,1] \longrightarrow X$ satisfying $\gamma(0)=x$, $\gamma(1)=y$, and $d(\gamma(s),\gamma(t)) = |s-t|d(x,y)$ for all $s,t \in [0,1]$. In this case, there are a number of characterizations of the Gromov hyperbolicity, most notably by the δ-*slimness* of geodesic triangles (see, e.g., [6, §III.H.1]). We remark that, by [4, Theorem 4.1], every δ-hyperbolic metric space can be isometrically embedded into a complete δ-hyperbolic geodesic space.

In a δ-hyperbolic space, for any $x,y,z \in X$ and any geodesic $\gamma\colon [0,1] \longrightarrow X$ from y to z, we have

$$d(x,\gamma) - 2\delta \le (y|z)_x \le d(x,\gamma), \tag{2.2}$$

where $d(x,\gamma) := \min_{s\in[0,1]} d(x,\gamma(s))$ (see [38, 2.33]). As a corollary, one can readily see the following; we give a proof for completeness (cf. [6, Lemma III.H.1.15]).

Lemma 2.2. *Let (X,d) be a δ-hyperbolic space and $x,y \in X$. Then, for any two geodesics $\gamma,\eta\colon [0,1] \longrightarrow X$ from x to y, we have*

$$d\big(\gamma(s),\eta(s)\big) \le 4\delta$$

for all $s \in [0,1]$.

Proof. Fix $s \in [0,1]$ and take $s' \in [0,1]$ such that $d(\gamma(s),\eta) = d(\gamma(s),\eta(s'))$. It follows from (2.2) and $(x|y)_{\gamma(s)} = 0$ that $d(\gamma(s),\eta(s')) \le 2\delta$. For $s'' \in [0,1]$ with $|s-s''|d(x,y) > 2\delta$, the triangle inequality implies

$$d\big(\gamma(s),\eta(s'')\big) \ge \big|d\big(x,\gamma(s)\big) - d\big(x,\eta(s'')\big)\big| = |s-s''|d(x,y) > 2\delta.$$

Hence, we find $|s-s'|d(x,y) \le 2\delta$ and

$$d\big(\gamma(s),\eta(s)\big) \le d\big(\gamma(s),\eta(s')\big) + |s-s'|d(x,y) \le 2\delta + 2\delta = 4\delta.$$

□

2.2 Gromov boundary

Next, we introduce the *Gromov boundary* ∂X of a proper, geodesic δ-hyperbolic space (X,d) (the *properness* means that every bounded closed set is compact). We refer to [6, §III.H.3] for further details, as well as to [38, §5] and [9, §3.4] for the more general non-proper, non-geodesic situation (see Remark 2.4 below).

A *ray* $\xi\colon [0,\infty) \longrightarrow X$ is a geodesic of unit speed, i.e., $d(\xi(s),\xi(t)) = |s-t|$ for all $s,t \ge 0$. Two rays $\xi,\zeta\colon [0,\infty) \longrightarrow X$ are said to be *asymptotic* if

$$\sup_{t\ge 0} d\big(\xi(t),\zeta(t)\big) < \infty.$$

Being asymptotic is an equivalence relation on the set of rays, and we denote by ∂X the associated equivalence classes. The equivalence class of a ray ξ will be denoted by $\xi(\infty) \in \partial X$. For any $p \in X$ and $v \in \partial X$, there exists a ray ξ with $\xi(0) = p$ and $\xi(\infty) = v$ (see [6, Lemma III.H.3.1]).

We set $\overline{X} := X \sqcup \partial X$ (called the *Gromov closure* or *bordification* of X). To endow $\overline{X}$ with a topology, we fix a point $p \in X$ and consider geodesics $\xi\colon [0,l) \longrightarrow X$ of unit speed with $\xi(0) = p$. If $l = \infty$, then ξ is a ray. If $l < \infty$, then we extend ξ by letting $\xi(t) := \xi(l)$ for $t > l$, and put $\xi(\infty) := \xi(l)$. In either case, we call $\xi\colon [0,\infty) \longrightarrow X$ a *generalized ray*. We say that two generalized rays $\xi, \zeta\colon [0,\infty) \longrightarrow X$ (emanating from p) are equivalent if $\xi(\infty) = \zeta(\infty)$ (in X or ∂X). Then, the set of equivalence classes of generalized rays is identified with $\overline{X}$. A sequence $(\xi_i)_{i\in\mathbb{N}}$ of generalized rays is said to converge to a generalized ray ξ if ξ_i converges to ξ uniformly on each bounded interval in $[0,\infty)$. This defines a topology of $\overline{X}$. Note that, thanks to the properness, this topology restricted to X coincides with the original topology of X by the Arzelà–Ascoli theorem.

