



Thibaut Le Gouic · Quentin Paris · Philippe Rigollet · Austin J. Stromme

# Fast convergence of empirical barycenters in Alexandrov spaces and the Wasserstein space

Received February 14, 2020

**Abstract.** This work establishes fast rates of convergence for empirical barycenters over a large class of geodesic spaces with curvature bounds in the sense of Alexandrov. More specifically, we show that parametric rates of convergence are achievable under natural conditions that characterize the bi-extendibility of geodesics emanating from a barycenter. These results largely advance the state-of-the-art on the subject both in terms of rates of convergence and the variety of spaces covered. In particular, our results apply to infinite-dimensional spaces such as the 2-Wasserstein space, where bi-extendibility of geodesics translates into regularity of Kantorovich potentials.

**Keywords.** Wasserstein barycenters, Alexandrov spaces, rates of convergence

## 1. Introduction

The notion of *average* is paramount across all of statistics. It is fundamental not only in theory but also in practice as it arises in nearly every known estimation method: the method of moments, empirical risk minimization, and more. The law of large numbers, sometimes dubbed the fundamental law of statistics, ensures that the average of independent and identically distributed random variables converges to their common expected value. While such results guarantee asymptotic validity of the average, modern statistics and machine learning focuses on non-asymptotic performance guarantees that hold for every sample size. The goal of this paper is to provide such non-asymptotic guarantees for a generalized notion of average over metric spaces with the Wasserstein space as a prime example.

---

Thibaut Le Gouic: Mathematics Department, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139-4307, USA; [thibaut.le\\_gouic@math.cnrs.fr](mailto:thibaut.le_gouic@math.cnrs.fr)

Quentin Paris: Computer Science Faculty, Higher School of Economics, Moscow, Russia; [qparis@hse.ru](mailto:qparis@hse.ru)

Philippe Rigollet: Mathematics Department, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139-4307, USA; [rigollet@mit.edu](mailto:rigollet@mit.edu)

Austin J. Stromme: Mathematics Department, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139-4307, USA; [astromme@mit.edu](mailto:astromme@mit.edu)

*Mathematics Subject Classification (2020):* Primary 62G05; Secondary 51F99

To illustrate our goal more precisely, consider the following example. Let  $X, X_1, \dots, X_n$  be i.i.d. (independent and identically distributed) copies of  $X$  in a vector space equipped with a distance  $d$ . Denote by  $\mathbb{E}[X]$  and  $\bar{X}$  the expectation of  $X$  and the average of the  $X_i$ 's respectively. In many instances,  $\bar{X}$  converges to  $\mathbb{E}[X]$  at a *dimension-independent and parametric rate*. Such convergence is characterized by the following inequality:

$$\mathbb{E}[d^2(\bar{X}, \mathbb{E}[X])] \leq c\sigma^2/n, \tag{1.1}$$

where  $c > 0$  and  $\sigma^2 = \mathbb{E}[d^2(X, \mathbb{E}[X])]$  denotes the variance of  $X$ . When the underlying space is a Hilbert space, and  $d$  is the natural Hilbert metric, this inequality is in fact an equality with  $c = 1$ . More generally, the bound (1.1) holds in a Banach space if and only if it is of type 2 [31], which is a property linked to the geometry of the Banach space [29].

Using a predominantly geometric approach, we tackle the question of extending these statistical guarantees to metric spaces that are devoid of a linear structure. In recent years, this problem has gone from a theoretical curiosity to a practical necessity. Indeed, data such as images, shapes, networks, point clouds and even distributions are now routinely collected and come with the need for new statistical methods with strong performance guarantees.

### 1.1. Barycenters

The barycenter is a natural extension of the notion of expectation on an abstract metric space, which we recall next. Let  $(S, d)$  be a metric space and  $\mathcal{P}_2(S)$  be the set of (Borel) probability measures  $P$  on  $S$  such that the *variance functional* defined by

$$b \mapsto \int d^2(b, x) dP(x)$$

is finite on  $S$ . For  $P \in \mathcal{P}_2(S)$ , a *barycenter* of  $P$  is any  $b^* \in S$  such that

$$b^* \in \operatorname{argmin}_{b \in S} \int d^2(b, x) dP(x).$$

Throughout, we always implicitly assume the existence of at least one barycenter. The question of existence has been largely addressed in the literature and shown to hold in most reasonable scenarios [1, 33].

Letting  $X_1, \dots, X_n$  be  $n$  random variables drawn independently from  $P$ , an *empirical barycenter* is defined as a barycenter of the empirical distribution  $P_n = (1/n) \sum_{i=1}^n \delta_{X_i}$ , i.e.,

$$b_n \in \operatorname{argmin}_{b \in S} \frac{1}{n} \sum_{i=1}^n d^2(b, X_i).$$

The main motivation for this work is to study the statistical behavior of empirical barycenters in the context where  $(S, d)$  is the 2-Wasserstein space of probability measures over  $\mathbb{R}^D$  that is defined by  $S = \mathcal{P}_2(\mathbb{R}^D)$  equipped with the 2-Wasserstein metric.

The Wasserstein space has recently played a central role in many applications including machine learning [7, 24], computer graphics [23, 43], statistics [37, 40] and computational biology [41].

Barycenters in this setting are called *Wasserstein barycenters* and were first studied in [2] who derive a dual formulation of the associated optimization problem and establish regularity properties of the solution. Despite computational advances [20, 44] and asymptotic consistency results [3, 33, 53], little is known about the rates of convergence of barycenters on Wasserstein spaces, let alone general metric spaces. Initial contributions have focused on the Wasserstein space over the real line which is isometric to a convex subset of a Hilbert space [3] or finite-dimensional subspaces of the Wasserstein space [19, 28]. The study of rates of convergence of Wasserstein barycenters in the general case was triggered in [4], where the truly geometric nature of the question was first brought forward. However, the rates obtained in that paper suffer from the curse of dimensionality: the rate  $n^{-1}$  in (1.1) is replaced with  $n^{-1/D}$  whenever the probability distributions are defined on  $\mathbb{R}^D$ ,  $D \geq 3$ , and their techniques yield the suboptimal rate  $n^{-2/3}$  in the flat case  $D = 1$ .

In this work, we aim to improve these results to optimality. To that end, we adopt a general framework and establish dimension-independent and parametric rates for empirical barycenters, of the form (1.1), for a large family of geodesic metric spaces with curvature bounds in the sense of Alexandrov. In particular these spaces need not be finite-dimensional. More specifically, our results apply to geodesic spaces with positive curvature under a compelling synthetic geometric condition: the bi-extendibility of all geodesics emanating from a barycenter.

The 2-Wasserstein space over  $\mathbb{R}^D$  is itself a geodesic space with positive curvature [6] and parametric rates for this case follow as a simple consequence of our general result. In addition, we show that the bi-extendibility condition translates into simple regularity conditions on the Kantorovich potentials.

We end by mentioning references related to our work. General properties of barycenters in metric spaces with bounded curvature are studied in [36, 45, 51, 52]. Specific examples of spaces with negative curvature are studied in [26]. The asymptotic properties of empirical barycenters, in the case where  $S$  is a Riemannian manifold, are addressed in [9, 10, 27]. The statistical properties of empirical barycenters in abstract metric spaces have only been considered in [42] (for the case of negatively curved spaces) and [4].

## 1.2. Overview of main results

Our results rely on the notion of *extendible geodesics*. For  $\lambda_{\text{in}}, \lambda_{\text{out}} > 0$  we say that a constant-speed geodesic  $\gamma : [0, 1] \rightarrow S$  is  $(\lambda_{\text{in}}, \lambda_{\text{out}})$ -*extendible* if there exists a constant-speed geodesic  $\gamma^+ : [-\lambda_{\text{in}}, 1 + \lambda_{\text{out}}] \rightarrow S$  such that  $\gamma$  is the restriction of  $\gamma^+$  to  $[0, 1]$ . We now state our main result and refer the reader to Section 3 for omitted definitions. The *variance*  $\sigma^2$  of a distribution  $P \in \mathcal{P}_2(S)$  is given by

$$\sigma^2 = \inf_{b \in S} \int d^2(b, x) dP(x) = \int d^2(b^*, x) dP(x),$$

where  $b^*$  is a barycenter of  $P$ .

Our main theorem is stated below informally and follows directly from Theorems 3.3 and 4.2.

**Theorem 1.1** (Main theorem, informal). *Let  $(S, d)$  be a geodesic space of nonnegative curvature. Let  $P \in \mathcal{P}_2(S)$  be such that, for any  $x$  in the support of  $P$ , there exists a geodesic from  $b^*$  to  $x$  that is  $(\lambda_{\text{in}}, \lambda_{\text{out}})$ -extendible. Define*

$$k := \frac{\lambda_{\text{out}}}{1 + \lambda_{\text{out}}} - \frac{1}{\lambda_{\text{in}}}.$$

*If  $k > 0$ , then  $b^*$  is unique and for any empirical barycenter  $b_n$ , we have*

$$\mathbb{E}[d^2(b_n, b^*)] \leq \frac{4\sigma^2}{kn}.$$

Note that the constant  $k$  is a geometric characteristic of the metric measure space  $(S, d, P)$ . Since the seminal work of Cheeger and Colding [15–18], and culminating with the Lott–Sturm–Villani theory of Ricci curvature lower bounds for geodesic spaces [34, 46, 47], such triples are central to modern geometry.

To illustrate the meaning of the condition  $k > 0$ , consider the canonical example of the Euclidean unit sphere (in any dimension). In this setup, the  $(\lambda_{\text{in}}, \lambda_{\text{out}})$ -extendibility imposes that, for all  $x$  in the support of  $P$ ,

$$d(b^*, x)(1 + \lambda_{\text{in}} + \lambda_{\text{out}}) \leq \pi.$$

In addition, the condition  $k > 0$  requires that

$$1 + \lambda_{\text{in}} + \lambda_{\text{out}} \geq 2 + \lambda_{\text{out}} + \frac{1}{\lambda_{\text{out}}} \geq 4,$$

so that the support of  $P$  is necessarily contained in the spherical ball of center  $b^*$  and radius  $\pi/4$ . This bound aligns with known results, on the sphere, stating that we cannot hope for  $1/n$  rates when mass is allowed to accumulate around a point antipodal to  $b^*$  [21]. The above theorem thus requires stringent conditions on the support of  $P$  that can be largely relaxed using a local notion of extendibility where  $\lambda_{\text{in}}$  and  $\lambda_{\text{out}}$  may depend on  $x$ , and  $k$  is only required to be positive on average (see Theorem 3.5).

