# Barycenters and a law of large numbers in Gromov hyperbolic spaces

Shin-ichi Ohta

**Abstract.** We investigate barycenters of probability measures on Gromov hyperbolic spaces, toward development of convex optimization in this class of metric spaces. We establish a contraction property (the Wasserstein distance between probability measures provides an upper bound of the distance between their barycenters), a deterministic approximation of barycenters of uniform distributions on finite points, and a kind of law of large numbers. These generalize the corresponding results on CAT(0)-spaces, up to additional terms depending on the hyperbolicity constant.

## 1. Introduction

This article is a continuation of [27], in which we studied discrete-time gradient flows for geodesically convex functions on (geodesic, proper) Gromov hyperbolic spaces. The theory of gradient flows for convex functions possesses fundamental importance in analysis, geometry and optimization theory, and has been well investigated in some classes of "Riemannian" metric spaces including CAT(0)-*spaces* (nonpositively curved metric spaces in the sense of triangle comparison; we refer to [4]). For "non-Riemannian" metric spaces such as normed spaces and Finsler manifolds, however, much less is known and, in fact, there is a large gap between properties of gradient flows in Riemannian and Finsler manifolds (see [27, 30] for further discussions).

Intending to develop optimization theory in possibly non-Riemannian spaces, in [27] we studied discrete-time gradient flows for geodesically convex functions on Gromov hyperbolic spaces, and showed some contraction estimates. Gromov hyperbolic spaces are metric spaces negatively curved in large-scale, and it is known that some non-Riemannian Finsler manifolds can be Gromov hyperbolic (see Example 2.2). The class of geodesically convex functions seems, however, restrictive when one has in mind the local flexibility of the Gromov hyperbolicity condition. Thereby, it is desirable to generalize the theory of gradient flows to a wider class of "convex functions" (we refer to Section 3.4 in [27] for related discussions). As an initial step toward such a generalization, in this article, we study the (squared) distance function on a Gromov hyperbolic space, which should be included in the class of generalized convex functions (in view of Lemmas 3.2 and 4.2).

Given a probability measure $\mu$ on a metric space $(X, d)$ with finite second moment, its *barycenter* is defined as a minimizer of the function $x \mapsto \int_X d^2(x, z)\, \mu(dz)$. If the (squared) distance function is sufficiently convex, then barycenters enjoy a number of fine properties. Specifically, in a CAT(0)-space $(X, d)$, for which $d^2$ is strictly convex by definition, every $\mu$ admits a unique barycenter $\beta_\mu \in X$ and we have a *contraction property* $d(\beta_\mu, \beta_\nu) \leq W_1(\mu, \nu)$ in terms of the $L^1$-Wasserstein distance $W_1$ (see [32]). Moreover, a kind of *law of large numbers* providing the almost sure convergence to barycenters by recursive applications of the proximal (resolvent) operator was established in [32] (we refer to [28, 38] for some generalizations).

A metric space $(X, d)$ is said to be *Gromov hyperbolic* (attributed to [16]) if it is $\delta$-*hyperbolic* for some $\delta \geq 0$ in the sense that

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

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

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

is the *Gromov product*. This is a large-scale notion of negative curvature and hence, on the one hand, it is natural to expect some variants of the aforementioned results in CAT(0)-spaces. On the other hand, since the Gromov hyperbolicity provides no local control (up to the hyperbolicity constant $\delta$), one cannot expect very sharp estimates as in the case of CAT(0)-spaces. Accordingly, our results will have additional terms (compared with the case of CAT(0)-spaces) that tend to 0 as $\delta \to 0$.

For a probability measure $\mu$ on a (geodesic) $\delta$-hyperbolic space $(X, d)$, its barycenter is not unique but lives in a bounded set (whose diameter tends to 0 as $\delta \to 0$; see Proposition 3.3). For this reason, we introduce the set

$$\mathcal{B}(\mu, \varepsilon) := \left\{ x \in X \;\middle|\; \int_X d^2(x, z)\, \mu(dz) \leq \inf_{y \in X} \int_X d^2(y, z)\, \mu(dz) + \varepsilon \right\}$$

for $\varepsilon \geq 0$, and call it a *barycentric set* (here we consider only probability measures of finite second moment for simplicity). Then, we show a contraction property of the form

$$d(x, y) \leq W_1(\mu, \nu) + \sqrt{6\varepsilon} + O(\delta^{1/4})$$

for $x \in \mathcal{B}(\mu, \varepsilon)$ and $y \in \mathcal{B}(\nu, \varepsilon)$; see Theorem 4.5 for the precise statement.

How to find (or approximate) a barycenter of a given probability measure $\mu$ is a fundamental problem. In this respect, we show the following law of large numbers (see Theorem 6.1 for the precise statement): given a sequence $(Z_i)_{i \geq 1}$ of independent, identically distributed random variables with distribution $\mu$, and an arbitrary initial point $S_0 \in X$, we consider a sequence $(S_k)_{k \geq 0}$ recursively chosen as $S_{k+1} = \gamma((2\tau + 1)^{-1})$ for a minimal geodesic $\gamma \colon [0, 1] \to X$ from $Z_{k+1}$ to $S_k$. Then, for any $\varepsilon > 0$, we have

$$\mathbb{E}[d^2(p, S_{k_0})] \leq \varepsilon + O(\sqrt{\delta})$$

for some $k_0 \leq C/(\varepsilon\sqrt{\delta})$ by taking $\tau$ proportional to $\sqrt{\delta}$, where $p \in \mathcal{B}(\mu, 0)$ is a barycenter of $\mu$. We remark that the construction of $S_{k+1}$ from $S_k$ is written by the proximal operator

$S_{k+1} \in \mathsf{J}_\tau^{f_k}(S_k)$ for the distance function $f_k = d^2(Z_{k+1}, \cdot)$ (see (5.1)), and that such an operation is meaningful only for large $\tau$ compared with $\delta$ due to the local flexibility of $\delta$-hyperbolic spaces. By a similar analysis, we prove in Theorem 5.2 a *deterministic approximation* of barycenters of uniform distributions on finite points, as a generalization of [23] in CAT(0)-spaces (we refer to [2, 3, 17, 28] for some related results).

We briefly mention related results on *hyperbolic groups* (discrete groups whose Cayley graphs are Gromov hyperbolic). Laws of large numbers can be formulated in terms of the behavior of the distance function $d(g_1 g_2 \cdots g_k(x_0), x_0)$ for a sequence $(g_i)_{i \geq 1}$ of independent, identically distributed random variables taking values in the group of isometric transformations acting on a metric space $(X, d)$ and a base point $x_0 \in X$. Indeed, Kingman's subadditive ergodic theorem [21, 22] ensures that

$$\lim_{k \to \infty} \frac{d(g_1 g_2 \cdots g_k(x_0), x_0)}{k}$$

exists almost surely and is constant almost everywhere. We refer to [19] and the references therein for more details, as well as for refined results including the case of hyperbolic groups. In this context, *central limit theorems* in hyperbolic groups were also established in [5, 7]. Our law of large numbers (Theorem 6.1) associated with a probability measure $\mu$ on a Gromov hyperbolic space $X$ is concerned with a more general setting, and provides a direct approximation of barycenters. In this generality, it is difficult even to formulate central limit theorems.

This article is organized as follows. In Section 2, we review the basics of Gromov hyperbolic spaces and some facts on barycenters in CAT(0)-spaces. In Section 3, we introduce and analyze barycentric sets for probability measures on Gromov hyperbolic spaces. Then we discuss the Wasserstein contraction property, a deterministic approximation, and a law of large numbers in Sections 4, 5, and 6, respectively.

## 2. Preliminaries

We review the basics of Gromov hyperbolic spaces, as well as some facts on barycenters in CAT(0)-spaces related to our results. For $a, b \in \mathbb{R}$, we set $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$.

### 2.1. Gromov hyperbolic spaces

Besides the original paper [16], we refer to [9, 10, 13, 31, 34] for the basics and various applications of the Gromov hyperbolicity.

Let $(X, d)$ be a metric space. For three points $x, y, z \in X$, we define the *Gromov product* $(y|z)_x$ by

$$(y|z)_x := \frac{1}{2} \{d(x, y) + d(x, z) - d(y, z)\}.$$

Observe from the triangle inequality that $0 \leq (y|z)_x \leq d(x, y) \wedge d(x, z)$. In the Euclidean plane $\mathbb{R}^2$, $(y|z)_x$ is understood as the distance from $x$ to the intersection of the

triangle $\triangle xyz$ and its inscribed circle. If $x, y, z$ are the endpoints of a tripod, then $(y|z)_x$ coincides with the distance from $x$ to the branching point.

**Definition 2.1** (Gromov hyperbolic spaces). For $\delta \geq 0$, a metric space $(X, d)$ is said to be $\delta$-*hyperbolic* if

$$(2.1) \qquad\qquad (x|z)_p \geq (x|y)_p \wedge (y|z)_p - \delta$$

holds for all $p, x, y, z \in X$. We say that $(X, d)$ is *Gromov hyperbolic* if it is $\delta$-hyperbolic for some $\delta \geq 0$.

Since the Gromov product does not exceed the diameter $\mathrm{diam}(X) := \sup_{x,y \in X} d(x, y)$, if $\mathrm{diam}(X) \leq \delta$, then $(X, d)$ is $\delta$-hyperbolic. This also means that the local structure of $X$ (up to size $\delta$) is not influential in the $\delta$-hyperbolicity. Another fact worth mentioning is that, if (2.1) holds for some $p \in X$ and all $x, y, z \in X$, then $(X, d)$ is $2\delta$-hyperbolic (see Corollary 1.1.B in [16]).