One can see that $\overline{X}$ is compact (see [6, Proposition III.H.3.7], [9, Proposition 3.4.18]).

Proposition 2.3. *Let (X,d) be a proper, geodesic Gromov hyperbolic space. Then, $\overline{X}$ and ∂X are compact.*

Moreover, $\overline{X}$ is metrizable, though we will not use it (see [6, Exercise III.H.3.18(4)], [9, Proposition 3.6.13]).

Remark 2.4 (Gromov sequences). Alternatively, one can introduce ∂X by considering sequences $(x_i)_{i\in\mathbb{N}}$ in X such that $\lim_{i,j\to\infty}(x_i|x_j)_p = \infty$, where $p \in X$ is an arbitrarily fixed point. Such a sequence is called a *Gromov sequence*. Two Gromov sequences $(x_i)_{i\in\mathbb{N}}$ and $(y_i)_{i\in\mathbb{N}}$ are defined to be *equivalent* if $\lim_{i\to\infty}(x_i|y_i)_p = \infty$. Then, there exists a natural bijection from the equivalence classes of Gromov sequences to ∂X (see [6, Lemma III.H.3.13]). Precisely, each Gromov sequence $(x_i)_{i\in\mathbb{N}}$ converges to a point in ∂X, and its inverse map is given by associating a ray ξ with the Gromov sequence $(\xi(i))_{i\in\mathbb{N}}$. We stress that, for non-proper, non-geodesic Gromov hyperbolic spaces, these two notions of the boundary may not coincide (see [38, Remark 5.5], [9, Remark 3.4.4]).

3 Unbounded convex functions

Let (X,d) be a proper, geodesic δ-hyperbolic space, and $f\colon X \longrightarrow \mathbb{R}$ be an L-Lipschitz function for some $L > 0$ (i.e., $|f(x) - f(y)| \le Ld(x,y)$ for all $x, y \in X$). We say that f is (*weakly, geodesically*) *convex* if, for any pair of points $x, y \in X$, there is a geodesic $\gamma\colon [0,1] \longrightarrow X$ from x to y such that

$$f\big(\gamma(s)\big) \le (1-s)f(x) + sf(y) \tag{3.1}$$

for all $s \in [0,1]$. We remark that, iteratively choosing a midpoint satisfying (3.1), one can actually find a geodesic γ such that $f \circ \gamma$ is convex on $[0,1]$.

We are interested in the case where $\inf_X f = -\infty$. Then, to investigate the asymptotic behavior of f at infinity, we shall utilize the Gromov boundary ∂X.

3.1 Asymptotic slope

Define the *descending slope* of f at $x \in X$ by

$$|\nabla^- f|(x) := \limsup_{y\to x} \frac{[f(x) - f(y)] \vee 0}{d(x,y)} \in [0,\infty].$$

For $v \in \partial X$ represented by a ray $\xi\colon [0,\infty) \longrightarrow X$, we define the *asymptotic slope*

$$\partial_\infty f(v) := \lim_{t\to\infty} \frac{f(\xi(t))}{t} = \lim_{t\to\infty} \frac{f(\xi(t)) - f(\xi(0))}{t} \in (-\infty, \infty].$$

Note that, as a subsequential limit of geodesics $\gamma_i\colon [0,i] \longrightarrow X$ from $\xi(0)$ to $\xi(i)$ along which f is convex, there is a ray ζ such that $\zeta(0) = \xi(0)$, $\zeta(\infty) = v$, and $f \circ \zeta$ is convex (recall Lemma 2.2). Then, the function

$$t \longmapsto \frac{f(\zeta(t)) - f(\zeta(0))}{t} \tag{3.2}$$

is non-decreasing, thereby the limit as $t \to \infty$ indeed exists. Moreover, since ζ is asymptotic to ξ (i.e., $d(\xi(t), \zeta(t))$ is bounded), we have

$$\left| \frac{f(\xi(t))}{t} - \frac{f(\zeta(t))}{t} \right| \le \frac{Ld(\xi(t), \zeta(t))}{t} \to 0$$

as $t \to \infty$. Thus, $\partial_\infty f$ is well-defined. In what follows, we will always choose a ray ξ such that $f \circ \xi$ is convex.

We summarize necessary properties of the asymptotic slope $\partial_\infty f$ in the next proposition (cf. [18, Lemma 3.2], [16, Proposition 2.1] on CAT(0)-spaces). We will consider the case where $\inf_{v\in\partial X} \partial_\infty f(v) < 0$.