It is known that extendibility of geodesics in the 2-Wasserstein space is related to the regularity of Kantorovich potentials [4]. In this case, we get the following corollary.

**Corollary 1.2.** *Let  $P \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^D))$  be a probability measure on the 2-Wasserstein space and let  $\mu^* \in \mathcal{P}_2(\mathbb{R}^D)$  be a barycenter of  $P$ . Let  $\alpha, \beta > 0$  and suppose that every  $\mu \in \text{supp}(P)$  is the pushforward of  $\mu^*$  by the gradient of an  $\alpha$ -strongly convex and  $\beta$ -smooth function  $\varphi_{\mu^* \rightarrow \mu}$ , i.e.,  $\mu = (\nabla \varphi_{\mu^* \rightarrow \mu})\# \mu^*$ . If  $\beta - \alpha < 1$ , then  $\mu^*$  is unique and any empirical barycenter  $\mu_n$  of  $P$  satisfies*

$$\mathbb{E}W_2^2(\mu_n, \mu^*) \leq \frac{4\sigma^2}{(1 - \beta + \alpha)n}.$$

While our results are resolutely focused on spaces with nonnegative curvature, we obtain the following strong result about spaces of nonpositive curvature as a byproduct of the techniques developed in this paper.

**Theorem 1.3.** *Suppose  $(S, d)$  is a geodesic space satisfying  $\kappa \leq \text{curv}(S) \leq 0$  for some  $\kappa \leq 0$ . Then any distribution  $P \in \mathcal{P}_2(S)$  has a unique barycenter  $b^*$  and any empirical barycenter  $b_n$  satisfies*

$$\mathbb{E}d^2(b_n, b^*) \leq \sigma^2/n.$$

The above parametric rate should be contrasted with the slower rates obtained in [4] and [42] but that hold without curvature lower bounds.

## 2. Geometry and barycenters

### 2.1. Preliminaries

We gather important notation and standard facts from metric geometry for reference. We refer the reader to [5, 13] for an introduction.

Throughout, let  $(S, d)$  be a (Polish) geodesic metric space (see Remark 2.3). For all  $x, y \in S$ , we define a (constant-speed) *geodesic* connecting  $x$  to  $y$  to be any continuous path  $\gamma : [a, b] \rightarrow S$  such that  $\gamma(a) = x$ ,  $\gamma(b) = y$  and, for all  $a \leq s \leq t \leq b$ ,

$$d(\gamma(s), \gamma(t)) = \frac{t-s}{b-a}d(x, y).$$

Given  $\kappa \in \mathbb{R}$ , we denote by  $(M_\kappa^2, d_\kappa)$  the complete and simply connected 2-dimensional Riemannian manifold with constant sectional curvature  $\kappa$ . The diameter  $D_\kappa$  of  $M_\kappa^2$  is  $D_\kappa = +\infty$  if  $\kappa \leq 0$  and  $D_\kappa = \pi/\sqrt{\kappa}$  if  $\kappa > 0$ . We define a *triangle* in  $S$  as any set of three points  $\{p, x, y\} \subset S$ . For  $\kappa \in \mathbb{R}$ , a *comparison triangle* for  $\{p, x, y\} \subset S$  in  $M_\kappa^2$  is an isometric embedding  $\{\bar{p}, \bar{x}, \bar{y}\} \subset M_\kappa^2$  of  $\{p, x, y\}$  in  $M_\kappa^2$ . Such a comparison triangle always exists and is unique (up to isometry) provided  $\text{peri}\{p, x, y\} = d(p, x) + d(p, y) + d(x, y) < 2D_\kappa$ . We recall the definition of curvature bounds for reference.

**Definition 2.1.** Let  $\kappa \in \mathbb{R}$ . We say that  $\text{curv}(S) \geq \kappa$  if for any triangle  $\{p, x, y\} \subset S$  with  $\text{peri}\{p, x, y\} < 2D_\kappa$ , any comparison triangle  $\{\bar{p}, \bar{x}, \bar{y}\} \subset M_\kappa^2$  and any geodesic  $\gamma : [0, 1] \rightarrow S$  connecting  $x$  to  $y$  in  $S$ , we have

$$d(p, \gamma(t)) \geq d_\kappa(\bar{p}, \bar{\gamma}(t)) \tag{2.1}$$

for all  $t \in [0, 1]$ , where  $\bar{\gamma} : [0, 1] \rightarrow M_\kappa^2$  is the unique geodesic connecting  $\bar{x}$  to  $\bar{y}$  in  $M_\kappa^2$ . We say that  $\text{curv}(S) \leq \kappa$  if the same holds with converse inequality in (2.1).

Comparison angles allow one to provide useful characterisations of curvature bounds. Given  $p, x, y \in S$  with  $p \notin \{x, y\}$  and  $\text{peri}\{p, x, y\} < 2D_\kappa$ , we define the *comparison angle*  $\sphericalangle_p^\kappa(x, y) \in [0, \pi]$  at  $p$  by

$$\cos \angle_p^\kappa(x, y) := \begin{cases} \frac{d^2(p, x) + d^2(p, y) - d^2(x, y)}{2d(p, x)d(p, y)} & \text{if } \kappa = 0, \\ \frac{c_\kappa(d(x, y)) - c_\kappa(d(p, x)) \cdot c_\kappa(d(p, y))}{\kappa \cdot s_\kappa(d(p, x))s_\kappa(d(p, y))} & \text{if } \kappa \neq 0, \end{cases}$$

where, for  $r \geq 0$ ,

$$s_\kappa(r) := \begin{cases} \sin(r\sqrt{\kappa})/\sqrt{\kappa} & \text{if } \kappa > 0, \\ \sinh(r\sqrt{-\kappa})/\sqrt{-\kappa} & \text{if } \kappa < 0, \end{cases} \tag{2.2}$$

and  $c_\kappa(r) = s'_\kappa(r)$ .

The next result can be found in [14].

**Theorem 2.2.** *Let  $(S, d)$  be a geodesic space and  $\kappa \in \mathbb{R}$ . Then the following statements are equivalent.*

(1)  $\text{curv}(S) \geq \kappa$  in the sense of Definition 2.1.

(2) For all  $p, x, y, z \in S$  such that  $p \notin \{x, y, z\}$ ,

$$\angle_p^\kappa(x, y) + \angle_p^\kappa(x, z) + \angle_p^\kappa(y, z) \leq 2\pi.$$

(3) For any  $p, x, y \in S$  with  $p \notin \{x, y\}$  and any geodesics  $\gamma_x, \gamma_y : [0, 1] \rightarrow S$  from  $p$  to  $x$  and from  $p$  to  $y$  respectively, the function

$$(s, t) \mapsto \angle_p^\kappa(\gamma_x(s), \gamma_y(t))$$

is nonincreasing in each variable when the other is fixed.

Tangent cones are fundamental objects around which revolve many ideas of the paper. While they reduce to more familiar tangent spaces when  $S$  is a smooth manifold, tangent cones provide a natural proxy for tangent spaces in the context of abstract metric structures. Given  $p \in S$ , let  $\Gamma_p$  be the set of all geodesics  $\gamma : [0, 1] \rightarrow S$  with  $\gamma(0) = p$ . For  $\gamma, \sigma \in \Gamma_p$ , the *Alexandrov angle*  $\angle_p(\gamma, \sigma)$  is given by

$$\angle_p(\gamma, \sigma) = \lim_{s, t \rightarrow 0} \angle_p^0(\gamma(s), \sigma(t)).$$

Whenever  $\text{curv}(S) \geq \kappa$  or  $\text{curv}(S) \leq \kappa$ , for  $\kappa \in \mathbb{R}$ , the angle  $\angle_p : \Gamma_p^2 \rightarrow [0, \pi]$  exists and is a pseudo-metric on  $\Gamma_p$ . Hence, it induces a metric, still denoted  $\angle_p$ , on the quotient space  $\Gamma_p/\sim$  where  $\gamma \sim \sigma$  if and only if  $\angle_p(\gamma, \sigma) = 0$ . The completion  $(\Sigma_p, \angle_p)$  of  $(\Gamma_p/\sim, \angle_p)$  is called the *space of directions* at  $p$ . The *tangent cone*  $T_p S$  of  $(S, d)$  at  $p$  is the Euclidean cone over the space of directions  $\Sigma_p$ , i.e., the quotient space  $\Sigma_p \times \mathbb{R}^+/\approx$  where  $(\gamma, s) \approx (\sigma, t)$  if and only if  $s = t = 0$  or  $(\gamma, s) = (\sigma, t)$ . For  $(\gamma, s) \in \Sigma_p \times \mathbb{R}^+$ , we denote by  $[\gamma, s]$  the corresponding element in  $T_p S$  and we call  $o_p = [\gamma, 0]$  the *tip* of the cone. For all  $u = [\gamma, s], v = [\sigma, t] \in T_p S$ , we set<sup>1</sup>

$$\|u - v\|_p^2 = s^2 + t^2 - 2st \cos \angle_p(\gamma, \sigma).$$

<sup>1</sup>Observe that we often denote paths and directions by the same letters. For any two directions  $\gamma, \sigma \in \Sigma_p$ , the angle  $\angle_p(\gamma, \sigma)$  is defined without ambiguity as the Alexandrov angle between any two representatives of these directions.

