



**HAL**  
open science

# On the rate of convergence of empirical barycentres in metric spaces: curvature, convexity and extendible geodesics

Adil Ahidar-Coutrix, Thibaut Le Gouic, Quentin Paris

## ► To cite this version:

Adil Ahidar-Coutrix, Thibaut Le Gouic, Quentin Paris. On the rate of convergence of empirical barycentres in metric spaces: curvature, convexity and extendible geodesics. 2018. hal-01810530v1

**HAL Id: hal-01810530**

**<https://hal.science/hal-01810530v1>**

Preprint submitted on 8 Jun 2018 (v1), last revised 17 Jun 2019 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On the rate of convergence of empirical barycentres in metric spaces: curvature, convexity and extendible geodesics

Adil Ahidar-Coutrix\*, Thibaut Le Gouic† and Quentin Paris‡

June 7, 2018

## Abstract

This paper provides rates of convergence for empirical barycentres of a Borel probability measure on a metric space under general conditions. Our results are given in the form of sharp oracle inequalities. Our main assumption, of geometrical nature, is shown to be satisfied at least in two meaningful scenarios. The first one is a form of weak curvature constraint of the underlying space referred to as  $(k, \alpha)$ -convexity, compatible with a positive upper curvature bound. The second scenario considers the case of a nonnegatively curved space on which geodesics, emanating from a barycentre, can be extended.

## 1 Introduction

Given a metric space  $(S, d)$ , define  $\mathcal{P}_2(S)$  as the set of Borel probability measures on  $S$  such that

$$\int_S d(x_0, x)^2 dP(x) < +\infty,$$

for some (hence all)  $x_0 \in S$ . A barycentre of  $P \in \mathcal{P}_2(S)$  (also called a Fréchet mean) is any element  $b^* \in S$  such that

$$b^* \in \arg \min_{b \in S} \int_S d(x, b)^2 dP(x).$$

A barycentre stands as a (one of the possibly many) natural analog of (and coincides with) the mean of a square integrable probability measure on  $\mathbf{R}^d$ . Extending the notion of mean value of a probability measure to non Euclidean (or non Hilbert) spaces  $S$  has a number of applications ranging from Geometry and Optimal Transport to Statistics and the context of abstract metric spaces provides a unifying framework encompassing most non-standard settings. Barycentres in metric spaces have been widely studied and are well understood in several scenarios including the case where  $S$  has nonpositive curvature (see, for instance, [Sturm, 2003](#)) or more generally upper bounded curvature ([Yokota, 2016, 2017](#)) (here and throughout, curvature of a metric spaces is understood in the sense of

---

\*Aix Marseille Univ., CNRS, Centrale Marseille, I2M, Marseille, France.  
Email:adil.ahidar@outlook.com

†Aix Marseille Univ., CNRS, Centrale Marseille, I2M, Marseille, France.  
Email:thibaut.le\_gouic@math.cnrs.fr

‡National Research University Higher School of Economics, Russian Federation & International Laboratory of Stochastic Algorithms and High-Dimensional Inference. This work has been funded by the Russian Academic Excellence Project '5-100'. Email:qparis@hse.ru

Alexandrov). Significant contributions have also focused on the more delicate context of spaces  $S$  with lower bounded curvature (Yokota, 2012; Ohta, 2012). Focus on this last context arises, in particular, from the increasing interest for questions related to Wasserstein spaces (and optimal transport), *i.e.* the case where  $(S, d) = (\mathcal{P}_2(E), W_2)$  (or more generally  $(\mathcal{P}_p(E), W_p)$  for  $p \geq 1$ ) for some auxiliary metric space  $E$ . Indeed, provided  $E$  has nonnegative curvature (which includes the important setting where  $E$  is a Hilbert space), the space  $(\mathcal{P}_2(E), W_2)$  is itself known to have nonnegative curvature (see Section 7.3 in Ambrosio et al., 2008 for the case where  $E$  is a Hilbert space and Appendix A in Lott and Villani, 2009 for the case where  $E$  is a compact Riemannian manifold).

The performance of estimators of barycentres in metric spaces is less fully understood. Two remarkable trends are however modifying rapidly the landscape. On the one hand, recent algorithmic breakthroughs in computational optimal transport (see, for instance, Cuturi, 2013; Benamou et al., 2015; Genevay et al., 2016, and the references therein) are opening the way for the efficient computation of barycentres of probability measures on Wasserstein spaces. From a more theoretical perspective, many contributions, old and new, have and are still improving our understanding of these algorithmic successes and the statistical behavior of estimating procedures in general (we refer to Bhattacharya and Patrangenaru, 2003; Kendall and Le, 2011; Le Gouic and Loubes, 2017; Altschuler et al., 2017; Dvurechensky et al., 2018; Bigot et al., 2018, for works in this direction).

Our present contribution is in the spirit of this second trend and focuses on providing non-asymptotic error bounds for empirical barycentres defined as follows. Throughout, we suppose that  $P$  has bounded support. Given  $X_1, \dots, X_n$  independent and identically distributed (i.i.d) with common distribution  $P$ , the empirical distribution  $P_n$  is constructed as

$$P_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}. \quad (1)$$

Then, we consider an empirical barycentre of  $P$  to be any point

$$b_n \in \arg \min_{b \in B} \int_S d(x, b)^2 dP_n(x) = \arg \min_{b \in B} \frac{1}{n} \sum_{i=1}^n d(X_i, b)^2, \quad (2)$$

for some  $B \subset S$  of bounded diameter. The need for restricting the computation of  $b_n$  over some set  $B$  of bounded diameter, as well as imposing the support of  $P$  to be bounded, appears as a limitation of our proof techniques. We believe that these limitations can be, in principle, removed via a truncation approach at the price of additional technicality. Avoiding these technicalities allows, however, to simply isolate the main geometrical insights and hopefully account for some practical considerations as the minimization step over  $S$  may be computationally unrealistic.

The paper is organized as follows. In Section 2, we provide a general convergence result with explicit rates of convergence for empirical barycentres under general conditions. Under the assumptions considered in this section, namely Assumptions **(A1)** and **(A2)**, the proof of our main result builds essentially upon the theory of empirical processes. Section 3 studies conditions, of geometrical nature, under which our main assumption, Assumption **(A1)**, is satisfied. The proofs of all results presented in the text are reported in Section 4. Appendix A provides additional bounds in expectation based of our main result, Theorem 1, avoided in the body of the article for better readability. Finally, the

material presented in Section 3 builds upon some background in metric geometry gathered in Appendix B.

## 2 General convergence result

Let  $(S, d)$  be a metric space. Let  $P \in \mathcal{P}_2(S)$  and let  $b^* \in S$  be a barycentre of  $P$ , i.e.

$$b^* \in \arg \min_{b \in S} \int_S d(x, b)^2 dP(x).$$

Consider i.i.d. random variables  $X_1, \dots, X_n$  with distribution  $P$ , and let  $P_n$  be the associated empirical measure defined as in (1). For any fixed subset  $B \subset S$  define  $b_n$  as in (2) and consider

$$\bar{b} \in \arg \min_{b \in B} \int_S d(x, b)^2 dP(x),$$

where the dependence of  $b_n$  and  $\bar{b}$  on  $B$  is omitted in the notation. Throughout the paper, we consider the following assumptions.