Proposition 3.1. *Let (X, d) be a proper, geodesic δ-hyperbolic space with $\partial X \neq \emptyset$, and $f\colon X \longrightarrow \mathbb{R}$ be an L-Lipschitz convex function.*

- (i) *If $\inf_{v\in\partial X} \partial_\infty f(v) > 0$, then f is bounded below and its minimum is attained at some point in X.*
- (ii) *We have*
$$\inf_{x\in X} |\nabla^- f|(x) \ge -\inf_{v\in\partial X} \partial_\infty f(v). \tag{3.3}$$
In particular, if $\inf_{v\in\partial X} \partial_\infty f(v) < 0$, then $\inf_{x\in X} |\nabla^- f|(x) > 0$.
- (iii) *$\partial_\infty f\colon \partial X \longrightarrow (-\infty, \infty]$ is lower semi-continuous.*
- (iv) *In the case of $\inf_{v\in\partial X} \partial_\infty f(v) < 0$, there exists unique $v_* \in \partial X$ such that $\partial_\infty f(v_*) < 0$.*

Proof. (i) On the contrary, suppose that there is a sequence $(x_i)_{i\in\mathbb{N}}$ such that $f(x_i) \to -\infty$. Then, by the Lipschitz continuity of f, $d(x_1, x_i) \to \infty$ necessarily holds. Fix $p \in X$. Since f is convex, we have $f \le f(p) \vee f(x_i)$ on some geodesic from p to x_i. Thanks to the compactness of $\overline{X}$, as a subsequential limit of those geodesics, we obtain a ray ξ such that $\xi(0) = p$ and $f(\xi(t)) \le f(p)$ for all $t > 0$. Hence, $\partial_\infty f(\xi(\infty)) \le 0$ holds, a contradiction.

The above argument actually shows that each sublevel set of f is bounded. Then, by the properness of X and the continuity of f, we can find a minimizer of f.

(ii) Fix $x \in X$ and $v \in \partial X$ represented by a ray ξ with $\xi(0) = x$. Then, we deduce from the convexity of f along ξ that

$$|\nabla^- f|(x) \ge \lim_{t\to 0} \frac{f(x) - f(\xi(t))}{t} \ge \frac{f(x) - f(\xi(t))}{t}$$

for all $t > 0$. Letting $t \to \infty$ shows $|\nabla^- f|(x) \geq -\partial_\infty f(v)$, which completes the proof.

(iii) Take $p \in X$ and a sequence $(v_i)_{i\in\mathbb{N}}$ in ∂X converging to some $v \in \partial X$, and let ξ_i and ξ be rays representing v_i and v with $\xi_i(0) = \xi(0) = p$, respectively.

We first assume that $\partial_\infty f(v) < \infty$. By the definition of $\partial_\infty f(v)$, for any $\varepsilon > 0$, we have

$$\frac{f(\xi(t_0)) - f(p)}{t_0} > \partial_\infty f(v) - \varepsilon$$

for some $t_0 > 0$. Since f is continuous and ξ_i uniformly converges to ξ on each bounded interval, we find $N \in \mathbb{N}$ such that, for all $i \geq N$,

$$\frac{f(\xi_i(t_0)) - f(p)}{t_0} > \partial_\infty f(v) - 2\varepsilon.$$

The monotonicity of the function (3.2) then implies $\partial_\infty f(v_i) > \partial_\infty f(v) - 2\varepsilon$, and hence

$$\liminf_{i\to\infty} \partial_\infty f(v_i) \geq \partial_\infty f(v) - 2\varepsilon.$$

Letting $\varepsilon \to 0$ completes the proof of the asserted lower semi-continuity:

$$\liminf_{i\to\infty} \partial_\infty f(v_i) \geq \partial_\infty f(v).$$

When $\partial_\infty f(v) = \infty$, the same argument shows that, for any $R > 0$,

$$\frac{f(\xi(t_0)) - f(p)}{t_0} \geq 2R$$

for some $t_0 > 0$, and

$$\partial_\infty f(v_i) \geq \frac{f(\xi_i(t_0)) - f(p)}{t_0} \geq R$$

for some $N \in \mathbb{N}$ and all $i \geq N$. Therefore, we have $\liminf_{i\to\infty} \partial_\infty f(v_i) \geq R$ and letting $R \to \infty$ yields $\liminf_{i\to\infty} \partial_\infty f(v_i) = \infty$ as desired.