We also use the notation  $\|u\|_p = \|u - o_p\|_p$  and  $\langle u, v \rangle_p = st \cos \angle_p(\gamma, \sigma)$  so that

$$\|u - v\|_p^2 = \|u\|_p^2 + \|v\|_p^2 - 2\langle u, v \rangle_p. \quad (2.3)$$

The map  $(u, v) \mapsto \|u - v\|_p$  is a metric on  $T_p S$  called the *cone metric*. The cone structure of  $T_p S$  is defined, for  $u = [\gamma, s]$  and  $\lambda \geq 0$ , by  $\lambda \cdot u = [\gamma, \lambda s]$ .

Let  $C_p \subset S$  denote the cut-locus of  $p$  and for all  $x \in S \setminus C_p$ , let  $\uparrow_p^x \in \Sigma_p$  denote the direction of the unique geodesic connecting  $p$  to  $x$ . Then the *log map* at  $p$  is the map from  $S \setminus C_p$  to  $T_p S$  that is defined by

$$\log_p(x) = [\uparrow_p^x, d(p, x)].$$

This definition can be extended to  $x \in C_p$  by selecting an arbitrary direction from  $p$  to  $x$ .

**Remark 2.3.** In the rest of this paper, we integrate functionals of the log map and now comment on its measurability. Note first that the map  $\log_p : S \rightarrow T_p S$  can be chosen to be measurable whenever the tangent cone  $T_p S$  is equipped with the  $\sigma$ -algebra  $\mathfrak{S}_B$  generated by the closed balls. This observation follows from a straightforward adaptation of [36, Lemmas 3.3 and 4.2] and implies Borel measurability of the log map if the tangent cone  $T_p S$  is separable, since, in that case,  $\mathfrak{S}_B$  coincides with the Borel  $\sigma$ -algebra.

In particular, when  $S = \mathcal{P}_2(X)$  is the 2-Wasserstein space over a separable space  $X$ , this condition follows from the characterization of  $T_p \mathcal{P}_2(X)$  in [6, Section 12.4]. It is unclear whether general conditions on  $S$  imply separability of the tangent cone—in fact, this question was raised in [36] in the case where  $S$  is proper—or even Borel measurability of the log map. Fortunately, it can be checked that our arguments require Borel measurability not of  $\log_p$  itself, but only of maps of the form  $\langle \log_p(\cdot), b \rangle_p$  for some  $b \in T_p S$ , which in turn follows from the  $\mathfrak{S}_B$ -measurability of  $\log_p$ .

A useful consequence of the monotonicity property, Theorem 2.2 (3), is the following. If  $\text{curv}(S) \geq 0$  (resp.  $\leq 0$ ), then for any  $x, y, p \in S$  the metric  $d$  and cone metric  $\|\cdot - \cdot\|_p$  satisfy

$$d(x, y) \leq \|\log_p(x) - \log_p(y)\|_p \quad (\text{resp. } \geq), \quad (2.4)$$

with equality if  $x = p$  or  $y = p$ .

## 2.2. Tangent cone at a barycenter

The notation  $\|\cdot\|_p$  and  $\langle \cdot, \cdot \rangle_p$  suggests that the cone  $T_p S$  possesses a Hilbert-like structure. However, the tangent cone may fail to be even geodesic whenever  $S$  has only lower bounded curvature (see [25]). To overcome this limitation, we gather facts on the structure of  $T_p S$  that are used in what follows, in particular when  $p$  is a barycenter.

**Theorem 2.4.** *Let  $\kappa \in \mathbb{R}$  and let  $(S, d)$  be a geodesic space with  $\text{curv}(S) \geq \kappa$ . Then the following holds:*

- (1) *For any  $p \in S$ , the tangent cone  $T_p S$  can be isometrically embedded in a geodesic space  $\mathcal{T}_p S$  such that  $\text{curv}(\mathcal{T}_p S) \geq 0$ .*

(2) For any  $P \in \mathcal{P}_2(S)$  and any barycenter  $b^* \in S$  of  $P$ ,

$$\iint \langle \log_{b^*}(x), \log_{b^*}(y) \rangle_{b^*} dP(x) dP(y) = 0.$$

(3) For any  $P \in \mathcal{P}_2(S)$  and any barycenter  $b^*$  of  $P$ , there exists a subset  $\mathcal{L}_{b^*} S \subset T_{b^*} S$  which is a Hilbert space when equipped with the restricted cone metric, and such that  $\log_{b^*}(\text{supp}(P)) \subset \mathcal{L}_{b^*} S$ .

(4) For any  $P \in \mathcal{P}_2(S)$ , any barycenter  $b^*$  of  $P$ , any  $Q \in \mathcal{P}_2(S)$  with  $\log_{b^*}(\text{supp}(Q)) \subset \mathcal{L}_{b^*} S$  and any  $b \in S$ , we have

$$\int \langle \log_{b^*}(x), \log_{b^*}(b) \rangle_{b^*} dQ(x) = \langle u, \log_{b^*}(b) \rangle_{b^*},$$

where  $u$  is the (Pettis) integral  $\int v d\bar{Q}(v)$  with  $\bar{Q} = (\log_{b^*})_{\#} Q$ .

*Proof.* (1) This statement follows directly from [5, Claim 3.3.2, Theorem 3.4.1 and Theorem 11.3.1]. Note that  $\mathcal{T}_p S$  is formally described as the ultralimit of blowups of  $S$  at  $p$  and called the *ultratangent cone*.

(2) Let  $X_1, \dots, X_n$  be i.i.d. with distribution  $P$  and let  $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ . Let  $f : S^2 \rightarrow \mathbb{R}$  be defined by  $f(x, y) = \langle \log_{b^*}(x), \log_{b^*}(y) \rangle_{b^*}$  and observe that  $f \in L^2(P \otimes P)$ . Moreover, explicit computations show that

$$\mathbb{E}|(P \otimes P)f - (P_n \otimes P_n)f|^2 \xrightarrow{n \rightarrow \infty} 0,$$

so that there exists a subsequence  $(P_{n_k})$  such that  $(P_{n_k} \otimes P_{n_k})f$  converges almost surely to  $(P \otimes P)f$ . Since  $P_{n_k}$  is finitely supported, it follows from [30, Proposition 3.2] that  $(P_{n_k} \otimes P_{n_k})f \geq 0$ . Therefore,  $(P \otimes P)f \geq 0$ . Note that this first inequality holds even when  $b^*$  is not a barycenter.

To complete the proof, we prove the reverse inequality.

Without loss of generality, suppose that  $\kappa < 0$ . Then, from [5, Claim 1.1.1.d and Lemma 5.3.1], we deduce there exists a constant  $c > 0$  (depending only on  $\kappa$ ) such that, for all  $x, y \in S$ ,

$$\langle_{b^*}^{\kappa}(x, y) \leq \langle_{b^*}^0(x, y) \leq \langle_{b^*}^{\kappa}(x, y) + cd(b^*, x)d(b^*, y). \tag{2.5}$$

Now for all  $x, y \in S$ , let  $\gamma_x$  and  $\gamma_y$  be two geodesics connecting  $b^*$  to  $x$  and  $y$  respectively and with respective directions  $\uparrow_{b^*}^x$  and  $\uparrow_{b^*}^y$ . Then, letting

$$\langle x, y \rangle_{b^*}^0 = d(b^*, x)d(b^*, y) \cos \langle_{b^*}^0(x, y),$$

we get

$$\begin{aligned} \langle \log_{b^*}(x), \log_{b^*}(y) \rangle_{b^*} &= d(b^*, x)d(b^*, y) \cos \langle_{b^*}(\uparrow_{b^*}^x, \uparrow_{b^*}^y) \\ &= \lim_{s,t \rightarrow 0} d(b^*, x)d(b^*, y) \cos \langle_{b^*}^0(\gamma_x(s), \gamma_y(t)) \\ &\leq \lim_{s,t \rightarrow 0} d(b^*, x)d(b^*, y) \cos \langle_{b^*}^{\kappa}(\gamma_x(s), \gamma_y(t)) \\ &\leq d(b^*, x)d(b^*, y) \cos \langle_{b^*}^{\kappa}(x, y) \leq \langle x, y \rangle_{b^*}^0 + cd^2(b^*, x)d^2(b^*, y), \end{aligned}$$



according to Theorem 2.2 (3), inequality (2.5) and the Lipschitz continuity of  $\cos(\cdot)$ . Now observe that, for all  $x \in S$  and all  $t \in (0, 1)$ ,

$$\begin{aligned} d^2(x, \gamma_y(t)) - d^2(b^*, x) &= d^2(b^*, \gamma_y(t)) - 2\langle x, y \rangle_{b^*}^0 \\ &\leq d^2(b^*, \gamma_y(t)) - 2\langle \log_{b^*}(x), \log_{b^*}(\gamma_y(t)) \rangle_{b^*} + cd^2(b^*, x)d^2(b^*, \gamma_y(t)) \\ &= t^2d^2(b^*, y) - 2t\langle \log_{b^*}(x), \log_{b^*}(y) \rangle_{b^*} + ct^2d^2(b^*, x)d^2(b^*, y). \end{aligned}$$

Integrating with respect to  $dP(x)$ , using the definition of a barycenter and letting  $t$  go to 0, we deduce that

$$\int \langle \log_{b^*}(x), \log_{b^*}(y) \rangle_{b^*} dP(x) \leq 0.$$

Since this holds for all  $y \in S$ , integrating with respect to  $dP(y)$  gives the result.

(3) Given  $p \in S$  and two elements  $u = [\gamma, s]$ ,  $v = [\sigma, t] \in T_p S$ , we write  $u + v = 0$  and say that  $u$  and  $v$  are *opposite* to each other if  $(s = t = 0)$  or  $(s = t \text{ and } \angle_p(\gamma, \sigma) = \pi)$ . Then we set

$$\mathcal{L}_p S = \{u \in T_p S : \exists v \in T_p S, u + v = 0\}.$$

Given  $u \in \mathcal{L}_p S$ , there is a unique vector opposite to  $u$  denoted  $-u$ . The fact that  $\mathcal{L}_p S$  is a Hilbert space for the restricted cone metric follows from [5, Theorem 11.6.4]. The inclusion  $\log_{b^*}(\text{supp}(P)) \subset \mathcal{L}_{b^*} S$  follows from statement (2) and [32].