**(A1)** *There exists constants  $K > 0$  and  $\beta \in (0, 1]$  such that, for all  $b \in B$ ,*

$$d(b, \bar{b})^2 \leq K \left( \int_S (d(b, x)^2 - d(\bar{b}, x)^2) dP(x) \right)^\beta.$$

**(A2)** *There exists constants  $C > 0$  and  $D > 0$  such that, for all  $0 < \varepsilon \leq r \leq \text{diam}(B)$ ,*

$$\log N(B \cap B(\bar{b}, r), d, \varepsilon) \leq \left( \frac{Cr}{\varepsilon} \right)^D.$$

**Note:** For any set  $A \subset S$  and any  $\varepsilon > 0$ , notation  $N(A, d, \varepsilon)$  stands for the  $\varepsilon$ -covering number of  $A$  in  $S$ , defined as the minimal number  $m$  of points  $x_1, \dots, x_m \in S$  needed for condition

$$A \subset \bigcup_{i=1}^m B(x_i, \varepsilon),$$

to hold, with  $B(x, \varepsilon) := \{y \in S : d(x, y) \leq \varepsilon\}$ .

Next is the main result of this section.

**Theorem 1.** *Suppose that  $(S, d)$  is a separable metric space and that the support of  $P$ , and the set  $B$ , have bounded diameter. Suppose that Assumptions **(A1)** and **(A2)** hold. Then for all  $n \geq 1$  and all  $t > 0$ ,*

$$d(b_n, \bar{b})^{\frac{2}{\beta}} \leq \int_S (d(b_n, x)^2 - d(\bar{b}, x)^2) dP(x) \leq A \cdot \max \left\{ \theta_n, \left( \frac{t}{n} \right)^{\frac{1}{2-\beta}}, \frac{t}{n} \right\},$$

with probability at least  $1 - 2e^{-t}$ , where

$$\theta_n = \begin{cases} n^{-\frac{1}{2-\beta}} & \text{if } D < 2, \\ (\log n)^2 n^{-\frac{1}{2-\beta}} & \text{if } D = 2, \\ n^{-\frac{2}{D(2-\beta)}} & \text{if } D > 2, \end{cases}$$

and  $A$  is an explicit constant available in the proof depending only on  $K, \beta, C, D, \text{diam}(B)$  and  $\max\{d(x, y) : x \in B, y \in \text{supp}(P)\}$ .

Note that more detailed results, namely Statements 19, 20 and 21 are provided in the proof, as well as a bound in expectation reported in Appendix A. The bounds of Theorem 1 may be simply translated into so called oracle inequalities involving the approximation properties of the chosen set  $B$  relative to  $b^*$ . Indeed, it follows directly from the previous bounds, and the definition of  $\bar{b}$ , that

$$\begin{aligned} & \int_S (d(b_n, x)^2 - d(b^*, x)^2) dP(x) \\ &= \inf_{b \in B} \int_S (d(b, x)^2 - d(b^*, x)^2) dP(x) + \int_S (d(b_n, x)^2 - d(\bar{b}, x)^2) dP(x) \\ &\leq \inf_{b \in B} \int_S (d(b, x)^2 - d(b^*, x)^2) dP(x) + A \cdot \max \left\{ \theta_n, \left( \frac{t}{n} \right)^{\frac{1}{2-\beta}}, \frac{t}{n} \right\}, \end{aligned}$$

and

$$d(b_n, b^*) \leq d(\bar{b}, b^*) + d(b_n, \bar{b}) \leq d(\bar{b}, b^*) + A^{\frac{\beta}{2}} \cdot \max \left\{ \theta_n, \left( \frac{t}{n} \right)^{\frac{1}{2-\beta}}, \frac{t}{n} \right\}^{\frac{\beta}{2}},$$

hold with probability  $1 - 2e^{-t}$ . The approximation terms on the right hand side of these inequalities, namely

$$\inf_{b \in B} \int_S (d(b, x)^2 - d(b^*, x)^2) dP(x) \quad \text{and} \quad d(\bar{b}, b^*),$$

account for the misspecification of set  $B$ . In particular, both terms vanish if  $b^* \in B$  since one may take  $\bar{b} = b^*$  in this case by definition.

While Assumption **(A1)** might seem esoteric at first sight, this condition is interesting in several respects. For the reader more familiar with metric geometry, this condition will appear to be in the same flavour as so called variance inequalities in spaces with bounded curvature (see, for instance, Proposition 4.4 in [Sturm, 2003](#)). At the same time, the author familiar with modern developments in the theory of empirical processes, will identify in the proof of Theorem 1 (see Remark 18) that this condition is instrumental in obtaining fast rates of convergence as it implies a so called Bernstein-type condition on the class of functions indexing our empirical process. In particular, this assumption may be understood as a version of the Mammen-Tsybakov low-noise assumption ([Mammen and Tsybakov, 1999](#)) used in binary classification (See also section 5.2 in the survey by [Boucheron et al., 2005](#)). A significant part of the present paper is precisely devoted to studying conditions (connected to the geometry of sets  $S$ ,  $B$  and  $\text{supp}(P)$ ) allowing for this inequality to hold. Assumption **(A2)** is imposing a rather mild complexity restriction on the neighbourhood of  $\bar{b}$  in  $B$ . This condition allows for a wide range of possibilities including obviously the case where  $B$  has finite (metric) dimension but also meaningful examples where  $B$  has infinite dimension.

**Remark 2.** *The flexibility provided by Assumption **(A2)** could be of special interest in the context where  $S = \mathcal{P}_2(\mathbf{R}^d)$  is equipped with the Wasserstein metric  $W_2$  and  $B = \mathcal{P}_2(B(x_0, \rho))$  for some fixed  $x_0 \in \mathbf{R}^d$  and some fixed radius  $\rho > 0$ . To illustrate this statement, note that by combining the result of Appendix A in [Bolley et al. \(2007\)](#) with a classical bound on the covering number of Euclidean balls, it follows that for all  $0 < \varepsilon \leq \rho$ ,*

$$\log N(B, W_2, \varepsilon) \leq 2 \left( \frac{6\rho}{\varepsilon} \right)^d \log \left( \frac{8e\rho}{\varepsilon} \right).$$

In particular, for any  $D > d$ , there exists  $C_D > 0$  such that, for all  $0 < \varepsilon \leq \rho$ ,

$$\log N(B, W_2, \varepsilon) \leq \left( \frac{C_D \rho}{\varepsilon} \right)^D.$$

To complete the discussion, observe that Assumption **(A2)** is actually imposing that, for all  $0 < \varepsilon \leq r \leq \rho$ ,

$$\log N(B \cap B(\bar{b}, r), W_2, \varepsilon) \leq \left( \frac{Cr}{\varepsilon} \right)^D,$$

where it is understood here that  $B(\bar{b}, r)$  is a ball in  $\mathcal{P}_2(\mathbf{R}^d)$  of radius  $r$  around  $\bar{b} \in B$ . The previous discussion therefore guarantees that a sufficient condition for this property to hold is that there exists  $C' > 0$  and points  $(x_r)_{0 < r \leq \rho}$  in  $\mathbf{R}^d$  such that, for all  $0 < \varepsilon \leq r \leq \rho$ ,

$$B \cap B(\bar{b}, r) \subset \mathcal{P}_2(B(x_r, C'r)).$$

We leave the identification of bounded subsets  $B \subset \mathcal{P}_2(\mathbf{R}^d)$  for which this last (sufficient) condition holds as an interesting question for future work. To conclude, note however that a simple modification of our proof reveals that results in the spirit of Theorem 1 can actually be obtained imposing only a more classical bound of the type

$$\log N(B, d, \varepsilon) \leq \left( \frac{C\rho}{\varepsilon} \right)^D,$$

for all  $0 < \varepsilon \leq \rho = \text{diam}(B)$ . This adaptation is left to the reader.