The Gromov hyperbolicity can be regarded as a large-scale notion of negative (sectional) curvature. We recall some fundamental examples (see also Section 1 in [16]).

**Example 2.2.** (a) Complete, simply connected Riemannian manifolds of sectional curvature $\leq -1$ (or, more generally, CAT$(-1)$-spaces) are Gromov hyperbolic (see, for example, Section 1.5 in [16]).

(b) An important difference between the class of CAT$(-1)$-spaces and that of Gromov hyperbolic spaces is that the latter admits 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 [20] and Section 6.5 in [26]).

(c) The definition (2.1) makes sense for discrete spaces. In fact, 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 [9, 16], and Part III of [10]). In the sequel, however, we do not consider discrete spaces.

(d) Assume that a metric space $(X, d_X)$ admits a map $\phi : T \to X$ from a tree $(T, d_T)$ such that $d_X(\phi(a), \phi(b)) = d_T(a, b)$ for all $a, b \in T$ and the $\delta$-neighborhood $B(\phi(T), \delta)$ of $\phi(T)$ covers $X$. Then, since $(T, d_T)$ is $0$-hyperbolic, we can easily see that $(X, d_X)$ is $6\delta$-hyperbolic.

We call $(X, d)$ a *geodesic space* if any two points $x, y \in X$ are connected by a *minimal geodesic* $\gamma : [0, 1] \to X$ satisfying $\gamma(0) = x$, $\gamma(1) = y$, and $d(\gamma(s), \gamma(t)) = |s - t| \cdot d(x, y)$ for all $s, t \in [0, 1]$. In this case, there are several characterizations of the Gromov hyperbolicity (see, for example, Section 6 in [16] and Section III.H.1 in [10]). We also remark that, by Theorem 4.1 in [8], every $\delta$-hyperbolic metric space can be isometrically embedded into a complete, geodesic $\delta$-hyperbolic space.

## 2.2. CAT(0)-spaces

A geodesic space $(X, d)$ is called a CAT$(0)$-*space* if, for any $x, y, z \in X$ and any minimal geodesic $\gamma : [0, 1] \to X$ from $x$ to $y$,

$$(2.2) \qquad d^2(z, \gamma(t)) \leq (1 - t)\, d^2(z, x) + t\, d^2(z, y) - (1 - t)t\, d^2(x, y)$$

holds for all $t \in (0, 1)$. A complete, simply connected Riemannian manifold is a CAT(0)-space if and only if its sectional curvature is nonpositive everywhere. Moreover, there are a number of rich classes of non-smooth CAT(0)-spaces such as Euclidean buildings, trees, phylogenetic tree spaces, and the orthoscheme complexes of modular lattices (see [4, 6, 12, 18]).

The CAT(0)-inequality (2.2) can be regarded as the uniform strict convexity of the squared distance function $d^2(z, \cdot)$, and such a convexity is known to be quite useful to study barycenters. In fact, as we mentioned in the introduction, every probability measure $\mu$ on $X$ with finite second (or first) moment admits a unique barycenter $\beta_\mu \in X$, and the contraction property $d(\beta_\mu, \beta_\nu) \leq W_1(\mu, \nu)$ holds. Moreover, a law of large numbers was established in [32], followed by many variants and generalizations [3, 17, 23, 28, 38] (see also [15, 24] for related works on a different notion of barycenter).

## 3. Barycentric sets

Henceforth, throughout the article, let $(X, d)$ be a geodesic $\delta$-hyperbolic space. We first recall the Wasserstein distance between probability measures (we refer to [35] for further reading).

Denote by $\mathcal{P}^p(X)$ ($p \in [1, \infty)$) the set of Borel probability measures on $X$ of finite $p$-th moment, that is, $\mu \in \mathcal{P}^p(X)$ if

$$\int_X d^p(x_0, x) \, \mu(dx) < \infty$$

for some (and hence any) $x_0 \in X$. For $\mu, \nu \in \mathcal{P}^p(X)$, the $L^p$-*Wasserstein distance* (or the *Kantorovich distance*) between $\mu$ and $\nu$ is defined by

$$W_p(\mu, \nu) := \inf_\pi \left( \int_{X \times X} d^p(x, y) \, \pi(dx \, dy) \right)^{1/p},$$

where $\pi$ runs over all *couplings* of $(\mu, \nu)$ (namely probability measures on $X \times X$ with marginals $\mu$ and $\nu$). A coupling $\pi$ attaining the above infimum is called an $L^p$-*optimal coupling* of $(\mu, \nu)$. Observe that

$$W_p(\delta_x, \mu) = \left( \int_X d^p(x, y) \, \mu(dy) \right)^{1/p}$$

for all $x \in X$ and $\mu \in \mathcal{P}^p(X)$, where $\delta_x$ denotes the Dirac mass at $x$.

Note that $\mathcal{P}^2(X) \subset \mathcal{P}^1(X)$ by the Hölder (or Cauchy–Schwarz) inequality. According to [32], we will consider barycenters of probability measures not only in $\mathcal{P}^2(X)$ but also in $\mathcal{P}^1(X)$. Fix an arbitrary point $x_0 \in X$. For $\mu \in \mathcal{P}^1(X)$, we define

$$(3.1) \qquad \mathbf{V}_{x_0}(\mu) := \inf_{x \in X} \int_X \{d^2(x, z) - d^2(x_0, z)\} \, \mu(dz).$$

We remark that the integral above is well-defined since, by the triangle inequality,

$$\int_X |d^2(x, z) - d^2(x_0, z)| \, \mu(dz) \leq \int_X d(x, x_0) \{d(x, z) + d(x_0, z)\} \, \mu(dz)$$
$$= d(x, x_0) \{W_1(\delta_x, \mu) + W_1(\delta_{x_0}, \mu)\} < \infty.$$

In a complete CAT(0)-space, every $\mu \in \mathcal{P}^1(X)$ admits a unique point $x \in X$ attaining the infimum in (3.1). Such a point $x$ is independent of the choice of $x_0$ and called the *barycenter* of (or the *center of mass* for) $\mu$ (see Proposition 4.3 in [32], and [36] for the case of CAT(1)-spaces of small radii). More precisely, what we consider is an $L^2$-barycenter involving the squared distance. We refer to [1, 37] for related works on $L^p$-barycenters.

In a $\delta$-hyperbolic space, however, we do not have a unique barycenter. For this reason, we introduce the following set:

$$(3.2) \qquad \mathcal{B}(\mu, \varepsilon) := \left\{ x \in X \;\middle|\; \int_X \{d^2(x, z) - d^2(x_0, z)\} \, \mu(dz) \leq \mathbf{V}_{x_0}(\mu) + \varepsilon \right\}$$

for $\mu \in \mathcal{P}^1(X)$ and $\varepsilon \geq 0$. We shall call $\mathcal{B}(\mu, \varepsilon)$ a *barycentric set*. We remark that $\mathcal{B}(\mu, \varepsilon)$ is independent of the choice of $x_0$. Note also that $\mathcal{B}(\mu, 0)$ may be empty (unless $X$ is proper), while $\mathcal{B}(\mu, \varepsilon) \neq \emptyset$ for any $\varepsilon > 0$.

Our goal in this section is to estimate the diameter of $\mathcal{B}(\mu, \varepsilon)$ in terms of $\varepsilon$ and $\delta$. To this end, we first generalize the CAT(0)-inequality (2.2) to $\delta$-hyperbolic spaces, with an inevitable additional term depending on $\delta$.

**Lemma 3.1** (CAT(0) + $\delta$). *For any $x, y, z \in X$ and any midpoint $w$ between $x$ and $y$, we have*

$$(3.3) \quad d^2(z, w) \leq \frac{d^2(z, x)}{2} + \frac{d^2(z, y)}{2} - \frac{d^2(x, y)}{4} + 2\delta \{d(z, x) + d(z, y)\} + 4\delta^2.$$

*Proof.* Since $w$ is a midpoint of $x$ and $y$ (namely $d(x, w) = d(w, y) = d(x, y)/2$), we have $(x|y)_w = 0$. Then the $\delta$-hyperbolicity (2.1) implies

$$0 \geq (x|z)_w \wedge (z|y)_w - \delta = \frac{1}{2} \left\{ d(z, w) + \frac{d(x, y)}{2} - d(z, x) \vee d(z, y) \right\} - \delta.$$

Hence,

$$d^2(z, w) \leq \left( d(z, x) \vee d(z, y) - \frac{d(x, y)}{2} + 2\delta \right)^2.$$

Setting $a = d(z, x) \vee d(z, y)$ and $b = d(z, x) \wedge d(z, y)$, we observe that

$$\begin{aligned}
\left( a - \frac{d(x, y)}{2} \right)^2 &= a^2 + \frac{d^2(x, y)}{4} - a\,d(x, y) \\
&= \frac{a^2 + b^2}{2} - \frac{d^2(x, y)}{4} - \frac{1}{2} \{d(x, y) - (a - b)\} \{(a + b) - d(x, y)\} \\
&\leq \frac{d^2(z, x)}{2} + \frac{d^2(z, y)}{2} - \frac{d^2(x, y)}{4},
\end{aligned}$$

since $a + b \geq d(x, y) \geq a - b$ by the triangle inequality. Moreover, we find

$$0 \leq a - \frac{d(x, y)}{2} \leq a - \frac{a - b}{2} = \frac{d(z, x) + d(z, y)}{2}.$$

Combining these, we obtain, as desired,

$$\begin{aligned}
d^2(z, w) &\leq \left( a - \frac{d(x, y)}{2} \right)^2 + 4\delta \left( a - \frac{d(x, y)}{2} \right) + 4\delta^2 \\
&\leq \frac{d^2(z, x)}{2} + \frac{d^2(z, y)}{2} - \frac{d^2(x, y)}{4} + 2\delta \{d(z, x) + d(z, y)\} + 4\delta^2. \qquad \blacksquare
\end{aligned}$$

We also present the corresponding inequality for general intermediate points between $x$ and $y$.