(4) Fix a barycenter  $b^*$  of  $P$  and some  $b \in S$ . Note that if  $b$  is in the support of  $P$ , it follows readily from (3) that (4) holds. The main purpose of this proof is to show that the same holds for any  $b \in S$ , not necessarily in  $\text{supp}(P)$ . For brevity we use the following notation:

$$|x - y| = \|\log_{b^*}(x) - \log_{b^*}(y)\|_{b^*}, \quad |x| = \|\log_{b^*}(x)\|_{b^*}$$

and

$$\langle x, y \rangle = \langle \log_{b^*}(x), \log_{b^*}(y) \rangle_{b^*} \quad \forall x, y \in S.$$

In addition, whenever  $u \in T_{b^*} S$  and  $y \in S$ , the notation  $|u - y|$ ,  $|u|$  and  $\langle u, y \rangle$  should be understood as  $\|u - \log_{b^*}(y)\|_{b^*}$ ,  $\|u\|_{b^*}$  and  $\langle u, \log_{b^*}(y) \rangle_{b^*}$  respectively.

We first prove that  $\mathcal{L}_{b^*} S \ni u \mapsto \langle u, b \rangle$  is a convex function. To that end, fix  $t \in (0, 1)$  and  $u_0, u_1 \in \mathcal{L}_{b^*} S$ , and set  $u_t = (1 - t)u_0 + tu_1 \in \mathcal{L}_{b^*} S$ . The path  $[0, 1] \ni t \mapsto u_t$  defines a geodesic in  $\mathcal{T}_{b^*} S$  according to (1) and (3). Since  $\text{curv}(\mathcal{T}_{b^*} S) \geq 0$  we see from Definition 2.1 that, for all  $y \in S$  and  $t \in [0, 1]$ ,

$$|u_t - y|^2 \geq (1 - t)|u_0 - y|^2 + t|u_1 - y|^2 - t(1 - t)|u_0 - u_1|^2,$$

with equality if  $y = b^*$  since  $u_t, u_0$  and  $u_1$  belong to  $\mathcal{L}_{b^*} S$ . Hence, by (2.3), we have

$$\begin{aligned} 2\langle u_t, b \rangle &= |u_t|^2 + |b|^2 - |u_t - b|^2 \\ &\leq (1 - t)(|u_0|^2 + |b|^2 - |u_0 - b|^2) + t(|u_1|^2 + |b|^2 - |u_1 - b|^2) \\ &= (1 - t)2\langle u_0, b \rangle + t2\langle u_1, b \rangle, \end{aligned}$$

so that  $u \mapsto \langle u, b \rangle$  is a convex function on the Hilbert space  $\mathcal{L}_{b^\star} S$ . Therefore, for all  $u \in \mathcal{L}_{b^\star} S$ , there exists  $g = g(u)$  such that for all  $v \in \text{supp } Q$ ,

$$\langle \log_{b^\star}(v), b \rangle \geq \langle u, b \rangle + \langle \log_{b^\star}(v) - u, g \rangle.$$

Choosing  $u$  to be the Pettis integral  $\int \log_{b^\star}(v) \, dQ(v)$  and integrating  $v$  with respect to  $Q$  yields

$$\int \langle \log_{b^\star}(v), b \rangle \, dQ(v) \geq \langle u, b \rangle, \tag{2.6}$$

The same arguments apply to the map  $u \mapsto \langle -u, b \rangle$ , so that

$$\int \langle -\log_{b^\star}(v), b \rangle \, dQ(v) \geq \langle -u, b \rangle. \tag{2.7}$$

To conclude the proof, it remains to show that  $\langle -v, b \rangle = -\langle v, b \rangle$  for all  $v \in \mathcal{L}_{b^\star} S$  (in particular, for  $u \in \mathcal{L}_{b^\star} S$ ). To that end, select  $v = [\gamma, s] \in \mathcal{L}_{b^\star} S$  and let  $-v = [\sigma, s]$ . Then on the one hand, using the definition of opposite points and the triangle inequality on the space of directions  $\Sigma_p$ , it follows that  $\pi = \angle_{b^\star}(\gamma, \sigma) \leq \angle_{b^\star}(\gamma, \uparrow_{b^\star}^b) + \angle_{b^\star}(\uparrow_{b^\star}^b, \sigma)$ . On the other hand, the fact that  $\text{curv}(\mathcal{T}_p S) \geq 0$  implies that

$$\angle_{b^\star}(\gamma, \sigma) + \angle_{b^\star}(\gamma, \uparrow_{b^\star}^b) + \angle_{b^\star}(\uparrow_{b^\star}^b, \sigma) \leq 2\pi,$$

according to Theorem 2.2(2). Combining these inequalities we find that  $\angle_{b^\star}(\gamma, \uparrow_{b^\star}^b) + \angle_{b^\star}(\uparrow_{b^\star}^b, \sigma) = \pi$  so that, by definition of the bracket  $\langle \cdot, \cdot \rangle$  and using elementary trigonometry, we get  $\langle -v, b \rangle = -\langle v, b \rangle$ . ■

### 3. Main results

We now turn to the proof of our main result: parametric and dimension-free rates for empirical barycenters on Alexandrov spaces with curvature bounded below. We study the case of positively curved spaces in more detail, and in particular the 2-Wasserstein space.

#### 3.1. Hugging

We begin with a useful identity that controls the average curvature of the square distance around its minimum.

Given  $P \in \mathcal{P}_2(S)$  and a barycenter  $b^\star$  of  $P$ , define for all  $x, b \in S$  the *hugging function* at  $b^\star$  by

$$k_{b^\star}^b(x) = 1 - \frac{\|\log_{b^\star}(x) - \log_{b^\star}(b)\|_{b^\star}^2 - d^2(x, b)}{d^2(b, b^\star)}. \tag{3.1}$$

Note that if  $S$  is a Hilbert space, then  $k_{b^\star}^b(x) = 1$  for all  $x, b \in S$ . The hugging function  $k_{b^\star}^b$  measures the proximity of  $S$  to its tangent cone  $T_{b^\star} S$ . The next result is a generalization of [4, Theorem 3.2] and demonstrates the central role of the hugging coefficient in the context of barycenters: it precisely controls the quadratic growth of the variance functional around its minimum.

**Theorem 3.1** (Variance equality). *Let  $\kappa \in \mathbb{R}$  and  $(S, d)$  be a geodesic space with  $\text{curv}(S) \geq \kappa$ . Let  $P \in \mathcal{P}_2(S)$  and  $b^*$  be a barycenter of  $P$ . Then, for all  $b \in S$ ,*

$$d^2(b, b^*) \int k_{b^*}^b(x) \, dP(x) = \int (d^2(x, b) - d^2(x, b^*)) \, dP(x), \quad (3.2)$$

where  $k_{b^*}^b(x)$  is as in (3.1).

*Proof.* We adopt the same notation as in the proof of Theorem 2.4(4). By definition of the cone metric  $|\cdot - \cdot|$  and bracket  $\langle \cdot, \cdot \rangle$  in  $T_{b^*}S$ , we get

$$\begin{aligned} k_{b^*}^b(x) d^2(b, b^*) &= d^2(b, b^*) + d^2(x, b) - |b - x|^2 \\ &= d^2(x, b) - d^2(x, b^*) + 2\langle x, b \rangle. \end{aligned}$$

The result follows by integrating both sides with respect to  $dP(x)$ , in Theorem 2.4(4) with  $Q = P$  and noticing that  $b^*$  is the Pettis integral of  $(\log_{b^*})_{\#}P$  in the Hilbert space  $\mathcal{L}_{b^*}S$ .  $\blacksquare$

**Remark 3.2.** By definition of a barycenter, the right-hand side of identity (3.2) is non-negative. It follows that

$$\int k_{b^*}^b(x) \, dP(x) \geq 0 \quad (3.3)$$

for all  $b \neq b^* \in S$ .

Note also that for any triple  $x, b, b^*$ , the curvature properties of  $S$  are directly reflected in  $k_{b^*}^b(x)$ . Indeed,  $k_{b^*}^b(x) \geq 1$  if  $S$  is non-positively curved and  $k_{b^*}^b(x) \leq 1$  if  $S$  is positively curved according to (2.4).

### 3.2. A master theorem for spaces with curvature lower bound

This section presents general statistical guarantees for empirical barycenters, in connection to lower bounds on the function  $k_{b^*}^b$ .

Throughout, we fix a probability measure  $P$  with barycenter  $b^*$  on a geodesic space  $(S, d)$ . Moreover, we denote by  $\sigma^2$  the variance of  $P$ , which is defined by

$$\sigma^2 = \int d^2(b^*, x) \, dP(x).$$

**Theorem 3.3.** *Suppose that  $\text{curv}(S) \geq \kappa$  for some  $\kappa \in \mathbb{R}$ . If there exists a constant  $k_{\min} > 0$  such that  $k_{b^*}^b(x) \geq k_{\min}$  for all  $x, b \in S$ , then  $b^*$  is unique and any empirical barycenter  $b_n$  satisfies*

$$\mathbb{E}d^2(b_n, b^*) \leq \frac{4\sigma^2}{nk_{\min}^2}.$$

*Proof.* Let  $k(x) := k_{b^*}^{b_n}(x)$  and recall the convention introduced in the proof of Theorem 2.4 concerning the use of the cone metric  $\|\cdot - \cdot\|_{b^*} = |\cdot - \cdot|$  and the associated

bracket  $\langle \cdot, \cdot \rangle_{b^*} = \langle \cdot, \cdot \rangle$ . We also use the following standard notation:

$$Pf(\bullet) = \int f \, dP \quad \text{and} \quad P_n f(\bullet) = \frac{1}{n} \sum_{i=1}^n f(X_i).$$

On the one hand, from the variance equality we have

$$\begin{aligned} P d^2(b_n, \bullet) - P d^2(b^*, \bullet) &= d^2(b_n, b^*) P k(\bullet), \\ P_n d^2(b^*, \bullet) - P_n d^2(b_n, \bullet) &= d^2(b_n, b^*) P_n k_{b_n}^{b^*}(\bullet). \end{aligned}$$