### 3 Convexity, curvature and extendible geodesics

This section investigates scenarios in which our main assumption **(A1)** holds, connecting it to notions of convexity, curvature and the possibility of extending geodesics. Note that the results considered next rely on ideas from metric geometry outlined in Appendix B for convenience. We start with a few definitions.

**Definition 3** (Convex set). *Suppose that  $(S, d)$  is a geodesic space. A subset  $B \subset S$  is said to be convex if for any two points  $x, y \in B$ , and any (constant speed) shortest path  $\gamma_{xy} : [0, 1] \rightarrow S$  joining  $x$  to  $y$ , we have  $\gamma_{xy}([0, 1]) \subset B$ .*

**Definition 4** ( $(k, \alpha)$ -convex function). *Suppose that  $(S, d)$  is a geodesic space and  $B \subset S$  is convex in the sense of Definition 3. Then, for  $k \geq 0$  and  $\alpha \geq 0$ , a function  $f : B \rightarrow \mathbf{R}$  is said  $(k, \alpha)$ -convex if for any two points  $x, y \in B$ , and any (constant speed) shortest path  $\gamma_{xy} : [0, 1] \rightarrow S$  joining  $x$  to  $y$ , the function*

$$t \in [0, 1] \mapsto f(\gamma_{xy}(t)) - \frac{k}{2}t^2d(x, y)^\alpha,$$

*is convex on  $[0, 1]$ .*

Note that, equivalently, the function  $f : B \rightarrow \mathbf{R}$  is said  $(k, \alpha)$ -convex if for any two points  $x, y \in B$ , and any (constant speed) shortest path  $\gamma_{xy} : [0, 1] \rightarrow S$  joining  $x$  to  $y$ , we have

$$\forall t \in [0, 1], \quad f(\gamma_{xy}(t)) \leq (1-t)f(x) + tf(y) - \frac{k}{2}t(1-t)d(x, y)^\alpha.$$

**Definition 5** ( $(k, \alpha)$ -convex set). *Suppose that  $(S, d)$  is a geodesic space. For  $k \geq 0$ ,  $\alpha \geq 0$  and  $B \subset S$ , we say that set  $B$  is a  $(k, \alpha)$ -convex set if it is convex in the sense of Definition 3 and if, for all  $p \in B$ , the function  $d(p, \cdot)^2 : B \rightarrow [0, +\infty)$  is  $(k, \alpha)$ -convex in the sense of Definition 4.*

Definition 5 is a slight generalisation of the definition of a  $k$ -convex space introduced by Ohta (2007). A  $k$ -convex metric space  $(S, d)$ , in the sense of Ohta (2007), is a  $(k, 2)$ -convex set in the sense of Definition 5. Hence, we simply introduce an additional degree of freedom through the introduction of parameter  $\alpha \geq 0$  and the possibility of focusing on convex subsets  $B$  of an arbitrary metric space  $S$ . Note that, if  $\text{diam}(B) \leq 1$  and  $\alpha > 2$ , every  $(k, 2)$ -convex set  $B$  is also  $(k, \alpha)$ -convex. Hence  $\alpha > 2$  introduces more flexibility for small sets compared to Ohta's definition. On the contrary if  $\text{diam}(B) \geq 1$  and  $\alpha > 2$ ,  $(k, \alpha)$ -convexity is actually imposing a more restrictive condition on geodesics connecting points far apart so that there are a priori no possible comparison between  $(k, 2)$  and  $(k, \alpha)$  convex sets of diameter at least 1. Finally, note that when  $\alpha < 2$ , the concavity of  $x \geq 0 \mapsto x^\alpha$  makes the notion of  $(k, \alpha)$ -convexity less relevant. The next result connects the notion of  $(k, \alpha)$ -convexity to Assumption (A1).

**Theorem 6.** *If  $B$  is  $(k, \alpha)$ -convex in the sense of Definition 5 for some  $k > 0$  and some  $\alpha \geq 2$ , then Assumption (A1) holds with  $K = 2/k$  and  $\beta = 2/\alpha$ .*

As clearly pointed out by Ohta (2007) in the context of a  $(k, 2)$ -convex set, the  $(k, \alpha)$ -convexity of a set  $B$  is connected to its curvature properties. For instance, recall that a geodesic space  $(S, d)$  is said to have curvature bounded from above by 0 in the large or globally (in this case  $S$  is also called a CAT(0) space) if  $S$  is a  $(2, 2)$ -convex set in the sense of Definition 5. To insist on the practical interest of the previous result, we therefore mention without proof the following.

**Corollary 7.** *Assumption (A1), with  $K = 1$  and  $\beta = 1$ , is therefore satisfied in particular for any convex subset of a CAT(0) (or global NPC) space such as a Banach space, the hyperbolic plane, a riemannian manifold with sectional curvature upper bounded by 0 or any (metric) tree.*

The next result, in the spirit of Proposition 3.1 in Ohta (2007), exhibits further the link between curvature and the notion of  $(k, \alpha)$ -convexity.

**Theorem 8.** *Suppose that  $B$  is convex, complete, with  $d := \text{diam}(B) < \pi/2$ , and with curvature upper bounded by some 1 in the large (i.e.  $B$  is a CAT(1) space). Then for any  $\alpha \geq 2$ ,  $B$  is  $(k(\alpha), \alpha)$ -convex for*

$$k(\alpha) = \frac{8d}{d^{\alpha-2}} \tan\left(\frac{\pi}{2} - d\right).$$

*Hence, for all  $\alpha \geq 2$ , Assumption (A1) is in this case satisfied for  $K = 2/k(\alpha)$  and  $\beta = 2/\alpha$ .*

The next result (in the spirit of Theorem 4.9 in Sturm, 2003) connects  $(k, 2)$ -convexity to a series of comparable statements.

**Theorem 9.** *Let  $(S, d)$  be a geodesic space and fix  $k > 0$ . Then we have (a)  $\Rightarrow$  (b)  $\Rightarrow$  (c)  $\Rightarrow$  (d), where*

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

$$d(b, b^*)^2 \leq \int_S d(x, b)^2 dP(x) - \frac{k}{2} \int_S d(x, b^*)^2 dP(x). \quad (3)$$

(b)  $S$  is  $(k, 2)$ -convex.

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

$$\frac{k}{2}d(b, b^*)^2 \leq \int_S (d(x, b)^2 - d(x, b^*)^2) dP(x). \quad (4)$$

(d) For all  $x, y, b \in S$ , any constant speed shortest path  $\gamma_{xy} : [0, 1] \rightarrow S$  from  $x$  to  $y$  and all  $t \in [0, 1]$ ,

$$\frac{k}{2}d(b, \gamma_{xy}(t))^2 \leq (1-t)d(b, x)^2 + td(b, y)^2 - t(1-t)d(x, y)^2.$$

In addition, if  $k \leq 2$  and if  $d(b, b^*)^2 \leq \int_S d(x, b^*)^2 dP(x)$ , then (4)  $\Rightarrow$  (3).