(iv) The existence is obvious, thereby it suffices to prove the uniqueness. Assume, on the contrary, that there are distinct $v_1, v_2 \in \partial X$ such that

$$\partial_\infty f(v_1) \vee \partial_\infty f(v_2) \leq -\varepsilon$$

for some $\varepsilon > 0$. Fix $p \in X$ and let ξ_1 and ξ_2 be rays representing v_1 and v_2 with $\xi_1(0) = \xi_2(0) = p$, respectively. Thanks to the monotonicity of the function (3.2), it follows that

$$\frac{f(\xi_1(t)) - f(p)}{t} \vee \frac{f(\xi_2(t)) - f(p)}{t} \leq -\varepsilon$$

for all $t > 0$. Given $t > 0$, let $\gamma_t \colon [0,1] \longrightarrow X$ be a geodesic from $\xi_1(t)$ to $\xi_2(t)$ along which f is convex, and x_t be a point on γ_t attaining $d(p, \gamma_t)$. Then, we deduce from the convexity of f along γ_t that

$$f(x_t) \leq f\big(\xi_1(t)\big) \vee f\big(\xi_2(t)\big) \leq f(p) - t\varepsilon.$$

Moreover, (2.2) yields

$$d(p, x_t) \leq \big(\xi_1(t)|\xi_2(t)\big)_p + 2\delta.$$

Now, since $v_1 \neq v_2$, there is $C > 0$ such that $(\xi_1(t)|\xi_2(t))_p \le C$ for all $t > 0$ (recall Remark 2.4). This implies $d(p, x_t) \le C + 2\delta$ for all $t > 0$; however,

$$\lim_{t\to\infty} f(x_t) \le \lim_{t\to\infty} \{f(p) - t\varepsilon\} = -\infty.$$

This contradicts the Lipschitz continuity of f. □

We remark that the uniqueness in (iv) is a consequence of negative curvature (cf. Example 3.3), and then we readily find that v_* satisfies $\partial_\infty f(v_*) = \inf_{\partial X} \partial_\infty f$ (without applying the lower semi-continuity of $\partial_\infty f$).

The inequality (3.3) is called the *weak duality* in [16, Lemma 2.2]. Then, the *strong duality* means that $\inf_X |\nabla^- f| > 0$ if and only if $\inf_{\partial X} \partial_\infty f < 0$. To show the only if part, in [18, Lemma 3.4] and [16, Theorems 3.1, 3.7], continuous-time gradient flows or a concavity condition on f are used. Intuitively speaking, in Gromov hyperbolic spaces, it seems not easy to control infinitesimal quantities like $|\nabla^- f|$.

3.2 Hadamard spaces

For the sake of comparison, here we briefly discuss the case of nonpositively curved spaces. We refer to [16, 18] for more details.

A geodesic space (Y, d) is called a CAT(0)-*space* if, for any $x, y, z \in Y$ and any geodesic $\gamma\colon [0,1] \longrightarrow Y$ from y to z, we have

$$d^2\big(x, \gamma(s)\big) \le (1-s)d^2(x,y) + sd^2(x,z) - (1-s)sd^2(y,z)$$

for all $s \in [0,1]$ (in other words, $d^2(x,\cdot)$ is 2-convex). A complete CAT(0)-space is called an *Hadamard space*. A complete, simply connected Riemannian manifold is a CAT(0)-space if and only if its sectional curvature is nonpositive everywhere (thus, an *Hadamard manifold*). Trees and Euclidean buildings are fundamental non-smooth examples of CAT(0)-spaces.

For an Hadamard space (Y, d), one can define ∂Y as the equivalence classes of rays in the same manner as Subsection 2.2. Moreover, ∂Y is equipped with a natural metric $\angle_T$ called the *Tits metric*, which can be defined by

$$2\sin\left(\frac{\angle_T(v_1, v_2)}{2}\right) = \lim_{t\to\infty} \frac{d(\xi_1(t), \xi_2(t))}{t}$$

for rays ξ_1, ξ_2 associated with v_1, v_2. We call $(\partial Y, \angle_T)$ the *Tits boundary*. For example, when Y is a Euclidean space $\mathbb{R}^n$, then its Tits boundary is isometric to $\mathbb{S}^{n-1}$. The Tits boundary of a hyperbolic space $\mathbb{H}^n$ is discrete ($\angle_T(v_1, v_2) = \pi$ for any distinct v_1, v_2).

Given a Lipschitz convex function $f\colon Y \longrightarrow \mathbb{R}$ on an Hadamard space, the asymptotic slope $\partial_\infty f\colon \partial Y \longrightarrow (-\infty, \infty]$ can be defined as in the previous subsection. In this case, $\partial_\infty f$ is Lipschitz with respect to $\angle_T$ and strictly convex on $\{v \in \partial Y \mid \partial_\infty f(v) < 0\}$. Therefore, when $\inf_{\partial Y} \partial_\infty f < 0$, we can find a unique minimizer $v_* \in \partial Y$ of $\partial_\infty f$, similarly to Proposition 3.1 (see [18, Lemma 3.2]).

Remark 3.2 (Recession functions). In [16], $\partial_\infty f$ (or its canonical extension to the Euclidean cone $C[\partial Y]$) is called a *recession function* in connection with convex analysis.