**Lemma 3.2** (General intermediate points). *For any $x, y, z \in X$ and any minimal geodesic $\gamma: [0, 1] \to X$ from $x$ to $y$, we have*

$$(3.4) \quad d^2(z, \gamma(t)) \leq (1-t)\, d^2(z, x) + t\, d^2(z, y) - (1-t)t\, d^2(x, y) \\ + 4\delta\, (d(z, x) \vee d(z, y)) + 4\delta^2$$

*for all $t \in (0, 1)$.*

*Proof.* Put $w = \gamma(t)$ and note that we have $(x|y)_w = 0$ again. Thus, it follows from the $\delta$-hyperbolicity (2.1) that

$$0 \geq (x|z)_w \wedge (z|y)_w - \delta \\ = \frac{1}{2}\left\{ d(z, w) + (t\, d(x, y) - d(z, x)) \wedge ((1-t)\, d(x, y) - d(z, y)) \right\} - \delta,$$

and hence

$$d^2(z, w) \leq \left\{ (d(z, x) - t\, d(x, y)) \vee (d(z, y) - (1-t)\, d(x, y)) + 2\delta \right\}^2.$$

Now, we claim that

$$(3.5) \quad \left\{ (d(z, x) - t\, d(x, y)) \vee (d(z, y) - (1-t)\, d(x, y)) \right\}^2 \\ \leq (1-t)\, d^2(z, x) + t\, d^2(z, y) - (1-t)t\, d^2(x, y)$$

holds. In fact, the inequality

$$(d(z, x) - t\, d(x, y))^2 \leq (1-t)\, d^2(z, x) + t\, d^2(z, y) - (1-t)t\, d^2(x, y)$$

can be rearranged as

$$t\, d^2(z, x) - 2t\, d(z, x)\, d(x, y) + t\, d^2(x, y) \leq t\, d^2(z, y),$$

which holds true by the triangle inequality. We can similarly show

$$(d(z, y) - (1-t)\, d(x, y))^2 \leq (1-t)\, d^2(z, x) + t\, d^2(z, y) - (1-t)t\, d^2(x, y),$$

thereby we obtain (3.5). Therefore, we deduce that

$$d^2(z, w) \leq (1-t)\, d^2(z, x) + t\, d^2(z, y) - (1-t)t\, d^2(x, y) \\ + 4\delta\, (d(z, x) \vee d(z, y)) + 4\delta^2. \quad \blacksquare$$

Note that, in the case of $\delta = 0$, (3.4) boils down to the CAT(0)-inequality (2.2). For $\delta = 0$, moreover, one can infer (3.4) from (3.3) by the standard subdivision argument (see, for example, (ii) $\Rightarrow$ (iii) of Theorem 1.3.3 in [4]). For $\delta > 0$, however, iterating subdivisions makes the additional term (depending on $\delta$) diverge, thereby we gave a direct argument to prove Lemma 3.2. We also remark that (3.4) is not meaningful for $t$ close to 0 or 1, since then the triangle inequality could give a better estimate.

We are ready to estimate the diameter of barycentric sets $\mathcal{B}(\mu, \varepsilon)$ defined in (3.2). Recall that $x_0 \in X$ is an arbitrary point fixed at the beginning of this section.

**Proposition 3.3** (Diameter of $\mathcal{B}(\mu, \varepsilon)$). *For any $\mu \in \mathcal{P}^1(X)$, $x, y \in X$ and any midpoint $w$ between $x$ and $y$, we have*

$$\int_X \{d^2(w, z) - d^2(x_0, z)\} \mu(dz) \leq \int_X \left\{ \frac{d^2(x, z)}{2} + \frac{d^2(y, z)}{2} - d^2(x_0, z) \right\} \mu(dz)$$
$$- \frac{d^2(x, y)}{4} + 2\delta \{W_1(\delta_x, \mu) + W_1(\delta_y, \mu)\} + 4\delta^2.$$

*In particular, for any $x \in \mathcal{B}(\mu, \varepsilon_1)$ and $y \in \mathcal{B}(\mu, \varepsilon_2)$ with $\varepsilon_1, \varepsilon_2 \geq 0$, we have*

$$(3.6) \qquad d(x, y) \leq \sqrt{8\delta \{W_1(\delta_x, \mu) + W_1(\delta_y, \mu)\} + 16\delta^2 + 2(\varepsilon_1 + \varepsilon_2)}.$$

*Proof.* The first assertion is shown by integrating (3.3) in $z$ with respect to $\mu$. Then, when $x \in \mathcal{B}(\mu, \varepsilon_1)$ and $y \in \mathcal{B}(\mu, \varepsilon_2)$, we find

$$\mathbf{V}_{x_0}(\mu) \leq \int_X \{d^2(w, z) - d^2(x_0, z)\} \mu(dz)$$
$$\leq \mathbf{V}_{x_0}(\mu) + \frac{\varepsilon_1 + \varepsilon_2}{2} - \frac{d^2(x, y)}{4} + 2\delta \{W_1(\delta_x, \mu) + W_1(\delta_y, \mu)\} + 4\delta^2.$$

Therefore, we obtain

$$d^2(x, y) \leq 8\delta \{W_1(\delta_x, \mu) + W_1(\delta_y, \mu)\} + 16\delta^2 + 2(\varepsilon_1 + \varepsilon_2).$$

This completes the proof. ∎

The second assertion (3.6) (with $\varepsilon_2 = 0$) can be regarded as a generalization of the *variance inequality* (see Proposition 4.4 in [32]; the reverse inequality under lower curvature bounds can be found in [25, 33]). Note that, when we are interested in the case of $\varepsilon_1 = \varepsilon_2 = 0$ (the set of barycenters), (3.6) implies $\text{diam}(\mathcal{B}(\mu, 0)) \leq O(\sqrt{\delta})$ as $\delta \to 0$.

**Remark 3.4** (When $\mu \in \mathcal{P}^2(X)$). In the case of $\mu \in \mathcal{P}^2(X)$, instead of $\mathbf{V}_{x_0}(\mu)$ as in (3.1), we can directly consider

$$\inf_{x \in X} W_2^2(\delta_x, \mu) = \inf_{x \in X} \int_X d^2(x, z) \mu(dz),$$

which is called the *variance* of $\mu$. One can simply write down the first assertion of Proposition 3.3 as

$$W_2^2(\delta_w, \mu) \leq \frac{W_2^2(\delta_x, \mu)}{2} + \frac{W_2^2(\delta_y, \mu)}{2} - \frac{d^2(x, y)}{4} + 2\delta \{W_1(\delta_x, \mu) + W_1(\delta_y, \mu)\} + 4\delta^2$$

and, if

$$W_2^2(\delta_x, \mu) \leq \inf_{p \in X} W_2^2(\delta_p, \mu) + \varepsilon_1 \quad \text{and} \quad W_2^2(\delta_y, \mu) \leq \inf_{p \in X} W_2^2(\delta_p, \mu) + \varepsilon_2,$$

then we have (3.6).

## 4. Wasserstein contraction property

We next consider a contraction property in terms of the Wasserstein distance, which in the case of (complete) CAT(0)-spaces means that

$$d(\beta_\mu, \beta_\nu) \le W_1(\mu, \nu)$$

holds for $\mu, \nu \in \mathcal{P}^1(X)$, where $\beta_\mu$ and $\beta_\nu$ are the (unique) barycenters of $\mu$ and $\nu$, respectively (see Theorem 6.3 in [32]). In the current setting of $\delta$-hyperbolic spaces, we shall estimate the distance between points in the barycentric sets.

We begin with a generalization of the *Busemann nonpositive curvature* (Busemann NPC for short). We say that a geodesic space $(Z, d_Z)$ has the Busemann NPC (or that $(Z, d_Z)$ is *convex*) if, for any geodesics $\xi, \zeta: [0, 1] \to Z$ with $\xi(0) = \zeta(0)$, we have that $d_Z(\xi(t), \zeta(t)) \le t \, d_Z(\xi(1), \zeta(1))$ for all $t \in [0, 1]$. Then, by the triangle inequality,

$$d_Z(\xi(t), \zeta(t)) \le (1 - t) \, d_Z(\xi(0), \zeta(0)) + t \, d_Z(\xi(1), \zeta(1))$$

holds for any geodesics $\xi, \zeta: [0, 1] \to Z$ (regardless of whether $\xi(0) = \zeta(0)$ or not). We refer to [10, 18] for further reading.

**Remark 4.1** (Busemann NPC versus CAT(0)). In his celebrated paper [11], Busemann showed that a complete, simply connected Riemannian manifold has the Busemann NPC if and only if its sectional curvature is nonpositive everywhere. Nonetheless, in general, the Busemann NPC is a weaker condition than the CAT(0)-property. On the one hand, it is easily seen that CAT(0)-spaces have the Busemann NPC. On the other hand, every strictly convex Banach space has the Busemann NPC, whereas it is a CAT(0)-space if and only if it is a Hilbert space.

Recall that $(X, d)$ will always denote a geodesic $\delta$-hyperbolic space. It is known that $\delta$-hyperbolic spaces have the Busemann NPC up to an additive constant depending only on $\delta$ (see Section 7.4 in [16]). We give an outline of the proof for completeness.