The next results of this section provide a different perspective in the context where the ambient space  $S$  has nonnegative curvature and builds upon the geometric properties of the tangent cone  $T_{b^*}S$  of  $S$  at a barycentre  $b^*$  of  $P \in \mathcal{P}_2(S)$ . The reader unfamiliar with the notion of a tangent cone is referred to Appendix B which in particular reports essential definitions and basic properties of this delicate object.

**Definition 10.** Let  $(S, d)$  be a geodesic space. For  $x \in S$ , denote  $C(x)$  the set of points  $y \in S$  such that there is strictly more than one constant speed shortest path connecting  $x$  to  $y$ .

Note that  $C(x) = \emptyset$  if  $S$  is a global NPC space (*i.e.* has curvature upper bounded by 0 in the large). However, for spaces of nonnegative curvature,  $C(x)$  may be nonempty. A notion, related to that of a barycentre, is that of an exponential barycentre (Émery and Mokobodzki, 1991) defined as follows.

**Definition 11** (Exponential barycentre). Let  $S$  be a global NNC space and  $P \in \mathcal{P}_2(S)$ . A point  $b \in S$  is said to be an exponential barycentre of  $P$  if

$$\int_S \int_S \langle \log_b(x), \log_b(y) \rangle_b dP(x)dP(y) = 0.$$

For the definition of  $\log_b : S \setminus C(x) \rightarrow T_bS$  and  $\langle \cdot, \cdot \rangle_b$  we refer the reader to Appendix B. The definition of an exponential barycentre mimics that of the Pettis-integral of a Hilbert-valued function and stands as one of the many possible ways to define the analog of the mean value of an element of  $\mathcal{P}_2(S)$ . An important result, Theorem A in Yokota (2012), states that if  $b$  is an exponential barycentre of  $P \in \mathcal{P}_2(S)$ , and  $S$  is a global NNC space, then the subset  $\text{supp}((\log_b)_\#P)$  of the tangent cone  $T_bS$  can be isometrically embedded into a Hilbert space and that the (closed) convex hull of that set in  $T_bS$  is also isometric to that of its image. This geometrical characterisation of the tangent cone at an exponential barycentre allows to conduct computations in a flat space, much easier to handle. As it turns out, any barycentre, in the sense considered so far, is also an exponential barycentre.

**Theorem 12.** Suppose that  $(S, d)$  is a global NNC space and let  $P \in \mathcal{P}_2(S)$ . Then if  $b^*$  is a barycentre of  $P$ , it is an exponential barycentre of  $P$ .

The next result exploits this last fact to provide a variance inequality in global NNC spaces.



**Theorem 13.** *Suppose  $(S, d)$  is a global NNC space. Let  $P \in \mathcal{P}_2(S)$ . Let  $b^* \in S$  be a barycentre of  $P$ . For all  $x, b \in \text{supp}(P) \setminus C(b^*)$ , define*

$$k_0(x, b, b^*) := 1 - \frac{\|\log(x) - \log(b)\|^2 - d^2(x, b)}{d^2(b, b^*)}.$$

where we have denoted  $\log := \log_{b^*}$  and  $\|\cdot\| := \|\cdot\|_{b^*}$ . Then, provided  $P(C(b^*)) = 0$ , we obtain

$$d(b, b^*)^2 \int_S k_0(x, b, b^*) dP(x) \leq \int_S (d(b, x)^2 - d(b^*, x)^2) dP(x).$$

The previous inequality indicates that, provided

$$\mathfrak{K}_0 := \inf_{b \in \text{supp}(P)} \int_S k_0(x, b, b^*) dP(x) > 0,$$

then Assumption **(A1)** holds with  $K = \mathfrak{K}_0^{-1}$ ,  $\beta = 1$  and  $B = S$ . It is worth noticing that in the context where  $S$  is a global NPC, the quantity  $k_0(x, b, b^*)$  can also be defined in a similar fashion and satisfies  $k_0(x, b, b^*) \geq 1$ . In particular, we recover in this case Proposition 4.4 in [Sturm \(2003\)](#). The last result of the section provides a sufficient condition for  $\mathfrak{K}_0 > 0$  to hold in a global NNC space.

**Theorem 14.** *Let  $(S, d)$  be a global NNC space. Let  $P \in \mathcal{P}_2(S)$  and let  $b^* \in S$  be a barycentre of  $P$ . Fix  $\lambda > 0$ . Suppose that, for  $P$ -almost every  $x$  in  $S$ , the two following properties holds.*

- *Any constant speed geodesic  $\gamma_x : [0, 1] \rightarrow S$  joining  $b^*$  to  $x$  can be extended to a function  $\gamma_x^+ : [0, 1 + \lambda] \rightarrow S$  that remains a shortest path between its endpoints.*
- *Let  $X$  be an  $S$ -valued random variable with distribution  $P$  and let  $P_\lambda$  be the distribution of  $\gamma_X^+(1 + \lambda)$  where  $\gamma_x^+$  is as above. Then  $b^*$  is the unique barycentre of  $P_\lambda$ .*

Then, for all  $b \in S$ ,

$$d(b, b^*)^2 \leq \frac{1 + \lambda}{\lambda} \int_S (d(b_n, x)^2 - d(b^*, x)^2) dP(x),$$

and Assumption **(A1)** holds for  $B = S$  with  $K = (1 + \lambda)/\lambda$  and  $\beta = 1$ .

## 4 Proofs

### 4.1 Proof of Theorem 1

The proof of Theorem 1 is based on three auxiliary results. The first lemma below provides an upper bound on the largest fixed point of a random nonnegative function. The proof follows from a combination of arguments presented in Theorem 4.1, Corollary 4.1 and Theorem 4.3 in [Koltchinskii \(2011\)](#).

**Lemma 15** ([Koltchinskii, 2011](#)). *Let  $\{\varphi(\delta) : \delta \geq 0\}$  be non-negative random variables (indexed by all deterministic  $\delta \geq 0$ ) such that, almost surely,  $\varphi(\delta) \leq \varphi(\delta')$  if  $\delta \leq \delta'$ . Let  $\{\beta(\delta, t) : \delta \geq 0, t \geq 0\}$ , be (deterministic) real numbers such that  $\beta(\delta, t) \leq \beta(\delta, t')$ , as soon as  $t \leq t'$ , and such that*

$$\mathbf{P}(\varphi(\delta) \geq \beta(\delta, t)) \leq e^{-t}.$$

Finally, let  $\hat{\delta}$  be a nonnegative random variable, a priori upper bounded by a constant  $\bar{\delta} > 0$ , and such that, almost surely,

$$\hat{\delta} \leq \varphi(\bar{\delta}).$$

Then, defining, for all  $t \geq 0$ ,

$$\beta(t) := \inf \left\{ \alpha > 0 : \sup_{\delta \geq \alpha} \frac{\beta(\delta, \frac{t\delta}{\alpha})}{\delta} \leq 1 \right\},$$

we obtain, for all  $t \geq 0$ ,

$$\mathbf{P}(\hat{\delta} \geq \beta(t)) \leq 2e^{-t}.$$

The second lemma we need is due to [Bousquet \(2002\)](#) and improves upon the work of [Talagrand \(1996\)](#) by providing explicit constants. Given a metric space  $(S, d)$  and a family  $\mathcal{F}$  of functions  $f : S \rightarrow \mathbf{R}$ , denote

$$|P - P_n|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} (P - P_n)f \quad \text{and} \quad \sigma_{\mathcal{F}}^2 := \sup_{f \in \mathcal{F}} P(f - Pf)^2.$$

**Lemma 16** ([Bousquet, 2002](#)). *Suppose that all functions in  $\mathcal{F}$  are  $[a, b]$ -valued, for some  $a < b$ . Then, for all  $n \geq 1$  and all  $t > 0$ ,*

$$|P - P_n|_{\mathcal{F}} \leq \mathbf{E}|P - P_n|_{\mathcal{F}} + \sqrt{\frac{2t}{n} (\sigma_{\mathcal{F}}^2 + 2(b-a) \mathbf{E}|P - P_n|_{\mathcal{F}})} + \frac{(b-a)t}{3n},$$

with probability larger than  $1 - e^{-t}$ .