Adding these two identities yields

$$(P - P_n)(d^2(b_n, \bullet) - d^2(b^*, \bullet)) = d^2(b_n, b^*)(P k(\bullet) + P_n k_{b_n}^{b^*}(\bullet)). \tag{3.4}$$

On the other hand, by the definition of  $k$ , we have

$$\begin{aligned} d^2(b_n, \bullet) - d^2(b^*, \bullet) &= (|b_n - \bullet|^2 - |b^* - \bullet|^2) + (d^2(b_n, \bullet) - |b_n - \bullet|^2) \\ &= (|b_n - \bullet|^2 - |b^* - \bullet|^2) + (k(\bullet) - 1)d^2(b_n, b^*). \end{aligned}$$

Hence, it follows that

$$\begin{aligned} (P - P_n)(d^2(b_n, \bullet) - d^2(b^*, \bullet)) \\ = (P - P_n)(|b_n - \bullet|^2 - |b^* - \bullet|^2) + d^2(b_n, b^*)(P - P_n)k(\bullet). \end{aligned} \tag{3.5}$$

Combining (3.4) and (3.5) yields

$$d^2(b_n, b^*) P_n [k(\bullet) + k_{b_n}^{b^*}(\bullet)] = (P - P_n)(|b_n - \bullet|^2 - |b^* - \bullet|^2). \tag{3.6}$$

We now focus on the right-hand side of (3.6). By definition of  $\langle \cdot, \cdot \rangle$ ,

$$|b_n - \bullet|^2 - |b^* - \bullet|^2 = |b_n|^2 - 2\langle \bullet, b_n \rangle.$$

Hence, by Theorem 2.4 (4) applied to  $Q = P$  and  $Q = P_n$ , we get

$$\begin{aligned} (P - P_n)(|b_n - \bullet|^2 - |b^* - \bullet|^2) &= 2P_n \langle \bullet, b_n \rangle = 2\langle b_n^*, b_n \rangle \\ &\leq 2|b_n^*| \cdot |b_n| = 2|b_n^*| d(b_n, b^*), \end{aligned} \tag{3.7}$$

where  $b_n^*$  stands for the barycenter of  $(\log_{b^*})_{\#}(P_n)$  in the Hilbert space  $\mathcal{L}_{b^*} S \subset T_{b^*} S$ . Combining (3.6) and (3.7) yields

$$d(b_n, b^*) P_n [k(\bullet) + k_{b_n}^{b^*}(\bullet)] \leq 2|b_n^*|. \tag{3.8}$$

Hence, since  $P_n k_{b_n}^{b^*}(\bullet) \geq 0$ , if  $k(x) \geq k_{\min} > 0$  for all  $x \in S$ , we see by assumption that

$$k_{\min}^2 d^2(b_n, b^*) \leq 4|b_n^*|^2.$$

Since  $(\log_{b^\star})_\#(P_n)$  is the empirical distribution associated to  $(\log_{b^\star})_\#(P)$ , we infer by properties of averages in Hilbert spaces that

$$\mathbb{E}|b_n^\star|^2 = \sigma^2/n, \tag{3.9}$$

where we also use the fact that the variance of  $(\log_{b^\star})_\#(P)$  is given by

$$\int \|\log_{b^\star}(x)\|_{b^\star}^2 dP(x) = \int d^2(x, b^\star) dP(x) = \sigma^2.$$

This completes the proof of the theorem. ■

As pointed out in Remark 3.2, the condition  $\text{curv}(S) \leq 0$  implies that  $k_{b_n^\star}^{b_n}(x) \geq 1$  and  $k_{b_n^\star}^{b_n}(x) \geq 1$  for all  $x \in S$ . The next result therefore follows readily from (3.8) and (3.9).

**Corollary 3.4.** *Suppose  $(S, d)$  is a geodesic space satisfying  $\kappa \leq \text{curv}(S) \leq 0$  for some  $\kappa \leq 0$ . Then any distribution  $P \in \mathcal{P}_2(S)$  has a unique barycenter  $b^\star$  and any empirical barycenter  $b_n$  satisfies*

$$\mathbb{E}d^2(b_n, b^\star) \leq \sigma^2/n.$$

A careful inspection of our proof techniques reveals that the previous result extends as well to complete and simply connected Riemannian manifolds with nonpositive sectional curvature, also known as *Hadamard manifolds*. Indeed, the use of curvature lower bounds in our results is twofold. First, it guarantees a Hilbert-like structure of the tangent cone at a barycenter. Second, we made use of the fact that the pushforward of  $P$  under  $\log_{b^\star}$  has the tip of the cone as its barycenter (i.e., the fact that any barycenter is a so-called *exponential barycenter* [22]). The former is satisfied by the manifold assumption, and the latter is the content of [9, Theorem 2.1(b, c)], at least when  $M$  is Hadamard.

Our proof technique does not extend, however, to the case of general geodesic spaces with nonpositive curvature and no curvature lower bound. Indeed, it can be easily checked that claim (2) in Theorem 2.4 is typically not satisfied by distributions supported on the tripod (gluing of three line segments) for example.

### 3.3. Tail bound

We now provide an extension of Theorem 3.3 to the case of more general bounds on  $k_{b^\star}^b$ . The next result shows in particular that the requirement, made in Theorem 3.3, that  $k_{b^\star}^b(x) \geq k_{\min} > 0$  for all  $x, b \in S$ , can be relaxed to  $k_{b^\star}^b(x) \geq k_{\min}$  for all  $x, b \in S$  and  $k_{\min}$  possibly negative, provided that  $Pk_{b^\star}(\cdot) > 0$  where

$$k_{b^\star}(x) = \min_b k_{b^\star}^b(x).$$

For simplicity, we prove this extension under an additional integrability assumption on  $P$ . We say that  $P$  (with barycenter  $b^\star$ ) is *subgaussian* with variance proxy  $\zeta^2 > 0$  if

$$\int \exp\left(\frac{d^2(b^\star, x)}{2\zeta^2}\right) dP(x) \leq 2. \tag{3.10}$$

Such conditions are commonly used in high-dimensional probability and statistics [39, 49].

**Theorem 3.5.** *Suppose that  $\text{curv}(S) \geq \kappa$  for some  $\kappa \in \mathbb{R}$ . Fix a subgaussian probability distribution  $P$  on  $S$  with variance proxy  $\zeta^2 > 0$  and barycenter  $b^*$ . Suppose that there exists  $k_{\min} < 0$  such that  $k_{b^*}^b(x) \geq k_{\min}$  for all  $x, b \in S$  and that  $Pk_{b^*}(\cdot) > 0$ . Then  $b^*$  is unique and, for any empirical barycenter  $b_n$  and any  $\delta \in (0, 1)$ , we have*

$$d^2(b_n, b^*) \leq \frac{c_1}{n} \log\left(\frac{2}{\delta}\right)$$

with probability at least  $1 - \delta - e^{-c_2 n}$  where  $c_1, c_2 > 0$  are independent of  $n$ .

*Proof.* Denote  $k(x) = k_{b^*}(x)$  for brevity. For all  $c \in (0, 1)$  and  $t > 0$ ,

$$\begin{aligned} \mathbb{P}(d(b_n, b^*) > t) &\leq \mathbb{P}(d(b_n, b^*) > t, P_n k \geq cPk) + \mathbb{P}(P_n k < cPk) \\ &\leq \mathbb{P}(d(b_n, b^*) P_n k > tcPk) + \mathbb{P}((P - P_n)k > (1 - c)Pk) \\ &\leq \mathbb{P}(2|b_n^*| > tcPk) + \mathbb{P}((P - P_n)k > (1 - c)Pk), \end{aligned}$$

where the last inequality follows from inequality (3.8). Using a one-sided variant of Bernstein’s inequality [12, (2.10)], we get

$$\mathbb{P}((P - P_n)k > (1 - c)Pk) \leq \exp\left(-\frac{n}{2} \frac{(1 - c)Pk}{|k_{\min}|} \left(\frac{(1 - c)|k_{\min}|Pk}{Pk^2} \wedge \frac{3}{2}\right)\right).$$

Using (3.10), and the classical properties of subgaussian variables, we get

$$\mathbb{P}(2|b_n^*| > tcPk) \leq 2 \exp\left(-\frac{nt^2 c^2 (Pk)^2}{8\zeta^2}\right). \tag{3.11}$$

The result then follows by considering, for any  $c \in (0, 1)$ ,

$$c_1 = \frac{c^2 (Pk)^2}{8\zeta^2} \quad \text{and} \quad c_2 = \frac{(1 - c)Pk}{2|k_{\min}|} \left(\frac{(1 - c)|k_{\min}|Pk}{Pk^2} \wedge \frac{3}{2}\right). \quad \blacksquare$$

To better appreciate the extent of this relaxation, recall the illustrative example of the sphere used in Section 1.2. Theorem 3.5 covers distributions whose support may be almost the entire sphere. However, note that the existence of a finite  $k_{\min}$  precludes a support that includes the cut-locus of the barycenter, even if  $P$  puts arbitrarily small probability on small neighborhoods of the cut-locus. While such cases are out of the scope of the present paper, we anticipate that careful truncation arguments should suffice to derive parametric rates of convergence in such favorable cases.

#### 4. Positively curved spaces

Theorem 3.3 guarantees a dimension-free parametric rate of convergence of  $b_n$  under a uniform positive lower bound on the function  $(x, b) \mapsto k_{b^*}^b(x)$ . Such a condition is closely linked to the better known notion of  $k$ -convexity which is connected to positive curvature

upper bounds [35]. In fact, it is not hard to check that assuming that  $k_{b^*}^b(x) \geq k_{\min}$  uniformly not only in  $b, x$  but also in  $b^* \in \text{supp}(P)$  is equivalent to  $k_{\min}$ -convexity of the support of  $P$ . As a result, the condition of Theorem 3.3 is weaker than  $k_{\min}$ -convexity of the support of  $P$  since it only imposes control on geodesics emanating from  $b^*$ . The rest of this section is devoted to developing a suitable relaxation via the notion of geodesic extendibility.