**Lemma 4.2** (Busemann NPC + $\delta$). *Let $x, y, p, q \in X$. For any pair of minimal geodesics $\gamma: [0, 1] \to X$ from $x$ to $p$ and $\eta: [0, 1] \to X$ from $y$ to $q$, we have*

$$(4.1) \qquad d(\gamma(t), \eta(t)) \le (1 - t) \, d(x, y) + t \, d(p, q) + 8\delta$$

*for all $t \in (0, 1)$.*

*Proof.* Let $\xi: [0, 1] \to X$ and $\zeta: [0, 1] \to X$ be minimal geodesics from $x$ to $q$ and from $p$ to $q$, respectively. Denote by $\triangle xpq$ the triangle formed by (the image of) $\gamma, \xi$ and $\zeta$. We can construct a map $T: \triangle xpq \to Y$ to a tripod $(Y, d_Y)$ with three edges of lengths $(p|q)_x$, $(x|q)_p$ and $(x|p)_q$ from the branching point such that the restrictions $T|_\gamma$, $T|_\xi$ and $T|_\zeta$ are isometric (see Figure 1, where $T(a) = T(b) = T(c)$ is the branching point $O$ of the tripod). We set $\tilde{x} := T(x)$, $\tilde{p} := T(p)$ and $\tilde{q} := T(q)$. Then $T$ is 1-Lipschitz (non-expanding) and

$$d_Y(T(u), T(v)) \ge d(u, v) - 4\delta$$
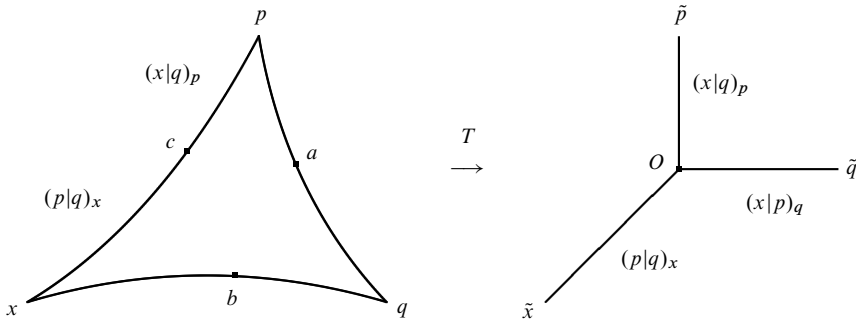
**Figure 1.** A 1-Lipschitz map from a triangle to a tripod

holds for all $u, v \in \triangle xpq$ by the triangle inequality and the tripod lemma (see, for instance, Lemma 2.15 in [34] and Lemma 2.3 in [27]). Together with the Busemann NPC of $(Y, d_Y)$, we deduce that

$$d(\gamma(t), \xi(t)) \leq d_Y\big(T\big(\gamma(t)\big), T(\xi(t))\big) + 4\delta \leq t\, d_Y(\tilde{p}, \tilde{q}) + 4\delta = t\, d(p, q) + 4\delta.$$

We similarly obtain $d(\xi(t), \eta(t)) \leq (1 - t)\, d(x, y) + 4\delta$, and hence the triangle inequality yields (4.1). $\blacksquare$

We remark that, similarly to (3.4) in Lemma 3.2, the inequality (4.1) does not give a meaningful estimate for $t$ close to 0 or 1. One can use a 1-Lipschitz map to a tripod also for showing a variant of the CAT(0)-inequality, whereas then the additional term seems to be necessarily dependent on the size of a triangle (as in (3.4)), since we take the square of the distance.

Now, let $\mathsf{d}$ be the $L^2$-distance function on $X \times X \times \mathbb{R}$, namely

$$\mathsf{d}((x, y, r), (p, q, s)) := \sqrt{d^2(x, p) + d^2(y, q) + (r - s)^2}.$$

The following subset will play a role:

$$A := \{(x, y, r) \in X \times X \times \mathbb{R} \mid d(x, y) \leq r\}.$$

If $(X, d)$ is a CAT(0)-space, then $A$ is a (geodesically) convex set thanks to the Busemann NPC.

We will make use of the nearest point projection to $A$ to prove the Wasserstein contraction property. We remark that $(X \times X \times \mathbb{R}, \mathsf{d})$ is not a Gromov hyperbolic space (it is a CAT(0)-space if so is $(X, d)$), thereby the contraction property of projection maps as in Lemma 7.3.D of [16] does not directly apply.

**Lemma 4.3.** *For* $(x, y, r) \in (X \times X \times [0, \infty)) \setminus A$, *we have*

$$\mathsf{d}((x, y, r), A) = \frac{d(x, y) - r}{\sqrt{3}}.$$

*Proof.* We first show $\mathsf{d}((x, y, r), A) \geq (d(x, y) - r)/\sqrt{3}$. Given $(p, q, s) \in A$, if $s < r$, then $(p, q, r) \in A$ and

$$\mathsf{d}((x, y, r), (p, q, r)) < \mathsf{d}((x, y, r), (p, q, s)).$$

Hence, to show $\mathsf{d}((x, y, r), (p, q, s)) \geq (d(x, y) - r)/\sqrt{3}$, we can assume $s \geq r$ without loss of generality (otherwise we replace $(p, q, s)$ with $(p, q, r)$). Then we have, together with the triangle inequality,

$$\begin{aligned}
\mathsf{d}((x, y, r), (p, q, s)) &= \sqrt{d^2(x, p) + d^2(y, q) + (r - s)^2} \\
&\geq \frac{1}{\sqrt{3}} \{d(x, p) + d(y, q) + (s - r)\} \\
&\geq \frac{1}{\sqrt{3}} \{d(x, y) - d(p, q) + (s - r)\} \geq \frac{1}{\sqrt{3}} \{d(x, y) - r\}.
\end{aligned}$$

This implies $\mathsf{d}((x, y, r), A) \geq (d(x, y) - r)/\sqrt{3}$.

To see the reverse inequality, we consider a minimal geodesic $\gamma : [0, 1] \to X$ from $x$ to $y$ and put

$$(p, q, s) := \left( \gamma\left( \frac{\lambda}{d(x, y)} \right), \gamma\left( 1 - \frac{\lambda}{d(x, y)} \right), r + \lambda \right), \quad \lambda := \frac{d(x, y) - r}{3} \in \left( 0, \frac{d(x, y)}{3} \right].$$

Then $(p, q, s) \in A$, since

$$d(p, q) = d(x, y) - 2\lambda = \frac{d(x, y) + 2r}{3} = s,$$

and we have $\mathsf{d}((x, y, r), (p, q, s)) = \sqrt{3} \, \lambda$. This completes the proof. ∎

The next lemma is an essential step for our contraction result.

**Lemma 4.4.** *For $(x, y, r) \in (X \times X \times [0, \infty)) \setminus A$ such that*

$$(4.2) \qquad\qquad \mathsf{d}((x, y, r), A) \geq \frac{8}{\sqrt{3}} \, \delta,$$

*let $(\tilde{x}, \tilde{y}, \tilde{r}) \in A$ be a point attaining $\mathsf{d}((x, y, r), A)$ given as in the proof of Lemma* 4.3. *Then, for any $(p, q, d(p, q)) \in A$, we have*

$$\begin{aligned}
\mathsf{d}^2 &((\tilde{x}, \tilde{y}, \tilde{r}), (p, q, d(p, q))) \\
&\leq \mathsf{d}^2((x, y, r), (p, q, d(p, q))) - \mathsf{d}^2((x, y, r), A) + 18 D_1 \sqrt{D_1 + \delta} \cdot \sqrt{\delta},
\end{aligned}$$

*where we set*

$$D_1 = D_1(x, y, p, q) := \mathrm{diam}((\{x, y, p, q\}, d)).$$

Observe that, if $\delta = 0$, then the assumption (4.2) is void, and the assertion shows that $\mathsf{d}((\tilde{x}, \tilde{y}, \tilde{r}), (p, q, d(p, q)))$ is strictly smaller than $\mathsf{d}((x, y, r), (p, q, d(p, q)))$.

*Proof.* Put $s := d(p, q)$ for brevity, and let $\gamma : [0, 1] \to X$ and $\eta : [0, 1] \to X$ be minimal geodesics from $\tilde{x}$ to $p$ and from $\tilde{y}$ to $q$, respectively. Note that Lemma 4.2 yields

$$d(\gamma(t), \eta(t)) \leq (1 - t) \, d(\tilde{x}, \tilde{y}) + t \, d(p, q) + 8\delta$$

for any $t \in [0, 1]$. Then we deduce from Lemma 4.3 that

$$\mathsf{d}\big((\gamma(t), \eta(t), (1 - t)\tilde{r} + ts), A\big)$$
$$\leq \frac{1}{\sqrt{3}} \left\{ (1 - t) \, (d(\tilde{x}, \tilde{y}) - \tilde{r}) + t \, (d(p, q) - s) \right\} + \frac{8\delta}{\sqrt{3}} = \frac{8}{\sqrt{3}} \, \delta.$$

Therefore, on the one hand,

$$\mathsf{d}\big((x, y, r), (\gamma(t), \eta(t), (1 - t)\tilde{r} + ts)\big)$$
$$\geq \mathsf{d}((x, y, r), A) - \mathsf{d}((\gamma(t), \eta(t), (1 - t)\tilde{r} + ts), A) \geq \mathsf{d}((x, y, r), A) - \frac{8}{\sqrt{3}} \, \delta.$$