Example 3.3 (Busemann functions). Given a ray $\xi\colon [0,\infty) \longrightarrow Y$ in an Hadamard space, define the associated *Busemann function* $b_\xi\colon Y \longrightarrow \mathbb{R}$ by

$$b_\xi(x) := \lim_{t\to\infty}\bigl\{d\bigl(x,\xi(t)\bigr) - t\bigr\}.$$

Inheriting the properties of the distance function, b_ξ is 1-Lipschitz and convex. Note also that $b_\xi(\xi(t)) = -t$. A level (resp. sublevel) set of b_ξ is called a *horosphere* (resp. *horoball*). If $Y = \mathbb{R}^n$, then we have $b_\xi(x) = -\langle x - \xi(0), \dot{\xi}(0)\rangle$ with the Euclidean inner product $\langle\cdot,\cdot\rangle$, and horospheres are hyperplanes perpendicular to (the extension of) ξ. When Y is a hyperbolic space of Poincaré model, each horosphere of b_ξ is drawn as a sphere tangent to the boundary at $\xi(\infty)$. In either case, v_* for b_ξ is given by $\xi(\infty)$.

4 Discrete-time gradient flows

As in Theorem 1.1, let (X,d) be a proper, geodesic δ-hyperbolic space with $\partial X \neq \emptyset$, and $f\colon X \longrightarrow \mathbb{R}$ be an L-Lipschitz convex function.

4.1 Proximal point algorithm

For $\tau > 0$ and $x \in X$, the *proximal* (or *resolvent*) *operator* is defined by

$$\mathsf{J}^f_\tau(x) := \operatorname*{arg\,min}_{y\in X}\left\{f(y) + \frac{d^2(x,y)}{2\tau}\right\}.$$

Roughly speaking, an element in $\mathsf{J}^f_\tau(x)$ can be regarded as an approximation of a point on the gradient curve of f at time τ from x. Note that $\mathsf{J}^f_\tau(x) \neq \emptyset$ by the properness of (X,d) (see also the beginning of [30, §3.1]).

For any $x \in X$ and $x_\tau \in \mathsf{J}^f_\tau(x)$, we infer from the L-Lipschitz continuity of f that

$$f(x_\tau) + \frac{d^2(x,x_\tau)}{2\tau} \le f(x) \le f(x_\tau) + Ld(x,x_\tau).$$

This implies

$$d(x,x_\tau) \le 2\tau L. \tag{4.1}$$

By using the convexity of f, we can also provide a lower bound of $d(x,x_\tau)$.

Lemma 4.1. *Suppose that $\alpha := -\inf_{\partial X}\partial_\infty f > 0$. Then, for any $x \in X$ and $x_\tau \in \mathsf{J}^f_\tau(x)$, we have*

$$d(x,x_\tau) \ge \bigl(L - \sqrt{L^2-\alpha^2}\bigr)\tau \ge \frac{\alpha^2\tau}{2L}. \tag{4.2}$$

Moreover,

$$f(x_\tau) \le f(x) - \frac{\alpha^4\tau}{8L^2}. \tag{4.3}$$

We remark that, by (3.3),

$$L \ge \inf_X |\nabla^- f| \ge -\inf_{\partial X}\partial_\infty f = \alpha.$$

Proof. Let $v_* \in \partial X$ be the unique element satisfying $\partial_\infty f(v_*) = -\alpha$, and ξ be a ray with $\xi(0) = x$ and $\xi(\infty) = v_*$. By the monotonicity of the function (3.2), we have

$$\frac{f(\xi(t)) - f(x)}{t} \le -\alpha$$

for all $t > 0$. Since $d(x, \xi(t)) = t$, the above inequality is rewritten as

$$f\big(\xi(t)\big) + \frac{d^2(x, \xi(t))}{2\tau} \le f(x) - \alpha t + \frac{t^2}{2\tau},$$

and the right hand side takes its minimum at $t = \alpha\tau$, yielding

$$f\big(\xi(\alpha\tau)\big) + \frac{d^2(x, \xi(\alpha\tau))}{2\tau} \le f(x) - \frac{\alpha^2\tau}{2}. \tag{4.4}$$

Now, for any $y \in X$ with $d(x, y) < (L - \sqrt{L^2 - \alpha^2})\tau$, we deduce from the L-Lipschitz continuity of f that

$$\begin{aligned} f(y) + \frac{d^2(x,y)}{2\tau} &\ge f(x) - Ld(x,y) + \frac{d^2(x,y)}{2\tau} \\ &= f(x) + \frac{1}{2\tau}\Big\{\big(L\tau - d(x,y)\big)^2 - L^2\tau^2\Big\} \\ &> f(x) + \frac{1}{2\tau}\Big\{(L^2 - \alpha^2)\tau^2 - L^2\tau^2\Big\} \\ &= f(x) - \frac{\alpha^2\tau}{2}. \end{aligned}$$

Comparing this with (4.4) shows that $y \notin \mathsf{J}_\tau^f(x)$, thereby the former inequality $d(x, x_\tau) \ge (L - \sqrt{L^2 - \alpha^2})\tau$ in (4.2) necessarily holds. The latter inequality in (4.2) is immediate.