#### 4.1. Extendible geodesics

We present a compelling synthetic geometric condition that implies this lower bound in the context of positively curved spaces: the extendibility, by a given factor, of all geodesics (emanating from and arriving at) the barycenter  $b^*$ . To formalize the notion of extendible geodesics, consider a (constant-speed) geodesic  $\gamma: [0, 1] \rightarrow S$ . For  $(\lambda_{\text{in}}, \lambda_{\text{out}}) \in \mathbb{R}_+^2$ , we say that  $\gamma$  is  $(\lambda_{\text{in}}, \lambda_{\text{out}})$ -extendible if there exists a (constant-speed) geodesic  $\gamma^+: [-\lambda_{\text{in}}, 1 + \lambda_{\text{out}}] \rightarrow S$  such that  $\gamma$  is the restriction of  $\gamma^+$  to  $[0, 1]$ . Before we state our sufficient condition we shall first need a fact given in [4].

**Theorem 4.1.** *Suppose that  $\text{curv}(S) \geq 0$ . Let  $P \in \mathcal{P}_2(S)$  with barycenter  $b^*$ . Suppose that, for each  $x \in \text{supp}(P)$ , there exists a geodesic  $\gamma_x: [0, 1] \rightarrow S$  connecting  $b^*$  to  $x$  which is  $(0, \lambda)$ -extendible. Suppose in addition that  $b^*$  remains a barycenter of distribution  $P_\lambda = (e_\lambda)_\#P$  where  $e_\lambda(x) = \gamma_x^+(1 + \lambda)$ . Then for all  $b \in S$ ,*

$$\frac{\lambda}{1 + \lambda} d^2(b, b^*) \leq \int (d^2(x, b) - d^2(x, b^*)) dP(x). \quad (4.1)$$

According to the variance equality of Theorem 3.1, inequality (4.1) is equivalent to the statement that, for all  $b \in S$ ,

$$\int k_{b^*}^b(x) dP(x) \geq \frac{\lambda}{1 + \lambda}.$$

We will use this observation next.

**Theorem 4.2.** *Suppose that  $\text{curv}(S) \geq 0$  and let  $x, b, b^* \in S$ . Suppose that, for some  $\lambda_{\text{in}}, \lambda_{\text{out}} > 0$ , there is a geodesic connecting  $b^*$  to  $x$  which is  $(\lambda_{\text{in}}, \lambda_{\text{out}})$ -extendible. Then*

$$k_{b^*}^b(x) \geq \frac{\lambda_{\text{out}}}{1 + \lambda_{\text{out}}} - \frac{1}{\lambda_{\text{in}}}.$$

*Proof.* Let  $\gamma: [0, 1] \rightarrow S$  be a  $(\lambda_{\text{in}}, \lambda_{\text{out}})$ -extendible geodesic connecting  $b^*$  to  $x$  and let  $\gamma^+: [-\lambda_{\text{in}}, 1 + \lambda_{\text{out}}] \rightarrow S$  be its extension. Let  $z = \gamma^+(-\xi)$  where  $\xi = \lambda_{\text{in}}/(1 + \lambda_{\text{out}})$ . Then it may be easily checked that  $b^*$  is a barycenter of the probability measure

$$P := \frac{\xi}{1 + \xi} \delta_x + \frac{1}{1 + \xi} \delta_z.$$

Now, we wish to apply Theorem 4.1 to  $P$ . To this end, note that the geodesic  $\gamma$  from  $b^*$  to  $x$  is  $(0, 1 + \lambda_{\text{out}})$ -extendible by assumption with  $e_{\lambda_{\text{out}}}(x) = \gamma^+(1 + \lambda_{\text{out}})$ . Similarly, we

check that the geodesic  $\sigma : [0, 1] \rightarrow S$  connecting  $b^*$  to  $z$  and defined by  $\sigma(t) = \gamma^+(-t\xi)$  is  $(0, 1 + \lambda_{\text{out}})$ -extendible by construction with  $e_{\lambda_{\text{out}}}(z) = \gamma^+(-\lambda_{\text{in}})$ . Finally, one checks that  $b^*$  remains a barycenter of the probability measure  $P_{\lambda_{\text{out}}} = (e_{\lambda_{\text{out}}})_{\#}P$ . As a result, Theorem 4.1 implies that

$$\frac{\lambda_{\text{out}}}{1 + \lambda_{\text{out}}} \leq Pk_{b^*}^b(\cdot) = \frac{\xi}{1 + \xi}k_{b^*}^b(x) + \frac{1}{1 + \xi}k_{b^*}^b(z).$$

Finally, the fact that  $\text{curv}(S) \geq 0$  implies that  $d(x, y) \leq \|\log_{b^*}(x) - \log_{b^*}(y)\|_{b^*}$  for all  $x, y \in S$ , which imposes that  $k_{b^*}^b(z) \leq 1$  for all  $b, z \in S$ . Hence, we obtain

$$k_{b^*}^b(x) \geq \frac{1 + \xi}{\xi} \left( \frac{\lambda_{\text{out}}}{1 + \lambda_{\text{out}}} - \frac{1}{1 + \xi} \right) = \frac{\lambda_{\text{out}}}{1 + \lambda_{\text{out}}} - \frac{1}{\lambda_{\text{in}}},$$

which completes the proof. ■

#### 4.2. The Wasserstein space

The case of geodesic spaces with lower curvature bound contains several spaces of interest such as the space of metric measure spaces equipped with the Gromov–Wasserstein distance [48] or the Wasserstein space over a positively curved space. In this subsection, we explore the extendible geodesics condition, described in the previous subsection, in the context of the Wasserstein space over a separable Hilbert space. This space exhibits a very particular structure that allows for a simple and transparent formulation of the extendibility of geodesics in terms of the regularity of Kantorovich potentials. Given a separable Hilbert space  $H$ , with inner product  $\langle \cdot, \cdot \rangle$  and norm  $|\cdot|$ , recall that the *subdifferential*  $\partial\varphi \subset H^2$  of a function  $\varphi : H \rightarrow \mathbb{R}$  is defined by  $\partial\varphi = \{(x, g) : \forall y \in H, \varphi(y) \geq \varphi(x) + \langle g, y - x \rangle\}$ . We denote  $\partial\varphi(x) = \{g \in H : (x, g) \in \partial\varphi\}$ . The function  $\varphi$  is called  *$\alpha$ -strongly convex* if, for all  $x \in H$ ,  $\partial\varphi(x) \neq \emptyset$  and

$$\langle g, x - y \rangle \geq \varphi(x) - \varphi(y) + \frac{\alpha}{2}|x - y|^2$$

for all  $g \in \partial\varphi(x)$  and all  $y \in H$ . We recall the following result.

**Theorem 4.3** ([4, Theorem 3.5]). *Let  $(S, d) = (\mathcal{P}_2(H), W_2)$  be the Wasserstein space over a separable Hilbert space  $H$ . Let  $\mu, \nu \in S$  and let  $\gamma : [0, 1] \rightarrow S$  be a geodesic connecting  $\mu$  to  $\nu$ . Then  $\gamma$  is  $(0, \lambda)$ -extendible if, and only if, the support of the optimal transport plan between  $\mu$  and  $\nu$  lies in the subdifferential  $\partial\varphi_{\mu \rightarrow \nu}$  of a  $\frac{\lambda}{1 + \lambda}$ -strongly convex map  $\varphi_{\mu \rightarrow \nu} : H \rightarrow \mathbb{R}$ .*

In the above theorem,  $\varphi_{\mu \rightarrow \nu}$  is defined as follows. Let  $f_{\mu \rightarrow \nu}, g_{\nu \rightarrow \mu}$  be optimal Kantorovich potentials, i.e., solutions of the dual Kantorovich problem

$$W_2^2(\mu, \nu) = \sup \left\{ \int f \, d\mu + \int g \, d\nu : f, g \in C_b(H), f(x) + g(y) \leq |x - y|^2 \right\},$$



where  $C_b(H)$  denotes the set of real-valued bounded continuous functions on  $H$ . Then, for all  $x \in H$ ,

$$f_{\mu \rightarrow \nu}(x) = 2|x|^2 - \varphi_{\mu \rightarrow \nu}(x), \quad g_{\nu \rightarrow \mu}(x) = 2|x|^2 - \varphi_{\nu \rightarrow \mu}(x),$$

and one checks that  $\varphi_{\nu \rightarrow \mu} = \varphi_{\mu \rightarrow \nu}^*$  is the Fenchel–Legendre conjugate of  $\varphi_{\mu \rightarrow \nu}$  [50, Theorem 5.10]. Our next result is in the same spirit as Theorem 4.3 and characterizes the  $(\lambda_{\text{in}}, 1 + \lambda_{\text{out}})$ -extendibility of geodesics in the Wasserstein space in a specific scenario. Recall that a convex function  $\varphi : H \rightarrow \mathbb{R}$  is called  $\beta$ -smooth if

$$\langle g_x, x - y \rangle \leq \varphi(x) - \varphi(y) + \frac{\beta}{2}|x - y|^2, \quad \forall g_x \in \partial\varphi(x),$$

for all  $x, y \in H$ . In fact, for  $\beta > 0$ , it is easy to check that convex  $\beta$ -smooth functions are differentiable. Therefore, a convex function  $\varphi : H \rightarrow \mathbb{R}$  is  $\beta$ -smooth if and only if it is differentiable and

$$\langle \nabla\varphi(x), x - y \rangle \leq \varphi(x) - \varphi(y) + \frac{\beta}{2}|x - y|^2.$$

It is known that a convex function is  $\beta$ -smooth if and only if its Fenchel–Legendre transform is  $1/\beta$ -strongly convex [8, Theorem 18.15]. Since  $\varphi_{\nu \rightarrow \mu} = \varphi_{\mu \rightarrow \nu}^*$ , the next result follows readily from Theorems 3.3, 4.2 and 4.3.