This implies that, since the right-hand side is nonnegative by the assumption (4.2) and $\mathsf{d}((x, y, r), A) \leq D_1 / \sqrt{3}$ by Lemma 4.3 and $r \geq 0$,

$$(4.3) \quad \mathsf{d}^2\big((x, y, r), (\gamma(t), \eta(t), (1 - t)\tilde{r} + ts)\big) \geq \mathsf{d}^2((x, y, r), A) - \frac{16}{3} \, D_1\delta + \frac{64}{3} \, \delta^2.$$

On the other hand, observe from (3.4) that

$$d^2(x, \gamma(t)) \leq (1 - t)d^2(x, \tilde{x}) + td^2(x, p) - (1 - t)td^2(\tilde{x}, p) + 4D_1\delta + 4\delta^2,$$
$$d^2(y, \eta(t)) \leq (1 - t)d^2(y, \tilde{y}) + td^2(y, q) - (1 - t)td^2(\tilde{y}, q) + 4D_1\delta + 4\delta^2$$

(recall that $\tilde{x}$ and $\tilde{y}$ are on a minimal geodesic between $x$ and $y$). Summing up, we obtain

$$\mathsf{d}^2\big((x, y, r), (\gamma(t), \eta(t), (1 - t)\tilde{r} + ts)\big)$$
$$\leq (1 - t)\mathsf{d}^2((x, y, r), (\tilde{x}, \tilde{y}, \tilde{r})) + t\mathsf{d}^2((x, y, r), (p, q, s))$$
$$- (1 - t)t\mathsf{d}^2((\tilde{x}, \tilde{y}, \tilde{r}), (p, q, s)) + 8D_1\delta + 8\delta^2.$$

Combining this with (4.3) shows

$$t\mathsf{d}^2((x, y, r), (p, q, s)) - (1 - t)t\mathsf{d}^2((\tilde{x}, \tilde{y}, \tilde{r}), (p, q, s))$$
$$\geq t\mathsf{d}^2((x, y, r), A) - \frac{40}{3} \, D_1\delta + \frac{40}{3} \, \delta^2.$$

By rearrangement, we find

$$\mathsf{d}^2((\tilde{x}, \tilde{y}, \tilde{r}), (p, q, s)) \leq \mathsf{d}^2((x, y, r), (p, q, s)) - \mathsf{d}^2((x, y, r), A)$$
$$+ t\mathsf{d}^2((\tilde{x}, \tilde{y}, \tilde{r}), (p, q, s)) + \frac{40}{3} \frac{D_1\delta}{t}.$$

Moreover, it follows from $d(x, \tilde{x}) = d(y, \tilde{y}) \leq d(x, y)/3$, $\tilde{r} = d(\tilde{x}, \tilde{y}) < d(x, y)$ and $s = d(p, q)$ that

$$\mathsf{d}^2((\tilde{x}, \tilde{y}, \tilde{r}), (p, q, s)) \leq \frac{16}{9} \, D_1^2 + \frac{16}{9} \, D_1^2 + D_1^2 = \frac{41}{9} \, D_1^2.$$

Finally, letting $t = \sqrt{\delta/(D_1 + \delta)} \in (0, 1)$, we obtain

$$\mathsf{d}^2((\tilde{x}, \tilde{y}, \tilde{r}), (p, q, s))$$

$$\leq \mathsf{d}^2((x, y, r), (p, q, s)) - \mathsf{d}^2((x, y, r), A) + \Big( \frac{41 D_1^2}{9 \sqrt{D_1 + \delta}} + \frac{40}{3} D_1 \sqrt{D_1 + \delta} \Big) \sqrt{\delta}$$

$$\leq \mathsf{d}^2((x, y, r), (p, q, s)) - \mathsf{d}^2((x, y, r), A) + 18 D_1 \sqrt{D_1 + \delta} \cdot \sqrt{\delta}. \qquad \blacksquare$$

We remark that, as is natural from the local flexibility of scale $\leq \delta$, the following contraction property is nontrivial only when $\delta$ is sufficiently small ($\delta \ll D_2$), and then we have $d(x, y) \leq W_1(\mu, \nu) + O(\delta^{1/4})$ for $\varepsilon_1 = \varepsilon_2 = 0$.

**Theorem 4.5** (Wasserstein contraction). *For any $\mu, \nu \in \mathcal{P}^1(X)$, $x \in \mathcal{B}(\mu, \varepsilon_1)$ and $y \in \mathcal{B}(\nu, \varepsilon_2)$ with $\varepsilon_1, \varepsilon_2 \geq 0$, we have*

$$d(x, y) \leq W_1(\mu, \nu) + 8\delta \vee \sqrt{54 D_2 \sqrt{D_2 + \delta} \sqrt{\delta} + 3(\varepsilon_1 + \varepsilon_2)},$$

*where*

$$D_2 = D_2(x, y, \mu, \nu) := \mathrm{diam}\big( (\{x, y\} \cup \mathrm{supp}\, \mu \cup \mathrm{supp}\, \nu, d) \big).$$

*Proof.* For arbitrary $\alpha > 0$, let $\pi \in \mathcal{P}(X \times X)$ be a coupling of $(\mu, \nu)$ with

$$r := \int_{X \times X} d(p, q) \, \pi(dp \, dq) \leq W_1(\mu, \nu) + \alpha.$$

We define a map

$$\Phi : X \times X \ni (p, q) \longmapsto (p, q, d(p, q)) \in X \times X \times \mathbb{R},$$

and put $\Pi := \Phi_* \pi$ (the push-forward of $\pi$ by $\Phi$). Note that $\mathrm{supp}\, \Pi \subset \Phi(\mathrm{supp}\, \pi) \subset A$. Since $x \in \mathcal{B}(\mu, \varepsilon_1)$ and $y \in \mathcal{B}(\nu, \varepsilon_2)$, together with the choice of $r$, we find

$$\int_{X \times X \times \mathbb{R}} \{ \mathsf{d}^2((x, y, r), (p, q, s)) - d^2(x_0, p) - d^2(x_0, q) - s^2 \} \, \Pi(dp \, dq \, ds)$$

$$= \int_X \{ d^2(x, p) - d^2(x_0, p) \} \, \mu(dp) + \int_X \{ d^2(y, q) - d^2(x_0, q) \} \, \nu(dq)$$

$$+ \int_{X \times X} \{ (r - d(p, q))^2 - d^2(p, q) \} \, \pi(dp \, dq)$$

$$\leq \inf_{\bar{x}, \bar{y} \in X, \bar{r} \in \mathbb{R}} \int_{X \times X \times \mathbb{R}} \{ \mathsf{d}^2((\bar{x}, \bar{y}, \bar{r}), (p, q, s)) - d^2(x_0, p) - d^2(x_0, q) - s^2 \} \, \Pi(dp \, dq \, ds)$$

$$+ \varepsilon_1 + \varepsilon_2.$$

That is to say, $(x, y, r)$ lives in the barycentric set $\mathcal{B}(\Pi, \varepsilon_1 + \varepsilon_2)$ in $(X \times X \times \mathbb{R}, \mathsf{d})$.

Then we deduce from Lemma 4.4 that $(x, y, r)$ necessarily satisfies

$$\mathsf{d}^2((x, y, r), A) \leq \frac{64}{3} \delta^2 \vee \big( 18 D_2 \sqrt{D_2 + \delta} \sqrt{\delta} + \varepsilon_1 + \varepsilon_2 \big).$$

Indeed, if not, then (4.2) is fulfilled and, for any $(p, q) \in \operatorname{supp} \pi$, $(\tilde{x}, \tilde{y}, \tilde{r}) \in A$ (given as in the proof of Lemma 4.3) satisfies

$$\mathsf{d}^2((\tilde{x}, \tilde{y}, \tilde{r}), (p, q, d(p, q)))$$
$$\leq \mathsf{d}^2((x, y, r), (p, q, d(p, q))) - \mathsf{d}^2((x, y, r), A) + 18 D_2 \sqrt{D_2 + \delta} \sqrt{\delta}$$
$$< \mathsf{d}^2((x, y, r), (p, q, d(p, q))) - (\varepsilon_1 + \varepsilon_2).$$

Integrating in $(p, q)$ with respect to $\pi$, we find a contradiction to $(x, y, r) \in \mathcal{B}(\Pi, \varepsilon_1 + \varepsilon_2)$.

Hence, it follows from Lemma 4.3 that

$$d(x, y) - W_1(\mu, \nu) \leq d(x, y) - r + \alpha$$
$$\leq 8\delta \vee \sqrt{54 D_2 \sqrt{D_2 + \delta} \sqrt{\delta} + 3(\varepsilon_1 + \varepsilon_2)} + \alpha.$$

Letting $\alpha \to 0$ completes the proof. ∎

## 5. Deterministic approximations of barycenters

We next discuss an approximation of barycenters of uniform distributions on finite points by the gradient flow method. Given a function $f : X \to \mathbb{R}$, we define the *proximal operator* (also called the *resolvent operator*) by

$$(5.1) \qquad \mathsf{J}_\tau^f(x) := \arg\min_{y \in X} \left\{ f(y) + \frac{d^2(x, y)}{2\tau} \right\}$$

for $x \in X$ and $\tau > 0$ (that is, $y \in \mathsf{J}_\tau^f(x)$ if $y$ attains the above minimum). Roughly speaking, an element in $\mathsf{J}_\tau^f(x)$ can be regarded as an approximation of a point on the gradient curve of $f$ at time $\tau$ from $x$. Thus the iteration of the proximal operator can be regarded as *discrete-time gradient flow* for $f$, which is expected to lead us to a minimizer of $f$. We refer to [27] for some contraction properties of discrete-time gradient flows for geodesically convex functions on Gromov hyperbolic spaces.