For background on empirical processes, including the proof of [Lemma 16](#), we refer the reader to [Giné and Nickl \(2015\)](#). Finally, the third result we need is the following generalized version of Dudley's entropy bound (see, for instance, [Theorem 5.31 in van Handel, 2016](#)).

**Lemma 17.** *Let  $(X_t)_{t \in E}$  be a real valued process indexed by a (pseudo) metric space  $(E, d)$ . Suppose that the three following conditions hold.*

- (Separability) *There exists a countable subset  $E' \subset E$  such that, for any  $t \in E$ ,*

$$X_t = \lim_{t' \rightarrow t, t' \in E'} X_{t'}, \quad a.s.$$

- (Subgaussian process) *For all  $s, t \in E$ ,  $X_s - X_t$  is subgaussian in the sense that*

$$\forall \theta \in \mathbf{R}, \quad \log \mathbf{E} \exp(\theta \{X_s - X_t\}) \leq \frac{\theta^2 d(s, t)^2}{2}.$$

- (Lipschitz property) *There exists a random variable  $C$  such that, for all  $s, t \in E$ ,*

$$|X_s - X_t| \leq Cd(s, t), \quad a.s.$$

Then, for any  $T \subset E$  and any  $\varepsilon \geq 0$ , we have

$$\mathbf{E} \sup_{t \in T} X_t \leq 2\varepsilon \mathbf{E}[C] + 12 \int_{\varepsilon}^{+\infty} \sqrt{\log N(T, d, u)} du.$$

We are now in position for proving Theorem 1.

*Proof of Theorem 1.* For any  $\delta \geq 0$ , denote

$$B(\delta) := \{b \in B : P(d(\cdot, b)^2 - d(\cdot, \bar{b})^2) \leq \delta\},$$

and

$$\varphi_n(\delta) := \sup\{(P - P_n)(d(\cdot, b)^2 - d(\cdot, \bar{b})^2) : b \in B(\delta)\}.$$

Let us abbreviate,

$$\delta_n := P(d(\cdot, b_n)^2 - d(\cdot, \bar{b})^2),$$

for simplicity. By definition of  $b_n$ ,  $P_n(d(\cdot, b_n)^2 - d(\cdot, \bar{b})^2) \leq 0$ . Hence,

$$\begin{aligned} \delta_n &\leq (P - P_n)(d(\cdot, b_n)^2 - d(\cdot, \bar{b})^2) \\ &\leq \varphi_n(\delta_n). \end{aligned} \tag{5}$$

As a result, in order to provide an upper bound on  $\delta_n$  holding with high probability, it is enough to provide an upper bound on  $\varphi_n(\delta)$ , for fixed  $\delta \geq 0$ , and apply Lemma 15. Invoking Lemma 16, denoting

$$\sigma^2(\delta) := \sup\{P(d(\cdot, b)^2 - d(\cdot, \bar{b})^2)^2 : b \in B(\delta)\},$$

letting  $\rho := 2\text{diam}(B) \max\{d(x, y) : x \in B, y \in \text{supp}(P)\}$  and observing that, for  $x \in \text{supp}(P)$ ,

$$|d(x, b)^2 - d(x, \bar{b})^2| \leq \rho,$$

the inequality

$$\varphi_n(\delta) \leq \mathbf{E}\varphi_n(\delta) + \sqrt{\frac{2t}{n} (\sigma^2(\delta) + 4\rho \mathbf{E}\varphi_n(\delta))} + \frac{2\rho t}{3n},$$

holds with probability at least  $1 - e^{-t}$ . Using that  $\sqrt{u+v} \leq \sqrt{u} + \sqrt{v}$  and that  $2\sqrt{ab} \leq a+b$ , we further deduce that

$$\varphi_n(\delta) \leq 2\mathbf{E}\varphi_n(\delta) + \sigma(\delta) \sqrt{\frac{2t}{n}} + \frac{8\rho t}{3n}, \tag{6}$$

with probability at least  $1 - e^{-t}$ . Let us now focus on upper bounding  $\sigma(\delta)$ . For  $x \in \text{supp}(P)$ ,

$$\begin{aligned} (d(x, b)^2 - d(x, \bar{b})^2)^2 &= (d(x, b) + d(x, \bar{b}))^2 (d(x, b) - d(x, \bar{b}))^2 \\ &\leq (d(x, b) + d(x, \bar{b}))^2 d(b, \bar{b})^2 \\ &\leq 4 \max\{d(u, y)^2 : u \in B, y \in \text{supp}(P)\} d(b, \bar{b})^2, \end{aligned}$$

and therefore, using Assumption **(A1)**,

$$\begin{aligned} P(d(\cdot, b)^2 - d(\cdot, \bar{b})^2)^2 &\leq 4 \max\{d(u, y)^2 : u \in B, y \in \text{supp}(P)\} d(b, \bar{b})^2 \\ &\leq 4RK(P(d(\cdot, b)^2 - d(\cdot, \bar{b})^2))^\beta, \end{aligned} \tag{7}$$

where we have denoted  $R := \max\{d(x, y) : x \in B, y \in \text{supp}(P)\}$ .

**Remark 18.** We briefly interrupt the proof for an important comment on a key step of the proof. Inequality (7) exhibits a property, sometimes referred to as the Bernstein property, of the functional class  $\{d(\cdot, b)^2 - d(\cdot, \bar{b})^2 : b \in B\}$  indexing our empirical process, allowing to control its second moment (on the left-hand side) by its first moment (the right-hand side). The importance of such a condition for obtaining fast convergence rates for empirical risk minimizers has been emphasized in one way or another by many authors including [Mammen and Tsybakov \(1999\)](#); [Massart \(2000\)](#); [Mendelson \(2002\)](#); [Blanchard et al. \(2003\)](#); [Bartlett et al. \(2005, 2006\)](#); [Koltchinskii \(2006\)](#) and [Bartlett and Mendelson \(2006\)](#) among others. Bernstein's condition is known to be sometimes connected to the convexity properties of the loss function in statistical learning theory. It is of interest to point out that, in the context of our study, this condition also has a very clear geometrical interpretation.