**Corollary 4.4.** *Let  $(S, d) = (\mathcal{P}_2(H), W_2)$  be the Wasserstein space over a separable Hilbert space  $H$ . Let  $P \in \mathcal{P}_2(S)$  with barycenter  $\mu^* \in S$ . Let  $\alpha, \beta > 0$  and suppose that every  $\mu \in \text{supp}(P)$  is the pushforward of  $\mu^*$  by the gradient of an  $\alpha$ -strongly convex and  $\beta$ -smooth function  $\varphi_{\mu^* \rightarrow \mu}$ , i.e.,  $\mu = (\nabla\varphi_{\mu^* \rightarrow \mu})\# \mu^*$ . If  $\beta - \alpha < 1$ , then any empirical barycenter  $\mu_n$  of  $P$  satisfies*

$$\mathbb{E}W_2^2(\mu_n, \mu^*) \leq \frac{4\sigma^2}{(1 - \beta + \alpha)^2 n}.$$

### 4.3. Examples

We complete this section by two concrete examples in which we instantiate sufficient conditions for the application of Corollary 4.4.

**4.3.1. Gaussians.** Consider first the metric space of Gaussian distributions over  $H = \mathbb{R}^D$  equipped with the 2-Wasserstein distance.

Assume that  $P$  is supported on nondegenerate Gaussians  $\mathcal{N}(m, \Sigma) \sim P$ .

It is a well-known fact that the optimal transport map between  $\mathcal{N}(m_0, \Sigma_0)$  and  $\mathcal{N}(m_1, \Sigma_1)$  is given by (see [38])

$$x \mapsto \Sigma_0^{-1/2} (\Sigma_0^{1/2} \Sigma_1 \Sigma_0^{1/2})^{1/2} \Sigma_0^{-1/2} (x - m_0) + m_1. \quad (4.2)$$

Using (4.2), the condition of geodesic extendibility translates, by Theorem 4.3, to control on the maximum and minimum eigenvalues of the matrix acting on  $x - m_0$ . A natural

means of establishing this control uniformly for the distribution  $P$  is to assume that each element  $\mathcal{N}(m, \Sigma)$  of its support has covariance matrix  $\Sigma$  with eigenvalues in  $[\kappa_0, \kappa_1]$ . In terms of the space of positive definite matrices, this can be interpreted as ensuring that the support of  $P$  lies away from the boundary while remaining bounded, respectively.

Under this assumption on the support of  $P$ , it can be shown [19, Prop. 15] that the barycenter of  $P$  exists, and is the unique Gaussian  $\mathcal{N}(m_\star, \Sigma_\star)$  where  $m_\star := \int m \, dP(m)$  and  $\Sigma_\star$  solves the first-order optimality equation

$$I_D = \int \Sigma_\star^{-1/2} (\Sigma_\star^{1/2} \Sigma \Sigma_\star^{1/2})^{1/2} \Sigma_\star^{-1/2} \, dP(\Sigma).$$

Using this equation one can directly calculate that  $\Sigma_\star$  must also have eigenvalues in  $[\kappa_0, \kappa_1]$ . By inspecting (4.2), it follows that the optimal transport maps between  $\mathcal{N}(m_\star, \Sigma_\star)$  and each element of the support of  $P$  are  $\kappa^{-1}$ -strongly convex and  $\kappa$ -smooth, where  $\kappa := \kappa_1/\kappa_0$ . Together, this yields the following application of our main result.

**Corollary 4.5.** *Let  $(S, d)$  denote the space of Gaussians metrized by the 2-Wasserstein distance. Fix  $\kappa_1 > \kappa_0 > 0$  and let  $P \in \mathcal{P}_2(S, d)$  be any distribution supported solely on Gaussians  $\mathcal{N}(m, \Sigma)$  with eigenvalues in  $[\kappa_0, \kappa_1]$ . Denote  $\kappa := \kappa_1/\kappa_0$ , and assume  $\kappa - \kappa^{-1} < 1$ . Then  $P$  has a unique barycenter  $\mu^\star = \mathcal{N}(m_\star, \Sigma_\star)$ , and the empirical barycenter  $\mu_n = \mathcal{N}(m_n, \Sigma_n)$  satisfies*

$$\mathbb{E} W_2^2(\mu_n, \mu^\star) \leq \frac{4\sigma^2}{(1 - \kappa + \kappa^{-1})^2 n}.$$

4.3.2. *Template deformation model.* Throughout this paper we have made minimal assumptions on  $P$ , apart from those that ensure convergence of its barycenter. It is customary to define models where the mean, and more generally the barycenter, of  $P$  is a parameter of interest. Template deformation models were introduced as convenient models where the Wasserstein barycenter is the parameter of interest. As we will see below, this model gives an equivalent but complementary perspective on Corollary 4.4.

Recall that since  $(\mathcal{P}_2(\mathbb{R}^D), W_2)$  is positively curved, the barycenter  $\mu^\star$  of a measure  $P \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^D))$  is also a barycenter of the measure  $P$  pushed forward to the tangent space at  $\mu^\star$ ; in other words, the barycenter is also an *exponential* barycenter [32, Corollary 2]. Moreover, the tangent space at a measure  $\mu^\star$  can be identified with the closure in  $L^2(\mu^\star)$  of the set of maps  $T$  that can be written as the gradient of a function—this follows from identifying a tangent vector  $\log_{\mu^\star}(\mu)$  with  $\nabla\varphi_{\mu^\star \rightarrow \mu} - \text{id}$  where  $\nabla\varphi_{\mu^\star \rightarrow \mu}$  is the optimal transport map from  $\mu^\star$  to  $\mu$ ; see [6, §8.4]. In particular,  $\mu^\star$  is an exponential barycenter for  $P$  if and only if

$$\int \nabla\varphi_{\mu^\star \rightarrow \mu}(x) \, dP(\mu) = x \quad \text{for } \mu^\star\text{-almost all } x. \tag{4.3}$$

Exponential barycenters correspond to critical points that are not necessarily global minima of the variance functional. Hence, (4.3) does not imply that  $\mu^\star$  is a barycenter of  $P$ . However, one can easily check that the variance equality (Theorem 3.1) holds also

when  $\mu^*$  is an exponential barycenter. It implies that if (4.3) holds and  $k_{\mu^*}^v(\mu) \geq 0$  for all  $v \in \mathcal{P}_2(\mathbb{R}^D)$  and  $P$ -almost all  $\mu$ , then  $\mu^*$  is also a barycenter of  $P$ .

Wasserstein exponential barycenters as in (4.3) arise naturally as the parameter of interest in the following template deformation model [11, 53]. Let  $T \in L^2(\mathbb{R}^D, \mathbb{R}^D, \mu^*)$  be a random function, often called a *warping function*. In a template deformation model, one observes independent copies of

$$\mu = T_{\#}\mu^*, \quad T \sim Q, \quad (4.4)$$

and the goal is to estimate  $\mu^*$ . Under some conditions on the distribution  $Q$  of  $T$ ,  $\mu^*$  is an exponential barycenter of the distribution of  $\mu$ . For example, if  $T$  is linear and identified to a positive semidefinite  $D \times D$  matrix, it is sufficient to assume the  $\mathbb{E}_Q[T] = I_D$ . More generally, we can assume that  $T$  is the gradient of a convex function. Since  $T_{\#}\mu^* = \mu$ , by Brenier's theorem we have  $T = \nabla\varphi_{\mu^* \rightarrow \mu}$ . In that case,  $\mathbb{E}_Q[T] = I_D$  is replaced with (4.3). Additional regularity conditions on the warping functions ensure that  $\mu^*$  is, in fact, a barycenter.

The following result follows readily from Corollary 4.4.

**Corollary 4.6.** *Fix  $\alpha, \beta > 0$  such that  $\beta - \alpha < 1$  and consider the template deformation model (4.4) where  $Q$  is supported on gradients of  $\alpha$ -strongly convex and  $\beta$ -smooth functions  $\varphi_{\mu^* \rightarrow \mu}$ . Moreover, assume that*

$$\int T(x) \, dQ(T) = x \quad \text{for } \mu^*\text{-almost all } x.$$

*Let  $\mu_n$  denote the empirical barycenter of  $n$  independent copies of  $\mu$  from the template deformation model (4.4). Then, for all  $n \geq 1$ ,*

$$\mathbb{E}W_2^2(\mu_n, \mu^*) \leq \frac{4\sigma^2}{(1 - \beta + \alpha)^2 n},$$

where

$$\sigma^2 = \mathbb{E}W_2^2(\mu, \mu^*) = \int \|T(x) - x\|^2 \, d\mu^*(x) \, dQ(T).$$

*Acknowledgements.* The authors thank Sinho Chewi and anonymous reviewers for helpful remarks. TLG, PR and QP acknowledge the hospitality of the Institute for Advanced Study where most of this work was carried out.

*Funding.* TLG and QP were supported by the Russian Academic Excellence Project 5-100. PR was supported by NSF awards IIS-1838071, DMS-1712596 and DMS-TRIPODS-1740751, grant 2018-182642 from the Chan-Zuckerberg Initiative DAF and the Schmidt Foundation.