The second assertion (4.3) follows from the choice of x_τ together with (4.2) as

$$f(x) \ge f(x_\tau) + \frac{d^2(x, x_\tau)}{2\tau} \ge f(x_\tau) + \frac{\alpha^4\tau}{8L^2}.$$

□

We remark that the δ-hyperbolicity was not used in the lemma above. It will come into play via the following estimate from [30] (we state only the case of $K = 0$).

Theorem 4.2 ([30]). *Let (X, d) be a proper, geodesic δ-hyperbolic space and $f\colon X \longrightarrow \mathbb{R}$ be an L-Lipschitz convex function. Then, for any $x \in X$, $y \in \mathsf{J}_\tau^f(x)$, and $p \in X$ with $f(p) \le f(y)$, we have*

$$d(p, y) \le d(p, x) - d(x, y) + 4\sqrt{2\tau L\delta}. \tag{4.5}$$

We remark that, in [30, Theorem 1.1], p was chosen as a minimizer of f; however, by having a look on its proof, it is sufficient to assume $f(p) \le f(y)$.

Remark 4.3 (No semi-contraction). We also obtained a kind of contraction property in [30], whereas it does not imply that the proximal operator is a semi-contraction. For $y_i \in \mathsf{J}_\tau^f(x_i)$ $(i = 1, 2)$ and any minimizer $p \in X$ of f, [30, Theorem 1.3(ii)] (with $K = 0$) asserts that

$$d(y_1, y_2) \le d(x_1, x_2) - (p|x_2)_{x_1} + C(L, D, \tau, \delta),$$

where $D := d(p, x_1) \vee d(p, x_2)$, provided $d(p, y_1) \le d(p, y_2) \wedge (x_1|x_2)_p$. If $x_2 \to x_1$, then $d(x_1, x_2) - (p|x_2)_{x_1} \to 0$, but the additional term $C(L, D, \tau, \delta)$ caused by the δ-hyperbolicity remains.

4.2 Proof of Theorem 1.1

We have shown (i) in Proposition 3.1, here we prove the remaining assertion (ii).

Proof of Theorem 1.1(ii). Fix an arbitrary initial point $x_0 \in X$ and recursively choose $x_k \in \mathsf{J}_\tau^f(x_{k-1})$ for $k \in \mathbb{N}$. We find from (4.3) and the Lipschitz continuity of f that $d(x_0, x_k) \to \infty$ as $k \to \infty$. By the compactness of $\overline{X}$, $(x_k)_{k\in\mathbb{N}}$ admits a subsequence converging to some $v \in \partial X$. We shall show that v necessarily coincides with v_* given in Proposition 3.1(iv).

It follows from (4.1) and (4.3) that

$$d(x_0, x_k) \le 2k\tau L, \qquad f(x_k) \le f(x_0) - k\frac{\alpha^4\tau}{8L^2},$$

and hence

$$\frac{f(x_k) - f(x_0)}{d(x_0, x_k)} \le -k\frac{\alpha^4\tau}{8L^2}\frac{1}{2k\tau L} = -\frac{\alpha^4}{16L^3}.$$

Note also that

$$d(x_0, x_k) \ge \frac{f(x_0) - f(x_k)}{L} \ge k\frac{\alpha^4\tau}{8L^3}. \tag{4.6}$$

Suppose that $v \neq v_*$. Let $(x_{k_i})_{i\in\mathbb{N}}$ be a subsequence converging to v, and ξ be a ray representing v_* with $\xi(0) = x_0$. Since $v \neq v_*$, we have

$$\Theta := \sup_{t>0,\, i\in\mathbb{N}} \big(\xi(t)|x_{k_i}\big)_{x_0} < \infty$$

(recall Remark 2.4). Together with (4.6), this implies that

$$d\big(\xi(t), x_{k_i}\big) \ge d\big(x_0, \xi(t)\big) + d(x_0, x_{k_i}) - 2\Theta \ge t + k_i\frac{\alpha^4\tau}{8L^3} - 2\Theta \tag{4.7}$$

for all $t > 0$, $i \in \mathbb{N}$. On the other hand, it follows from (4.5) and (4.2) that, for large $t > 0$ satisfying $f(\xi(t)) \le f(x_{k+1})$ (note that $f(\xi(t)) \to -\infty$ as $t \to \infty$),

$$d\big(\xi(t), x_{k+1}\big) - d\big(\xi(t), x_k\big) \le -d(x_k, x_{k+1}) + 4\sqrt{2\tau L\delta} \le -\frac{\alpha^2\tau}{2L} + 4\sqrt{2\tau L\delta}.$$