It follows from the definition of  $\sigma(\delta)$  and equation (7) that

$$\sigma(\delta) \leq 2R\sqrt{K\delta^\beta}. \quad (8)$$

Next, we provide an upper bound for  $\mathbf{E}\varphi_n(\delta)$ . Let  $\sigma_1, \dots, \sigma_n$  be a sequence of i.i.d. random signs, *i.e.* such that  $\mathbf{P}(\sigma_i = -1) = \mathbf{P}(\sigma_i = 1) = 1/2$ , independent from the  $X_i$ 's. A classical symmetrization argument (see, e.g., Lemma 7.4 in [van Handel, 2016](#)) indicates that

$$\begin{aligned} \mathbf{E}\varphi_n(\delta) &\leq 2\mathbf{E} \sup \left\{ \frac{1}{n} \sum_{i=1}^n \sigma_i (d(X_i, b)^2 - d(X_i, \bar{b})^2) : b \in B(\delta) \right\} \\ &= 2\mathbf{E} \sup \left\{ \frac{1}{n} \sum_{i=1}^n \sigma_i d(X_i, b)^2 : b \in B(\delta) \right\}. \end{aligned}$$

By the contraction principle (see Theorem 4.12 in [Ledoux and Talagrand, 1991](#)), the fact that  $d(X_i, b) \leq R$  almost surely and the fact that  $x \mapsto x^2$  is  $2R$ -Lipschitz on  $[0, R]$ , we further deduce that

$$\mathbf{E}\varphi_n(\delta) \leq 4R \mathbf{E} \sup \left\{ \frac{1}{n} \sum_{i=1}^n \sigma_i d(X_i, b) : b \in B(\delta) \right\}.$$

Now observe that, conditionally on the  $X_i$ 's, the process

$$\xi_b := \frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_i d(X_i, b), \quad b \in B$$

satisfies the separability condition of Lemma 17 due to the separability of  $S$ , is subgaussian with

$$\forall b, b' \in B, \forall \theta \in \mathbf{R}, \quad \log \mathbf{E}[\exp(\theta\{\xi_b - \xi_{b'}\}) | X_1, \dots, X_n] \leq \frac{\theta^2 d(b, b')^2}{2},$$

and satisfies almost surely the Lipschitz condition  $|\xi_b - \xi_{b'}| \leq \sqrt{n}d(b, b')$ . Hence, applying Lemma 17 we obtain

$$\mathbf{E}\varphi_n(\delta) \leq 4R \inf_{\varepsilon \geq 0} \left\{ 2\varepsilon + \frac{12}{\sqrt{n}} \int_{\varepsilon}^{+\infty} \sqrt{\log N(B(\delta), d, u)} du \right\}.$$

Using once again Assumption **(A1)** yields  $B(\delta) \subset B \cap B(\bar{b}, \sqrt{K\delta^\beta})$  so that

$$\begin{aligned} \mathbf{E}\varphi_n(\delta) &\leq 4R \inf_{\varepsilon \geq 0} \left\{ 2\varepsilon + \frac{12}{\sqrt{n}} \int_\varepsilon^{+\infty} \sqrt{\log N(B \cap B(\bar{b}, \sqrt{K\delta^\beta}), d, u)} du \right\} \\ &= 4R \inf_{\varepsilon \geq 0} \left\{ 2\varepsilon + \frac{12}{\sqrt{n}} \int_\varepsilon^{\sqrt{K\delta^\beta}} \sqrt{\log N(B \cap B(\bar{b}, \sqrt{K\delta^\beta}), d, u)} du \right\}. \end{aligned} \quad (9)$$

Now consider the case  $D \in (0, 2)$  where parameter  $D$  appears in Assumption **(A2)**. In this case, the integrand on the right-hand side of (9) is integrable at 0. Hence, taking  $\varepsilon = 0$  in (9) and using Assumption **(A2)**, it follows that

$$\begin{aligned} \mathbf{E}\varphi_n(\delta) &\leq \frac{48R}{\sqrt{n}} \int_0^{\sqrt{K\delta^\beta}} \left( \frac{C\sqrt{K\delta^\beta}}{u} \right)^{\frac{D}{2}} du \\ &= \frac{96RC^{\frac{D}{2}} K^{\frac{1}{2}}}{2-D} \sqrt{\frac{\delta^\beta}{n}}. \end{aligned} \quad (10)$$

Combining (6),(8) and (10) implies therefore that inequality

$$\varphi_n(\delta) \leq \beta_n(\delta, t) := c_1 \sqrt{\frac{\delta^\beta}{n}} + c_2 \sqrt{\frac{t\delta^\beta}{n}} + \frac{c_3 t}{n},$$

holds with probability at least  $1 - e^{-t}$  with  $c_1 := 96RC^{\frac{D}{2}} K^{\frac{1}{2}}/(2-D)$ ,  $c_2 = 2R\sqrt{2K}$  and  $c_3 = 16R\text{diam}(B)/3$ . Using the fact that  $\delta_n \leq \varphi_n(\delta_n)$  and Lemma 15, it follows that

$$\delta_n \leq \beta_n(t) := \inf \left\{ \tau > 0 : \sup_{\delta \geq \tau} \delta^{-1} \beta_n \left( \delta, \frac{t\delta}{\tau} \right) \leq 1 \right\},$$

with probability larger than  $1 - 2e^{-t}$ . It therefore remains to provide an upper bound for  $\beta_n(t)$ . Straightforward computations using the fact that  $\beta \in (0, 1]$  indicate that, for any  $\tau > 0$ ,

$$\sup_{\delta \geq \tau} \delta^{-1} \beta_n \left( \delta, \frac{t\delta}{\tau} \right) = \frac{c_1}{\sqrt{n}} \tau^{\frac{\beta}{2}-1} + c_2 \sqrt{\frac{t}{n}} \tau^{\frac{\beta}{2}-1} + \frac{c_3 t}{\tau n}.$$

Observing finally that for any nonincreasing functions  $h_j : [0, +\infty) \rightarrow [0, +\infty)$ , we have

$$\inf\{\tau > 0 : h_1(\tau) + \dots + h_m(\tau) \leq 1\} \leq \max_{1 \leq j \leq m} \inf\{\tau > 0 : h_j(\tau) \leq 1/m\},$$

it follows that

$$\beta_n(t) \leq \max \left\{ A_1 \left( \frac{1}{n} \right)^{\frac{1}{2-\beta}}, A_2 \left( \frac{t}{n} \right)^{\frac{1}{2-\beta}}, A_3 \left( \frac{t}{n} \right) \right\},$$

with

$$A_1 = (3c_1)^{\frac{2}{2-\beta}}, \quad A_2 = (3c_2)^{\frac{2}{2-\beta}} \quad \text{and} \quad A_3 = 3c_3.$$

**Statement 19** (Case  $D \in (0, 2)$ ). *To sum up, we have proven so far that under the assumptions of the theorem, and if  $D \in (0, 2)$ , then for all  $n \geq 1$  and all  $t > 0$ ,*

$$\int_S (d(x, b_n)^2 - d(x, \bar{b})^2) dP(x) \leq \max \left\{ A_1 \left( \frac{1}{n} \right)^{\frac{1}{2-\beta}}, A_2 \left( \frac{t}{n} \right)^{\frac{1}{2-\beta}}, A_3 \left( \frac{t}{n} \right) \right\},$$

with probability at least  $1 - 2e^{-t}$ , with

$$A_1 := \left( \frac{288RC^{\frac{D}{2}} K^{\frac{1}{2}}}{2-D} \right)^{\frac{2}{2-\beta}}, \quad A_2 := (6R\sqrt{2K})^{\frac{2}{2-\beta}}, \quad A_3 := 16R\text{diam}(B),$$

and  $R := \max\{d(x, y) : x \in B, y \in \text{supp}(P)\}$ .