We will apply the proximal operator only to squared distance functions $f = d^2(z, \cdot)$. In this case, each $y \in \mathsf{J}_\tau^f(x)$ is explicitly given as $y = \gamma((2\tau + 1)^{-1})$ for some minimal geodesic $\gamma : [0, 1] \to X$ from $z$ to $x$.

The next lemma (corresponding to Lemma 3.2 in [3] and Lemma 4.6(I) in [28]) provides a key estimate. Precisely, those in [3, 28] are concerned with $y \in \mathsf{J}_\tau^f(x)$ for a convex function $f$, and we shall generalize it to Gromov hyperbolic spaces for the squared distance function $f = d^2(z, \cdot)$.

**Lemma 5.1** (Key estimate). *For any $\tau > 0$, $w, x, z \in X$ and $y = \gamma((2\tau + 1)^{-1})$ on a minimal geodesic $\gamma : [0, 1] \to X$ from $z$ to $x$, we have*

$$(5.2) \qquad d^2(w, y) \leq d^2(w, x) - 2\tau \{d^2(z, y) - d^2(z, w)\} + \Theta \tau \delta,$$

*where*

$$\Theta := \{8 d(z, w) + 8\delta\} \vee \left[ \{4 d(w, y) + 8\tau d(z, y)\} \left( \frac{1}{\tau} \wedge \frac{2 d(z, y)}{\delta} \right) \right].$$

*Proof.* First of all, it follows from $y \in \mathsf{J}_\tau^f(x)$ with $f = d^2(z, \cdot)$ that

$$d^2(x, y) \leq d^2(x, w) - 2\tau \{d^2(z, y) - d^2(z, w)\}.$$

Hence, if $d(w, y) \leq d(x, y)$, then (5.2) holds (without the additional term $\Theta \tau \delta$).

We assume $d(w, y) > d(x, y)$ in the sequel. Note that the $\delta$-hyperbolicity implies

$$0 = (x|z)_y \geq (x|w)_y \wedge (z|w)_y - \delta.$$

If $(z|w)_y \leq \delta$, then we find

$$\{d(z, w) + 2\delta\}^2 \geq \{d(z, y) + d(w, y)\}^2 \geq d^2(z, y) + 2d(z, y)d(w, y).$$

Combining this with the triangle inequality $|d(w, y) - d(x, y)| \leq d(w, x)$ and $d(x, y) = 2\tau d(z, y)$, we obtain

$$\begin{aligned}
d^2(w, y) &\leq d^2(w, x) + 2d(x, y)d(w, y) - d^2(x, y) \\
&\leq d^2(w, x) + 4\tau d(z, y)d(w, y) \\
&\leq d^2(w, x) + 2\tau \{d^2(z, w) - d^2(z, y)\} + 2\tau \{4\delta d(z, w) + 4\delta^2\}.
\end{aligned}$$

In the other case of $(x|w)_y \leq \delta$, together with the triangle inequality, we have

$$d(x, y) + d(w, y) - d(w, x) \leq (2\delta) \wedge (2d(x, y)).$$

Thus we obtain

$$d(w, x) \geq d(w, y) + d(x, y) - 2(\delta \wedge d(x, y)),$$

and observe that the right-hand side is positive by the hypothesis $d(w, y) > d(x, y)$. Therefore, we deduce that

$$d^2(w, x) \geq d^2(w, y) + d^2(x, y) + 2d(w, y)d(x, y) - 4\{d(w, y) + d(x, y)\}(\delta \wedge d(x, y)).$$

Substituting $d(x, y) = 2\tau d(z, y)$ and $d(w, y) \geq d(z, y) - d(z, w)$ yields

$$\begin{aligned}
d^2(w, x) \geq{}& d^2(w, y) + 4\tau(\tau + 1)d^2(z, y) - 4\tau d(z, w)d(z, y) \\
&- 4\{d(w, y) + d(x, y)\}(\delta \wedge 2\tau d(z, y)).
\end{aligned}$$

Then we apply the elementary inequality $2(\tau + 1)a^2 - 2ab \geq a^2 - b^2$ with $a = d(z, y)$ and $b = d(z, w)$ to see

$$\begin{aligned}
d^2(w, x) \geq{}& d^2(w, y) + 2\tau \{d^2(z, y) - d^2(z, w)\} \\
&- 4\{d(w, y) + d(x, y)\}(\delta \wedge 2\tau d(z, y)).
\end{aligned}$$

This completes the proof. ∎

In the CAT(0)-case (see [3, 28]), we have

$$d^2(w, y) \leq d^2(w, x) - 2\tau \{d^2(z, y) - d^2(z, w)\}$$

without the additional term $\Theta\tau\delta$. Note that $\Theta\tau\delta$ in (5.2) tends to 0 not only as $\delta \to 0$ but also as $\tau \to 0$. This is the natural behavior, since $y$ tends to $x$ as $\tau \to 0$.

We also remark that, in the proof of Lemma 4.6(I) in [28], the Riemannian nature of CAT(0)-spaces plays an essential role. Precisely, the inequality (4.2) in [28] is a consequence of a Riemannian property called the *commutativity* as in equations (1.2) and (3.1) in [29].

By using Lemma 5.1, we establish the following deterministic approximation of barycenters by the iterative application of the proximal operator in the spirit of [17, 23] (so-called the *no dice* theorem). We also refer to Theorem 3.4 in [3] and Theorem 5.5 in [28] for generalizations to the sum of convex functions.

**Theorem 5.2** (Deterministic approximation). *Fix a finite sequence $(z_i)_{i=1}^n$ in $X$, put $f_i := d^2(z_i, \cdot)$, and let $p \in X$ be a minimizer of the function $f := \sum_{i=1}^n f_i$. Given $\tau > 0$ and an arbitrary initial point $y_0 \in X$, we recursively choose*

$$y_{kn+i} \in \mathsf{J}_\tau^{f_i}(y_{kn+i-1}) \quad \text{for } k \geq 0, 1 \leq i \leq n,$$

*and assume that $p$, $(z_i)$ and $(y_{kn+i})$ are all included in a bounded set $\Omega \subset X$. Then, for any $\varepsilon > 0$, there exists some $k_0 < d^2(p, y_0)/(2\tau\varepsilon)$ such that*

$$(5.3) \qquad f(y_{k_0 n}) \leq f(p) + \frac{n\Theta_\Omega \delta}{2} + 2n(n+1) D_\Omega^2 \tau + \varepsilon,$$

*where we set*

$$D_\Omega := \operatorname{diam}(\Omega) \quad \text{and} \quad \Theta_\Omega := (8D_\Omega + 8\delta) \vee \left[(4 + 8\tau) D_\Omega \left(\frac{1}{\tau} \wedge \frac{2D_\Omega}{\delta}\right)\right].$$

*Moreover, we have*

$$(5.4) \qquad d(p, y_{k_0 n}) \leq \sqrt{(16D_\Omega + \Theta_\Omega)\delta + 16\delta^2 + 4(n+1) D_\Omega^2 \tau + 2\varepsilon/n}.$$

We remark that, in a CAT(0)-space, we can employ as $\Omega$ a ball including $(z_i)_{i=1}^n$ and $y_0$. This is because balls are convex by the CAT(0)-inequality (or the Busemann NPC). In a $\delta$-hyperbolic space, however, balls are not necessarily convex, and it is unclear to the author how to control (the sum of) the additional terms in (3.4) (or (4.1)) during the recursive scheme $y_{kn+i} \in \mathsf{J}_\tau^{f_i}(y_{kn+i-1})$.

*Proof.* It follows from Lemma 5.1 that

$$d^2(p, y_{kn+i}) \leq d^2(p, y_{kn+i-1}) - 2\tau \{d^2(z_i, y_{kn+i}) - d^2(z_i, p)\} + \Theta_\Omega \tau\delta.$$

Summing up for $1 \leq i \leq n$, we have

$$d^2(p, y_{(k+1)n}) \leq d^2(p, y_{kn}) - 2\tau \{f(y_{kn}) - f(p)\}$$

$$(5.5) \qquad\qquad + 2\tau \left\{f(y_{kn}) - \sum_{i=1}^n d^2(z_i, y_{kn+i})\right\} + n\Theta_\Omega \tau\delta.$$

Concerning the third term in the right-hand side, we infer from the triangle inequality that

$$\sum_{i=1}^{n}\{d^2(z_i, y_{kn}) - d^2(z_i, y_{kn+i})\} = \sum_{i=1}^{n}\sum_{l=1}^{i}\{d^2(z_i, y_{kn+l-1}) - d^2(z_i, y_{kn+l})\}$$

$$\leq 2D_\Omega \sum_{i=1}^{n}\sum_{l=1}^{i} d(y_{kn+l-1}, y_{kn+l}).$$

Then, by the choice of $y_{kn+l} \in \mathsf{J}_\tau^{f_l}(y_{kn+l-1})$ with $f_l = d^2(z_l, \cdot)$, we find

$$(5.6) \qquad d(y_{kn+l-1}, y_{kn+l}) = \frac{2\tau}{2\tau + 1} d(z_l, y_{kn+l-1}) \leq 2D_\Omega \tau.$$

Hence, we obtain

$$\sum_{i=1}^{n}\{d^2(z_i, y_{kn}) - d^2(z_i, y_{kn+i})\} \leq 4D_\Omega^2 \tau \sum_{i=1}^{n} i = 2n(n + 1)D_\Omega^2 \tau.$$

Plugging this into (5.5) yields

$$(5.7) \quad d^2(p, y_{(k+1)n}) \leq d^2(p, y_{kn}) - 2\tau\left\{ f(y_{kn}) - f(p) - \frac{n\Theta_\Omega\delta}{2} - 2n(n + 1)D_\Omega^2\tau\right\}.$$

It immediately follows from (5.7) that

$$f(y_{k_0n}) - f(p) - \frac{n\Theta_\Omega\delta}{2} - 2n(n + 1)D_\Omega^2\tau \leq \varepsilon$$

necessarily holds for some $k_0 < d^2(p, y_0)/(2\tau\varepsilon)$. Indeed, otherwise we have

$$d^2(p, y_{\bar{k}n}) < d^2(p, y_0) - 2\tau\bar{k}\varepsilon \leq 0$$

with $\bar{k}$ the minimum integer not smaller than $d^2(p, y_0)/(2\tau\varepsilon)$, a contradiction.