Therefore, for $t > 0$ with $f(\xi(t)) \le f(x_{k_i})$,

$$d\big(\xi(t), x_{k_i}\big) \le d\big(\xi(t), x_0\big) - k_i\left(\frac{\alpha^2\tau}{2L} - 4\sqrt{2\tau L\delta}\right).$$

Combining this with (4.7) yields

$$k_i\left(\frac{\alpha^2\tau}{2L} - 4\sqrt{2\tau L\delta}\right) \le t - d\big(\xi(t), x_{k_i}\big) \le -k_i\frac{\alpha^4\tau}{8L^3} + 2\Theta,$$

thereby

$$k_i\sqrt{\tau}\left\{\left(\frac{\alpha^2}{2L} + \frac{\alpha^4}{8L^3}\right)\sqrt{\tau} - 4\sqrt{2L\delta}\right\} \le 2\Theta.$$

Since the hypothesis (1.4) means

$$\left(\frac{\alpha^2}{2L} + \frac{\alpha^4}{8L^3}\right)\sqrt{\tau} - 4\sqrt{2L\delta} > 0,$$

letting $i \to \infty$ induces a contradiction. Therefore, we obtain $v = v_*$. $\square$

4.3 Further problems

We close the article with some further problems, including those discussed at the end of [30].

(A) As we explained in Example 2.1(d), the Gromov hyperbolicity makes sense for discrete spaces as well. Therefore, it is interesting to explore some generalizations of the results in this article and [30] to discrete (non-geodesic) Gromov hyperbolic spaces. Then, it is a challenging problem to formulate and analyze convex functions on discrete Gromov hyperbolic spaces (possibly for some special classes such as hyperbolic groups). We refer to [26] for the theory of convex functions on $\mathbb{Z}^N$ (called *discrete convex analysis*), and to [13, 22] for some generalizations to graphs and trees, respectively.

(B) Even in geodesic Gromov hyperbolic spaces, it is worthwhile considering a certain "large-scale convexity" of functions, preserved by *quasi-isometries*, since the Gromov hyperbolicity is preserved by quasi-isometries between geodesic spaces (see, e.g., [6, Theorem III.H.1.9], [38, Theorem 3.18]).

(C) Since [16, 18, 20] are concerned with nonpositively curved spaces, it is natural to expect that our results can be extended to some class of metric spaces including both CAT(0)-spaces and Gromov hyperbolic spaces, probably defined through an appropriate relaxation (perturbation) of the CAT(0)-condition.

Acknowledgements. I would like to thank Hiroshi Hirai for stimulating discussions. This work was supported in part by the JSPS Grant-in-Aid for Scientific Research (KAKENHI) 22H04942, 24K00523, 24K21511.

References

[1] L. Ambrosio, N. Gigli and G. Savaré, Gradient flows in metric spaces and in the space of probability measures. Second edition. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2008.

[2] M. Bačák, The proximal point algorithm in metric spaces. Israel J. Math. **194** (2013), 689–701.

[3] M. Bačák, Convex analysis and optimization in Hadamard spaces. De Gruyter Series in Nonlinear Analysis and Applications, **22**. De Gruyter, Berlin, 2014.

[4] M. Bonk and O. Schramm, Embeddings of Gromov hyperbolic spaces. Geom. Funct. Anal. **10** (2000), 266–306.

[5] B. H. Bowditch, A course on geometric group theory. MSJ Memoirs, **16**. Mathematical Society of Japan, Tokyo, 2006.

[6] M. R. Bridson and A. Haefliger, Metric spaces of non-positive curvature. Grundlehren der mathematischen Wissenschaften, **319**. Springer-Verlag, Berlin, 1999.

[7] J. Brock and B. Farb, Curvature and rank of Teichmüller space. Amer. J. Math. **128** (2006), 1–22.

[8] P.-E. Caprace and A. Lytchak, At infinity of finite-dimensional CAT(0) spaces. Math. Ann. **346** (2010), 1–21.

[9] T. Das, D. Simmons and M. Urbański, Geometry and dynamics in Gromov hyperbolic metric spaces. With an emphasis on non-proper settings. Mathematical Surveys and Monographs, **218**. American Mathematical Society, Providence, RI, 2017.

[10] V. Georgoulas, J. W. Robbin and D. A. Salamon, The moment-weight inequality and the Hilbert–Mumford criterion—GIT from the differential geometric viewpoint. Lecture Notes in Math., **2297**, Springer, Cham, 2021.