The rest of the proof is devoted to the case  $D \geq 2$ . Suppose  $D = 2$ . Combining Assumption **(A2)** with inequality (9) yields in this case

$$\mathbf{E}\varphi_n(\delta) \leq 8R\varepsilon + 48RC\sqrt{\frac{K\delta^\beta}{n}} \log\left(\frac{\sqrt{K\delta^\beta}}{\varepsilon}\right),$$

for all  $\varepsilon > 0$ . Taking  $\varepsilon = \sqrt{K\delta^\beta/n}$  provides the upper bound

$$\mathbf{E}\varphi_n(\delta) \leq 8R\sqrt{\frac{K\delta^\beta}{n}} + 96RC(\log n)\sqrt{\frac{K\delta^\beta}{n}} = 8R\sqrt{K}(1 + 12C \log n)\sqrt{\frac{\delta^\beta}{n}}.$$

Then, the proof follows by applying the exact same steps as in the case  $D \in (0, 2)$  with  $c'_1 = 8R\sqrt{K}(1 + 12C \log n)$  instead of  $c_1$ .

**Statement 20** (Case  $D = 2$ ). *We have proven that under the assumptions of the theorem, and if  $D = 2$ , then for all  $n \geq e$  and all  $t > 0$ ,*

$$\int_S (d(x, b_n)^2 - d(x, \bar{b})^2) dP(x) \leq \max\left\{A'_1 \left(\frac{\log^2 n}{n}\right)^{\frac{1}{2-\beta}}, A_2 \left(\frac{t}{n}\right)^{\frac{1}{2-\beta}}, A_3 \left(\frac{t}{n}\right)\right\},$$

with probability at least  $1 - 2e^{-t}$ , with

$$A'_1 := (48R\sqrt{K} \max\{1, 12C\})^{\frac{2}{2-\beta}}, \quad A_2 := (6R\sqrt{2K})^{\frac{2}{2-\beta}}, \quad A_3 := 16R\text{diam}(B),$$

and  $R := \max\{d(x, y) : x \in B, y \in \text{supp}(P)\}$ .

Suppose finally  $D > 2$ . Combining Assumption **(A2)** with inequality (9) yields in this case

$$\begin{aligned} \mathbf{E}\varphi_n(\delta) &\leq 8R\varepsilon + \frac{96RC^{\frac{D}{2}} K^{\frac{D}{4}} \delta^{\frac{\beta D}{4}}}{D-2} \frac{1}{\sqrt{n}} \left[ \frac{1}{\varepsilon^{\frac{D-2}{2}}} - \frac{1}{(K\delta^\beta)^{\frac{D-2}{4}}} \right] \\ &\leq 8R\varepsilon + \frac{96RC^{\frac{D}{2}} K^{\frac{D}{4}} \delta^{\frac{\beta D}{4}}}{D-2} \frac{1}{\sqrt{n} \varepsilon^{\frac{D-2}{2}}}, \end{aligned}$$

for all  $\varepsilon > 0$ . Taking  $\varepsilon$  so that the two terms on the right-hand side coincide, *i.e.*

$$\varepsilon = C\sqrt{K} \left( \frac{12}{D-2} \right)^{\frac{2}{D}} \delta^{\frac{\beta}{2}} n^{-\frac{1}{D}},$$

we obtain

$$\mathbf{E}\varphi_n(\delta) \leq c''_1 \delta^{\frac{\beta}{2}} n^{-\frac{1}{D}} \quad \text{where} \quad c''_1 := 16RC\sqrt{K} \left( \frac{12}{D-2} \right)^{\frac{2}{D}},$$

and therefore

$$\varphi_n(\delta) \leq \beta_n(\delta, t) := c''_1 \delta^{\frac{\beta}{2}} n^{-\frac{1}{D}} + c_2 \sqrt{\frac{t\delta^\beta}{n}} + \frac{c_3 t}{n},$$

with probability at least  $1 - e^{-t}$  where  $c_2$  and  $c_3$  are as in the previous cases. The proof now ends using the same strategy as above and we obtain the following.

**Statement 21** (Case  $D > 2$ ). *Under the assumptions of the theorem and assuming  $D > 2$ , we obtain for all  $n \geq 1$  and all  $t > 0$ ,*

$$\int_S (d(x, b_n)^2 - d(x, \bar{b})^2) dP(x) \leq \max \left\{ A_1'' n^{-\frac{2}{D(2-\beta)}}, A_2 \left( \frac{t}{n} \right)^{\frac{1}{2-\beta}}, A_3 \left( \frac{t}{n} \right) \right\},$$

with probability at least  $1 - 2e^{-t}$ , with

$$A_1'' := (48CR\sqrt{K})^{\frac{2}{2-\beta}} \left( \frac{12}{D-2} \right)^{\frac{4}{D(2-\beta)}}, \quad A_2 := (6R\sqrt{2K})^{\frac{2}{2-\beta}}, \quad A_3 := 16R\text{diam}(B),$$

and  $R := \max\{d(x, y) : x \in B, y \in \text{supp}(P)\}$ .

The proof is complete. □

## 4.2 Proof of Theorem 6

By  $(k, \alpha)$ -convexity of  $B$ , we deduce that for all  $x, b \in B$ , any constant speed shortest path  $\gamma = \gamma_{\bar{b}b} : [0, 1] \rightarrow S$  joining  $\bar{b}$  to  $b$  and any  $t \in [0, 1]$ ,

$$d(x, \gamma(t))^2 \leq (1-t)d(x, \bar{b})^2 + td(x, b)^2 - \frac{k}{2}t(1-t)d(b, \bar{b})^\alpha.$$

After rearranging the terms, we obtain, for all  $t \in (0, 1)$ ,

$$\frac{k}{2}d(b, \bar{b})^\alpha \leq \frac{d(x, b)^2 - d(x, \bar{b})^2}{1-t} + \frac{d(x, \bar{b})^2 - d(x, \gamma(t))^2}{t(1-t)}.$$

Integrating this inequality with respect to  $P$ , using the definition of  $\bar{b}$  and the fact that according to the  $(k, \alpha)$ -convexity of  $B$  we have  $\gamma(t) \in B$  for all  $t$ , it follows that

$$\frac{k}{2}d(b, \bar{b})^\alpha \leq \frac{P(d(\cdot, b)^2 - d(\cdot, \bar{b})^2)}{1-t},$$

for all  $t \in (0, 1)$ . The result then follows by letting  $t$  tend to 0.