## References

- [1] Afsari, B.: Riemannian  $L^p$  center of mass: existence, uniqueness, and convexity. Proc. Amer. Math. Soc. **139**, 655–673 (2011) Zbl [1220.53040](#) MR [2736346](#)

- [2] Agueh, M., Carlier, G.: Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **43**, 904–924 (2011) Zbl [1223.49045](#) MR [2801182](#)
- [3] Agueh, M., Carlier, G.: Vers un théorème de la limite centrale dans l'espace de Wasserstein? *C. R. Math. Acad. Sci. Paris* **355**, 812–818 (2017) Zbl [1388.60061](#) MR [3673057](#)
- [4] Ahidar-Coutrix, A., Le Gouic, T., Paris, Q.: Convergence rates for empirical barycenters in metric spaces: curvature, convexity and extendable geodesics. *Probab. Theory Related Fields* **177**, 323–368 (2020) Zbl [1442.51004](#) MR [4095017](#)
- [5] Alexander, S., Kapovitch, V., Petrunin, A.: Alexandrov geometry: preliminary version no. 1. Book in preparation (2019)
- [6] Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures. *Lectures Math. ETH Zürich*, Birkhäuser, Basel (2005) Zbl [1090.35002](#) MR [2129498](#)
- [7] Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: *International Conference on Machine Learning*, 214–223 (2017)
- [8] Bauschke, H. H., Combettes, P. L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. 2nd ed., CMS Books in Math./Ouvrages de Math. de la SMC, Springer, Cham (2017) Zbl [1359.26003](#) MR [3616647](#)
- [9] Bhattacharya, R., Patrangenaru, V.: Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31**, 1–29 (2003) Zbl [1020.62026](#) MR [1962498](#)
- [10] Bhattacharya, R., Patrangenaru, V.: Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33**, 1225–1259 (2005) Zbl [1072.62033](#) MR [2195634](#)
- [11] Boissard, E., Le Gouic, T., Loubes, J.-M.: Distribution's template estimate with Wasserstein metrics. *Bernoulli* **21**, 740–759 (2015) Zbl [1320.62107](#) MR [3338645](#)
- [12] Boucheron, S., Lugosi, G., Massart, P.: *Concentration Inequalities*. Oxford Univ. Press, Oxford (2013) Zbl [1279.60005](#) MR [3185193](#)
- [13] Burago, D., Burago, Y., Ivanov, S.: *A Course in Metric Geometry*. Grad. Stud. Math. 33, Amer. Math. Soc, Providence, RI (2001) Zbl [0981.51016](#) MR [1835418](#)
- [14] Burago, Y., Gromov, M., Perelman, G.: A. D. Alexandrov spaces with curvature bounded below. *Russian Math. Surveys* **47**, no. 2, 1–58 (1992) Zbl [0802.53018](#) MR [1185284](#)
- [15] Cheeger, J., Colding, T. H.: Lower bounds on Ricci curvature and the almost rigidity of warped products. *Ann. of Math. (2)* **144**, 189–237 (1996) Zbl [0865.53037](#) MR [1405949](#)
- [16] Cheeger, J., Colding, T. H.: On the structure of spaces with Ricci curvature bounded below. I. *J. Differential Geom.* **46**, 406–480 (1997) Zbl [0902.53034](#) MR [1484888](#)
- [17] Cheeger, J., Colding, T. H.: On the structure of spaces with Ricci curvature bounded below. II. *J. Differential Geom.* **54**, 13–35 (2000) Zbl [1027.53042](#) MR [1815410](#)
- [18] Cheeger, J., Colding, T. H.: On the structure of spaces with Ricci curvature bounded below. III. *J. Differential Geom.* **54**, 37–74 (2000) Zbl [1027.53043](#) MR [1815411](#)
- [19] Chewi, S., Maunu, T., Rigollet, P., Stromme, A. J.: Gradient descent algorithms for Bures–Wasserstein barycenters. In: *Conference on Learning Theory*, PMLR, 1276–1304 (2020)
- [20] Cuturi, M., Doucet, A.: Fast computation of Wasserstein barycenters. In: *International Conference on Machine Learning*, 685–693 (2014)
- [21] Eltzner, B., Huckemann, S. F.: A smeary central limit theorem for manifolds with application to high-dimensional spheres. *Ann. Statist.* **47**, 3360–3381 (2019) Zbl [1436.60032](#) MR [4025745](#)
- [22] Émery, M., Mokobodzki, G.: Sur le barycentre d'une probabilité dans une variété. In: *Séminaire de Probabilités, XXV*, Lecture Notes in Math. 1485, Springer, Berlin, 220–233 (1991) Zbl [0753.60046](#) MR [1187782](#)
- [23] Feydy, J., Charlier, B., Vialard, F., Peyré, G.: Optimal transport for diffeomorphic registration. In: *MICCAI 2017, Proceedings, Part I*, 291–299 (2017)
- [24] Genevay, A., Peyré, G., Cuturi, M.: Learning generative models with Sinkhorn divergences. In: *International Conference on Artificial Intelligence and Statistics*, 1608–1617 (2018)

- [25] Halbeisen, S.: On tangent cones of Alexandrov spaces with curvature bounded below. *Manuscripta Math.* **103**, 169–182 (2000) Zbl [1021.53040](#) MR [1796313](#)
- [26] Hotz, T., Huckemann, S., Le, H., Marron, J. S., Mattingly, J. C., Miller, E., Nolen, J., Owen, M., Patrangenaru, V., Skwerer, S.: Sticky central limit theorems on open books. *Ann. Appl. Probab.* **23**, 2238–2258 (2013) Zbl [1293.60006](#) MR [3127934](#)
- [27] Kendall, W. S., Le, H.: Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables. *Braz. J. Probab. Statist.* **25**, 323–352 (2011) Zbl [1234.60025](#) MR [2832889](#)
- [28] Kroshnin, A., Spokoiny, V., Suvorikova, A.: Statistical inference for Bures–Wasserstein barycenters. *Ann. Appl. Probab.* **31**, 1264–1298 (2021) Zbl [1476.60051](#) MR [4278784](#)
- [29] Kwapien, S.: Isomorphic characterizations of inner product spaces by orthogonal series with vector valued coefficients. *Studia Math.* **44**, 583–595 (1972) Zbl [0256.46024](#) MR [341039](#)
- [30] Lang, U., Schroeder, V.: Kirschbraun’s theorem and metric spaces of bounded curvature. *Geom. Funct. Anal.* **7**, 535–560 (1997) Zbl [0891.53046](#) MR [1466337](#)
- [31] Ledoux, M., Talagrand, M.: *Probability in Banach Spaces*. *Ergeb. Math. Grenzgeb.* (3) **23**, Springer, Berlin (1991) Zbl [0748.60004](#) MR [1102015](#)
- [32] Le Gouic, T.: A note on flatness of non separable tangent cone at a barycenter. *C. R. Math. Acad. Sci. Paris* **358**, 489–495 (2020) Zbl [1452.60006](#) MR [4134259](#)
- [33] Le Gouic, T., Loubes, J.-M.: Existence and consistency of Wasserstein barycenters. *Probab. Theory Related Fields* **168**, 901–917 (2017) Zbl [1406.60019](#) MR [3663634](#)
- [34] Lott, J., Villani, C.: Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.* (2) **169**, 903–991 (2009) Zbl [1178.53038](#) MR [2480619](#)
- [35] Ohta, S.-i.: Convexities of metric spaces. *Geom. Dedicata* **125**, 225–250 (2007) Zbl [1140.52001](#) MR [2322550](#)
- [36] Ohta, S.-i.: Barycenters in Alexandrov spaces of curvature bounded below. *Adv. Geom.* **12**, 571–587 (2012) Zbl [1276.53073](#) MR [3005101](#)
- [37] Panaretos, V. M., Zemel, Y.: Statistical aspects of Wasserstein distances. *Ann. Rev. Statist. Appl.* **6**, 405–431 (2019) MR [3939527](#)
- [38] Peyré, G., Cuturi, M.: Computational optimal transport. *Foundations and Trends in Machine Learning* **11**, 355–607 (2019)
- [39] Rigollet, P., Hütter, J.-C.: *High-dimensional statistics*. Lecture notes (2017)
- [40] Rigollet, P., Weed, J.: Entropic optimal transport is maximum-likelihood deconvolution. *C. R. Math. Acad. Sci. Paris* **356**, 1228–1235 (2018) Zbl [1407.62211](#) MR [3907589](#)
- [41] Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., Lee, L., Chen, J., Brumbaugh, J., Rigollet, P., Hochedlinger, K., Jaenisch, R., Regev, A., Lander, E. S.: Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**, 928–943.e22 (2019)
- [42] Schötz, C.: Convergence rates for the generalized Fréchet mean via the quadruple inequality. *arXiv:1812.08037* (2018)
- [43] Solomon, J., de Goes, F., Peyré, G., Cuturi, M., Butscher, A., Nguyen, A., Du, T., Guibas, L.: Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Trans. Graph.* **34**, art. 66, 11 pp. (2015) Zbl [1334.68267](#)
- [44] Staib, M., Claiici, S., Solomon, J. M., Jegelka, S.: Parallel streaming Wasserstein barycenters. In: *Advances in Neural Information Processing Systems*, 2644–2655 (2017)
- [45] Sturm, K.-T.: Probability measures on metric spaces of nonpositive curvature. In: *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces* (Paris, 2002), *Contemp. Math.* **338**, Amer. Math. Soc., Providence, RI, 357–390 (2003) Zbl [1040.60002](#) MR [2039961](#)
- [46] Sturm, K.-T.: On the geometry of metric measure spaces. I. *Acta Math.* **196**, 65–131 (2006) Zbl [1105.53035](#) MR [2237206](#)
- [47] Sturm, K.-T.: On the geometry of metric measure spaces. II. *Acta Math.* **196**, 133–177 (2006) Zbl [1106.53032](#) MR [2237207](#)

- 
- [48] Sturm, K.-T.: The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces. arXiv:[1208.0434](#) (2012)
  - [49] Vershynin, R.: High-dimensional Probability. Cambridge Ser. Statist. Probab. Math. 47, Cambridge Univ. Press, Cambridge (2018) Zbl [1430.60005](#) MR [3837109](#)
  - [50] Villani, C.: Optimal Transport. Grundlehren Math. Wiss. 338, Springer, Berlin (2009) Zbl [1156.53003](#) MR [2459454](#)
  - [51] Yokota, T.: Convex functions and barycenter on CAT(1)-spaces of small radii. J. Math. Soc. Japan **68**, 1297–1323 (2016) Zbl [1351.53057](#) MR [3523548](#)
  - [52] Yokota, T.: Convex functions and  $p$ -barycenter on CAT(1)-spaces of small radii. Tsukuba J. Math. **41**, 43–80 (2017) Zbl [1378.53056](#) MR [3705774](#)
  - [53] Zemel, Y., Panaretos, V. M.: Fréchet means and Procrustes analysis in Wasserstein space. Bernoulli **25**, 932–976 (2019) Zbl [1431.62132](#) MR [3920362](#)