[11] M. Gromov, Hyperbolic groups. Essays in group theory, 75–263, Math. Sci. Res. Inst. Publ., **8**, Springer, New York, 1987.

[12] M. Hamada and H. Hirai, Computing the nc-rank via discrete convex optimization on CAT(0) spaces. SIAM J. Appl. Algebra Geom. **5** (2021), 455–478.

[13] H. Hirai, Discrete convex functions on graphs and their algorithmic applications. Combinatorial optimization and graph algorithms, 67–100, Springer, Singapore, 2017.

[14] H. Hirai, Convex analysis on Hadamard spaces and scaling problems. Found. Comput. Math. **24** (2024), 1979–2016.

[15] H. Hirai, A scaling characterization of nc-rank via unbounded gradient flow. Preprint (2025), available at arXiv:2502.05412

[16] H. Hirai and K. Sakabe, Gradient descent for unbounded convex functions on Hadamard manifolds and its applications to scaling problems. 65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, pages 2387–2402. IEEE, 2024.

[17] J. Jost, Convex functionals and generalized harmonic maps into spaces of nonpositive curvature. Comment. Math. Helv. **70** (1995), 659–673.

[18] M. Kapovich, B. Leeb and J. Millson, Convex functions on symmetric spaces, side lengths of polygons and the stability inequalities for weighted configurations at infinity. J. Differential Geom. **81** (2009) 297–354.

[19] A. Karlsson, Non-expanding maps and Busemann functions. Ergodic Theory Dynam. Systems **21** (2001), 1447–1457.

[20] A. Karlsson and G. Margulis, A multiplicative ergodic theorem and nonpositively curved spaces. Comm. Math. Phys. **208** (1999), 107–123.

[21] A. Karlsson and G. A. Noskov, The Hilbert metric and Gromov hyperbolicity. Enseign. Math. (2) **48** (2002), 73–89.

[22] V. Kolmogorov, Submodularity on a tree: unifying $L^\natural$-convex and bisubmodular functions. Mathematical foundations of computer science 2011, 400–411, Lecture Notes in Comput. Sci., **6907**, Springer, Heidelberg, 2011.

[23] A. Lytchak, Open map theorem for metric spaces. St. Petersburg Math. J. **17** (2006), 477–491.

[24] H. A. Masur and M. Wolf, Teichmüller space is not Gromov hyperbolic. Ann. Acad. Sci. Fenn. Ser. A I Math. **20** (1995), 259–267.

[25] U. F. Mayer, Gradient flows on nonpositively curved metric spaces and harmonic maps. Comm. Anal. Geom. **6** (1998), 199–253.

[26] K. Murota, Discrete convex analysis. SIAM Monographs on Discrete Mathematics and Applications. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2003.

[27] S. Ohta, Gradient flows on Wasserstein spaces over compact Alexandrov spaces. Amer. J. Math. **131** (2009), 475–516.

[28] S. Ohta, Comparison Finsler geometry. Springer Monographs in Mathematics. Springer, Cham, 2021.

[29] S. Ohta, Barycenters and a law of large numbers in Gromov hyperbolic spaces. Rev. Mat. Iberoam. **40** (2024), 1185–1206.

[30] S. Ohta, Discrete-time gradient flows in Gromov hyperbolic spaces. Israel J. Math. (to appear), available at arXiv:2205.03156

[31] S. Ohta, How does the contraction property fail for convex functions on normed spaces? Preprint (2023), available at arXiv:2311.15152

[32] S. Ohta and M. Pálfia, Discrete-time gradient flows and law of large numbers in Alexandrov spaces. Calc. Var. Partial Differential Equations **54** (2015), 1591–1610.

[33] S. Ohta and M. Pálfia, Gradient flows and a Trotter–Kato formula of semi-convex functions on CAT(1)-spaces. Amer. J. Math. **139** (2017), 937–965.

[34] S. Ohta and K.-T. Sturm, Non-contraction of heat flow on Minkowski spaces, Arch. Ration. Mech. Anal. **204** (2012), 917–944.

[35] A. Petrunin, Semiconcave functions in Alexandrov's geometry. Surveys in differential geometry. Vol. XI, 137–201, Surv. Differ. Geom., **11**, Int. Press, Somerville, MA, 2007.

[36] G. Savaré, Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds. C. R. Math. Acad. Sci. Paris **345** (2007), 151–154.

[37] K.-T. Sturm, Gradient flows for semiconvex functions on metric measure spaces—existence, uniqueness, and Lipschitz continuity. Proc. Amer. Math. Soc. **146** (2018), 3985–3994.

[38] J. Väisälä, Gromov hyperbolic spaces. Expo. Math. **23** (2005), 187–231.