Finally, the second assertion (5.4) is a consequence of Proposition 3.3. Putting $\mu = n^{-1}\sum_{i=1}^{n}\delta_{z_i}$, we deduce from (5.3) that

$$W_2^2(\delta_{y_{k_0n}}, \mu) \leq W_2^2(\delta_p, \mu) + \frac{\Theta_\Omega\delta}{2} + 2(n + 1)D_\Omega^2\tau + \frac{\varepsilon}{n}.$$

Hence, (3.6) (with $\varepsilon_1 = 0$) yields

$$d(p, y_{k_0n}) \leq \sqrt{8\delta\{W_1(\delta_p, \mu) + W_1(\delta_{y_{k_0n}}, \mu)\} + 16\delta^2 + \Theta_\Omega\delta + 4(n + 1)D_\Omega^2\tau + 2\varepsilon/n}$$

$$\leq \sqrt{16D_\Omega\delta + 16\delta^2 + \Theta_\Omega\delta + 4(n + 1)D_\Omega^2\tau + 2\varepsilon/n}. \qquad \blacksquare$$

Thanks to (5.3), up to $d^2(p, y_0)/(2\tau\varepsilon)$ iterations, $f(y_{kn})$ nearly achieves $\min_X f$ and $y_{kn}$ passes close to $p$ as in (5.4). We remark that the sublinear rate $\varepsilon < d^2(p, y_0)/(2\tau k_0)$ (following from $k_0 < d^2(p, y_0)/(2\tau\varepsilon)$) could be compared with Proposition 5.7 in [28].

Note also that we have an effective estimate on the resolvent operator only when $\tau$ is larger than $\delta$ (so-called "giant steps"; see [27] for a further discussion), thereby we did not consider $(\tau_k)_{k \geq 0}$ converging to 0 (compare this with Theorem 5.5 in [28]). If we assume $\delta \leq D_\Omega/2$ and take $\tau = \sqrt{\delta/D_\Omega}$, then we have $\Theta_\Omega = (4\sqrt{D_\Omega/\delta} + 8)D_\Omega$, and (5.4) shows

$$d(p, y_{k_0 n}) \leq \sqrt{\frac{2\varepsilon}{n}} + O(\delta^{1/4}).$$

## 6. A law of large numbers

We next consider a law of large numbers in our setting. Our formulation follows Sturm's Theorem 4.7 in [32] for CAT(0)-spaces. We refer to Theorem 6.7 in [28] and Theorem 3 in [38] for some generalizations to other (upper and lower) curvature bounds.

**Theorem 6.1** (Law of large numbers). *Let $(Z_i)_{i \geq 1}$ be a sequence of independent, identically distributed random variables on a probability space taking values in $X$ with distribution $\mu \in \mathcal{P}(X)$, and take $p \in \mathcal{B}(\mu, 0)$. Given $\tau > 0$ and an arbitrary initial point $S_0 \in X$, we define a sequence $(S_k)_{k \geq 0}$ in $X$ recursively by*

$$S_{k+1} \in \mathsf{J}_\tau^{f_k}(S_k), \qquad f_k := d^2(Z_{k+1}, \cdot).$$

*Assume that $p$, $\operatorname{supp}\mu$ and $(S_k)_{k \geq 0}$ are all included in a bounded set $\Omega \subset X$. Then, for any $\varepsilon > 0$, we have*

(6.1) $$\mathbb{E}[d^2(p, S_{k_0})] \leq 8 D_\Omega^2 \tau + (\Theta_\Omega + 16 D_\Omega + 16\delta)\delta + \varepsilon$$

*for some $k_0 < d^2(p, S_0)/(\tau\varepsilon)$, where $D_\Omega := \operatorname{diam}(\Omega)$ and $\Theta_\Omega$ is defined as in Theorem 5.2.*

*Proof.* We can apply a calculation similar to the proof of Theorem 5.2. It follows from Lemma 5.1 that

$$\begin{aligned}
d^2(p, S_{k+1}) &\leq d^2(p, S_k) - 2\tau\{d^2(Z_{k+1}, S_{k+1}) - d^2(Z_{k+1}, p)\} + \Theta_\Omega\tau\delta \\
&= d^2(p, S_k) - 2\tau\{d^2(Z_{k+1}, S_k) - d^2(Z_{k+1}, p)\} \\
&\quad + 2\tau\{d^2(Z_{k+1}, S_k) - d^2(Z_{k+1}, S_{k+1})\} + \Theta_\Omega\tau\delta \\
&\leq d^2(p, S_k) - 2\tau\{d^2(Z_{k+1}, S_k) - d^2(Z_{k+1}, p)\} + 8 D_\Omega^2\tau^2 + \Theta_\Omega\tau\delta,
\end{aligned}$$

where we used

$$d^2(Z_{k+1}, S_k) - d^2(Z_{k+1}, S_{k+1}) \leq 2 D_\Omega\, d(S_k, S_{k+1}) \leq 4 D_\Omega^2\tau$$

in the latter inequality (recall also (5.6)). Taking the expectations in $Z_{k+1}$ conditioned on $\mathcal{F}_k := \{Z_1, \dots, Z_k\}$ and applying the variance inequality (3.6) (with $x = p$, $y = S_k$, $\varepsilon_1 = 0$), we obtain

$$\begin{aligned}
\mathbb{E}[d^2(p, S_{k+1}) \mid \mathcal{F}_k] \\
&\leq d^2(p, S_k) - 2\tau\mathbb{E}[d^2(Z_{k+1}, S_k) - d^2(Z_{k+1}, p)] + 8 D_\Omega^2\tau^2 + \Theta_\Omega\delta\tau \\
&\leq d^2(p, S_k) - \tau\{d^2(p, S_k) - 16 D_\Omega\delta - 16\delta^2\} + 8 D_\Omega^2\tau^2 + \Theta_\Omega\delta\tau \\
&= (1 - \tau)\, d^2(p, S_k) + 8 D_\Omega^2\tau^2 + (\Theta_\Omega + 16 D_\Omega + 16\delta)\delta\tau.
\end{aligned}$$

Taking the expectations once again, we arrive at

$$(6.2) \quad \mathbb{E}[d^2(p, S_{k+1})] \leq (1 - \tau) \mathbb{E}[d^2(p, S_k)] + 8 D_\Omega^2 \tau^2 + (\Theta_\Omega + 16 D_\Omega + 16\delta)\delta\tau.$$

In the same way as in the proof of Theorem 5.2, we infer from (6.2) that

$$\mathbb{E}[d^2(p, S_{k_0})] \leq 8 D_\Omega^2 \tau + (\Theta_\Omega + 16 D_\Omega + 16\delta)\delta + \varepsilon$$

necessarily holds for some $k_0 < d^2(p, S_0)/(\tau\varepsilon)$. This completes the proof. ∎

When we assume $\delta \leq D_\Omega/2$ and choose $\tau = \sqrt{\delta/D_\Omega}$ as in the discussion after Theorem 5.2, (6.1) yields

$$\mathbb{E}[d^2(p, S_{k_0})] \leq \varepsilon + O(\sqrt{\delta}).$$

An advantage of the above recursive choice of $(S_k)_{k \geq 0}$ is that $S_{k+1}$ is concretely given as a point on a geodesic between $S_k$ and $Z_{k+1}$ without any knowledge about the construction of barycenters of probability measures (though minimal geodesics are not unique in Gromov hyperbolic spaces).

Employing "empirical means" instead of $(S_k)_{k \geq 0}$, one can also show the following version of law of large numbers (see Proposition 6.6 in [32]).

**Proposition 6.2** (Empirical law of large numbers). *Let $(X, d)$ be complete and separable, let $(Z_i)_{i \geq 1}$ be a sequence of independent, identically distributed random variables on a probability space taking values in $X$ with distribution $\mu \in \mathcal{P}(X)$ of bounded support, and take $p \in \mathcal{B}(\mu, 0)$. Then,*

$$\sigma_k \in \mathcal{B}\left(\frac{1}{k} \sum_{i=1}^k \delta_{Z_i}, 0\right)$$

*satisfies*

$$\limsup_{k \to \infty} d(p, \sigma_k) \leq 8\delta \vee \sqrt{54 D \sqrt{D + \delta}\sqrt{\delta}}$$

*almost surely, where we set $D := 3 \operatorname{diam}(\operatorname{supp} \mu)$.*

*Proof.* It follows from Varadarajan's theorem (see Theorem 11.4.1 in [14]) that

$$\nu_k := \frac{1}{k} \sum_{i=1}^k \delta_{Z_i}$$

weakly converges to $\mu$ almost surely, thereby $W_1(\mu, \nu_k) \to 0$ (see Theorem 7.12 in [35]). Observe that $d(p, \operatorname{supp} \mu) \leq \operatorname{diam}(\operatorname{supp} \mu)$ necessarily holds, and similarly,

$$d(\sigma_k, \operatorname{supp} \mu) \leq d(\sigma_k, \operatorname{supp} \nu_k) \leq \operatorname{diam}(\operatorname{supp} \nu_k) \leq \operatorname{diam}(\operatorname{supp} \mu).$$

Hence, we have $\operatorname{diam}((\{p, \sigma_k\} \cup \operatorname{supp} \mu, d)) \leq D$ and Theorem 4.5 (with $\varepsilon_1 = \varepsilon_2 = 0$) yields

$$\limsup_{k \to \infty} d(p, \sigma_k) \leq 8\delta \vee \sqrt{54 D \sqrt{D + \delta}\sqrt{\delta}}$$

by the choices of $p$ and $\sigma_k$. ∎

We finally discuss two possible directions of further research.

**Remark 6.3** (Further problems). (a) We studied in [27] discrete-time gradient flows for geodesically convex functions (namely, they are convex along geodesics). However, despite the negative curvature nature, distance functions on Gromov hyperbolic spaces are not geodesically convex due to the local flexibility. Therefore, it is an intriguing problem to introduce an appropriate notion of "roughly convex functions" on Gromov hyperbolic spaces, including the distance function $d(z, \cdot)$ or its square.

(b) Compared with the contraction estimates in [27] directly akin to trees, the results in the present article are generalizations from CAT(0)-spaces to $\delta$-hyperbolic spaces. Therefore, on the one hand, it may be possible to improve our estimates (for example, the order of $\delta$) via an analysis closer to the case of trees. Specifically, it is desirable that we can reduce the dependence on $D_2$ and $D_\Omega$ in Theorems 4.5, 5.2 and 6.1. On the other hand, there seems a room for further generalizations to metric spaces satisfying the CAT(0)-inequality with small perturbations in some way (cf. Lemma 3.2).

# References

[1] Afsari, B.: Riemannian $L^p$ center of mass: existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** (2011), no. 2, 655–673. Zbl 1220.53040 MR 2736346

[2] Bačák, M.: The proximal point algorithm in metric spaces. *Israel J. Math.* **194** (2013), no. 2, 689–701. Zbl 1278.49039 MR 3047087

[3] Bačák, M.: Computing medians and means in Hadamard spaces. *SIAM J. Optim.* **24** (2014), no. 3, 1542–1566. Zbl 1306.49046 MR 3264572

[4] Bačák, M.: *Convex analysis and optimization in Hadamard spaces*. De Gruyter Ser. Nonlinear Anal. Appl. 22, De Gruyter, Berlin, 2014. Zbl 1299.90001 MR 3241330

[5] Benua, I. and Kén, Z. F.: Central limit theorem on hyperbolic groups. *Izv. Ross. Akad. Nauk Ser. Mat.* **80** (2016), no. 1, 5–26; translation in *Izv. Math.* **80** (2016), no. 1, 3–23. Zbl 1376.20050 MR 3462675

[6] Billera, L. J., Holmes, S. P. and Vogtmann, K.: Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** (2001), no. 4, 733–767. Zbl 0995.92035 MR 1867931

[7] Björklund, M.: Central limit theorems for Gromov hyperbolic groups. *J. Theoret. Probab.* **23** (2010), no. 3, 871–887. Zbl 1217.60019 MR 2679960

[8] Bonk, M. and Schramm, O.: Embeddings of Gromov hyperbolic spaces. *Geom. Funct. Anal.* **10** (2000), no. 2, 266–306. Zbl 0972.53021 MR 1771428

[9] Bowditch, B. H.: *A course on geometric group theory*. MSJ Mem. 16, Mathematical Society of Japan, Tokyo, 2006. Zbl 1103.20037 MR 2243589

[10] Bridson, M. R. and Haefliger, A.: *Metric spaces of non-positive curvature*. Grundlehren Math. Wiss. 319, Springer, Berlin, 1999. Zbl 0988.53001 MR 1744486

[11] Busemann, H.: Spaces with non-positive curvature. *Acta Math.* **80** (1948), 259–310. Zbl 0038.10005 MR 0029531

[12] Chalopin, J., Chepoi, V., Hirai, H. and Osajda, D.: Weakly modular graphs and nonpositive curvature. *Mem. Amer. Math. Soc.* **268** (2020), no. 1309, vi+159 pp. Zbl 1475.05001 MR 4223822

[13] Das, T., Simmons, D. and Urbański, M.: *Geometry and dynamics in Gromov hyperbolic metric spaces*. Math. Surveys Monogr. 218, American Mathematical Society, Providence, RI, 2017. Zbl 1515.20008 MR 3558533

[14] Dudley, R. M.: *Real analysis and probability*. Cambridge Stud. Adv. Math. 74, Cambridge University Press, Cambridge, 2002. Zbl 1023.60001 MR 1932358

[15] Es-Sahib, A. and Heinich, H.: Barycentre canonique pour un espace métrique à courbure négative. In *Séminaire de Probabilités, XXXIII*, pp. 355–370. Lecture Notes in Math. 1709, Springer, Berlin, 1999. Zbl 0952.60010 MR 1768010

[16] Gromov, M.: Hyperbolic groups. In *Essays in group theory*, pp. 75–263. Math. Sci. Res. Inst. Publ. 8, Springer, New York, 1987. Zbl 0634.20015 MR 0919829

[17] Holbrook, J.: No dice: a deterministic approach to the Cartan centroid. *J. Ramanujan Math. Soc.* **27** (2012), no. 4, 509–521. Zbl 1325.47039 MR 3027448

[18] Jost, J.: *Nonpositive curvature: geometric and analytic aspects*. Lectures in Mathematics ETH Zürich, Birkhäuser, Basel, 1997. Zbl 0896.53002 MR 1451625

[19] Karlsson, A. and Ledrappier, F.: On laws of large numbers for random walks. *Ann. Probab.* **34** (2006), no. 5, 1693–1706. Zbl 1111.60005 MR 2271477

[20] Karlsson, A. and Noskov, G. A.: The Hilbert metric and Gromov hyperbolicity. *Enseign. Math. (2)* **48** (2002), no. 1-2, 73–89. Zbl 1046.53026 MR 1923418

[21] Kingman, J. F. C.: The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc. Ser. B* **30** (1968), 499–510. Zbl 0182.22802 MR 0254907

[22] Kingman, J. F. C.: Subadditive ergodic theory. *Ann. Probability* **1** (1973), 883–909. Zbl 0311.60018 MR 0356192

[23] Lim, Y. and Pálfia, M.: Weighted deterministic walks for the squares mean on Hadamard spaces. *Bull. Lond. Math. Soc.* **46** (2014), no. 3, 561–570. Zbl 1291.53050 MR 3210712

[24] Navas, A.: An $L^1$ ergodic theorem with values in a non-positively curved space via a canonical barycenter map. *Ergodic Theory Dynam. Systems* **33** (2013), no. 2, 609–623. Zbl 1277.37014 MR 3035300

[25] Ohta, S.: Barycenters in Alexandrov spaces of curvature bounded below. *Adv. Geom.* **12** (2012), no. 4, 571–587. Zbl 1276.53073 MR 3005101

[26] Ohta, S.: *Comparison Finsler geometry*. Springer Monogr. Math., Springer, Cham, 2021. Zbl 1487.53003 MR 4386101

[27] Ohta, S.: Discrete-time gradient flows in Gromov hyperbolic spaces. To appear in *Israel J. Math.*

[28] Ohta, S. and Pálfia, M.: Discrete-time gradient flows and law of large numbers in Alexandrov spaces. *Calc. Var. Partial Differential Equations* **54** (2015), no. 2, 1591–1610. Zbl 1335.51014 MR 3396425

[29] Ohta, S. and Pálfia, M.: Gradient flows and a Trotter–Kato formula of semi-convex functions on CAT(1)-spaces. *Amer. J. Math.* **139** (2017), no. 4, 937–965. Zbl 1372.49027 MR 3689321

[30] Ohta, S. and Sturm, K.-T.: Non-contraction of heat flow on Minkowski spaces. *Arch. Ration. Mech. Anal.* **204** (2012), no. 3, 917–944. Zbl 1257.53098 MR 2917125

[31] Roe, J.: *Lectures on coarse geometry*. Univ. Lecture Ser. 31, American Mathematical Society, Providence, RI, 2003. Zbl 1042.53027 MR 2007488

[32] Sturm, K.-T.: Probability measures on metric spaces of nonpositive curvature. In *Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002)*, pp. 357–390. Contemp. Math. 338, American Mathematical Society, Providence, RI, 2003. Zbl 1040.60002 MR 2039961

[33] Sturm, K.-T.: A semigroup approach to harmonic maps. *Potential Anal.* **23** (2005), no. 3, 225–277. Zbl 1078.58010 MR 2140797

[34] Väisälä, J.: Gromov hyperbolic spaces. *Expo. Math.* **23** (2005), no. 3, 187–231. Zbl 1087.53039 MR 2164775

[35] Villani, C.: *Topics in optimal transportation*. Grad. Stud. Math. 58, American Mathematical Society, Providence, RI, 2003. Zbl 1106.90001 MR 1964483

[36] Yokota, T.: Convex functions and barycenter on CAT(1)-spaces of small radii. *J. Math. Soc. Japan* **68** (2016), no. 3, 1297–1323. Zbl 1351.53057 MR 3523548

[37] Yokota, T.: Convex functions and *p*-barycenter on CAT(1)-spaces of small radii. *Tsukuba J. Math.* **41** (2017), no. 1, 43–80. Zbl 1378.53056 MR 3705774

[38] Yokota, T.: Law of large numbers in CAT(1)-spaces of small radii. *Calc. Var. Partial Differential Equations* **57** (2018), no. 2, article no. 35, 28 pp. Zbl 1395.53048 MR 3763113

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