## 4.3 Proof of Theorem 8

Let  $p, x, y \in B$  be arbitrary and let  $\bar{p}, \bar{x}, \bar{y} \in M_1^2$  be a comparison triangle in  $M_1^2$ , the Euclidean sphere of radius 1 in  $\mathbf{R}^3$ . Since  $B$  has curvature upper bounded by 1 in the large, it follows by definition of curvature and completeness of  $B$  that this property is equivalent to the fact that for any constant speed shortest path  $\gamma$  between  $x$  and  $y$  in  $B$  and any constant speed shortest path  $\bar{\gamma}$  between  $\bar{x}$  and  $\bar{y}$  in  $M_1^2$ , we have

$$d(p, \gamma(1/2)) \leq d_1(\bar{p}, \bar{\gamma}(1/2)),$$

where  $d_1$  denotes the angular metric on  $M_1^2$ . Hence, to show that  $B$  is  $(k, \alpha)$ -convex, it is enough to show that for any pairwise distinct  $\bar{p}, \bar{x}, \bar{y} \in M_1^2$  we have

$$kd_1(\bar{x}, \bar{y})^{\alpha-2} \leq \frac{8}{d_1(\bar{x}, \bar{y})^2} \left( \frac{1}{2}d_1(\bar{p}, \bar{x})^2 + \frac{1}{2}d_1(\bar{p}, \bar{y})^2 - d_1(\bar{p}, \bar{\gamma}(t))^2 \right).$$

Since  $\alpha \geq 2$ , it is enough for this property to hold that, for all pairwise distinct  $\bar{p}, \bar{x}, \bar{y} \in M_1^2$ , we have

$$kd^{\alpha-2} \leq \frac{8}{d_1(\bar{x}, \bar{y})^2} \left( \frac{1}{2}d_1(\bar{p}, \bar{x})^2 + \frac{1}{2}d_1(\bar{p}, \bar{y})^2 - d_1(\bar{p}, \bar{\gamma}(t))^2 \right).$$

As proven by [Ohta \(2007\)](#), the right-hand side is upper bounded by

$$8d \tan\left(\frac{\pi}{2} - d\right),$$

which concludes the proof.

#### 4.4 Proof of Theorem 9

(a)  $\Rightarrow$  (b): For any  $x, y \in S$  let  $\gamma_{xy} : [0, 1] \rightarrow S$  be a constant speed shortest path joining  $x$  to  $y$ . For any  $t \in [0, 1]$ , applying property (a) with  $P = (1 - t)\delta_x + t\delta_y$ , and noticing that  $b^* = \gamma_{xy}(t)$  is a barycentre of  $P$ , yields property (b).

(b)  $\Rightarrow$  (c): This implication follows from the proof of Theorem 6 with  $B = S$  and  $\alpha = 2$ .

(c)  $\Rightarrow$  (d): As for the first implication, the result follows from applying property (c) to the measure  $P = (1 - t)\delta_x + t\delta_y$  for all  $x, y \in S$  and all  $t \in [0, 1]$ .

The last statement follows from the fact that  $u + av \leq au + v$  whenever  $0 \leq a \leq 1$  and  $u \leq v$ . Indeed, applying this inequality with  $a = k/2$ ,  $u = d^2(b, b^*)$  and  $v = Pd^2(\cdot, b^*)$  yields the desired result.

#### 4.5 Proof of Theorem 12

In the proof, we implicitly identify a point  $x \in S$  and its image  $\log_{b^*}(x)$  in  $T_{b^*}S$  for readability. Since  $T_{b^*}S$  is a global NNC space by Proposition 31, it follows from Proposition 1.7 in [Sturm \(1999\)](#) that

$$\int_S \int_S \langle x, y \rangle_{b^*} dP(x)dP(y) \geq 0.$$

Moreover, Proposition 31 also ensures that  $d(x, y) \leq \|x - y\|_{b^*}$  for all  $x, y \in S$  with equality if  $x = b^*$  or  $y = b^*$ . In particular,  $b^*$  minimizes also

$$y \mapsto \int_S \|x - y\|_{b^*}^2 dP(x).$$

Thus, using the properties of  $\|\cdot\|_{b^*}$ , discussed in Appendix B, we get for all  $y \in S$  and  $t > 0$ ,

$$\begin{aligned} \int_S \|x\|_{b^*}^2 dP(x) &= \int_S \|x - b^*\|_{b^*}^2 dP(x) \\ &\leq \int_S \|x - ty\|_{b^*}^2 dP(x) \\ &= \int_S (\|x\|_{b^*}^2 + \|ty\|_{b^*}^2 - 2\langle x, ty \rangle_{b^*}) dP(x), \end{aligned}$$

so that

$$2 \int_S \langle x, y \rangle_{b^*} dP(x) \leq t\|y\|_{b^*}^2.$$

Letting  $t \rightarrow 0^+$ , we get

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

Then, integrating with respect to  $y$ , we obtain

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

Combining this observation with the first inequality of the proof yields the desired result.



#### 4.6 Proof of Theorem 13

In the proof, we implicitly identify a point  $x \in S$  and its image  $\log_{b^*}(x)$  in  $T_{b^*}S$  for readability. In particular, we also identify  $\text{supp}(P) \subset S$  with  $\text{supp}((\log_{b^*})_{\#}P) \subset T_{b^*}S$ . According to Theorem 12, any barycentre  $b^*$  of  $P$  is also an exponential barycentre of  $P$ . For  $b \in \text{span}(\text{supp}(P))$ , let  $b_t$  be a constant speed shortest path connecting  $b^*$  to  $b$ . Then,  $\text{span}(\text{supp}(P))$  being a flat space according to Theorem A in Yokota (2012), it follows that for all  $x \in S$ ,

$$\begin{aligned} & t(1-t)d^2(b, b^*) \\ &= (1-t)d^2(x, b^*) + t\|x - b\|_{b^*}^2 - \|x - b_t\|_{b^*}^2 \\ &= t(\|x - b\|_{b^*}^2 - d^2(x, b^*)) + (d^2(x, b^*) - \|x - b_t\|_{b^*}^2) \\ &= t(d^2(x, b) - d^2(x, b^*)) + (d^2(x, b^*) - \|x - b_t\|_{b^*}^2) + t(\|x - b\|_{b^*}^2 - d^2(x, b)) \\ &= t(d^2(x, b) - d^2(x, b^*)) + (d^2(x, b^*) - \|x - b_t\|_{b^*}^2) + t(1 - k_0(x, b, b^*))d^2(b, b^*). \end{aligned}$$

Then by dividing by  $t > 0$ , we have

$$(k_0(x, b, b^*) - t)d^2(b, b^*) = (d^2(x, b) - d^2(x, b^*)) + \frac{1}{t}(d^2(x, b^*) - \|x - b_t\|_{b^*}^2).$$

Using that  $S$  is a global NNC space, we deduce that

$$(k_0(x, b, b^*) - t)d^2(b, b^*) \leq (d^2(x, b) - d^2(x, b^*)) + \frac{1}{t}(d^2(x, b^*) - d^2(x, b_t)),$$

and integrating with respect to  $P$ , we obtain

$$(Pk_0(\cdot, b, b^*) - t)d^2(b, b^*) \leq P(d(\cdot, b)^2 - d(\cdot, b^*)^2) + \frac{P(d(\cdot, b^*)^2 - d(\cdot, b_t)^2)}{t}.$$

Since  $b^*$  is a barycentre of  $P$ , we get  $P(d(\cdot, b^*)^2 - d(\cdot, b_t)^2) \leq 0$  and thus the result follows by letting  $t \rightarrow 0$ .

#### 4.7 Proof of Theorem 14

We proceed to the same identifications as in the proof of the previous Theorem. First remark that since  $b^*$  is barycentre of  $P$ , it is also an exponential barycentre of  $P$  by Theorem 12. Since  $\text{supp}(P)$  is flat on  $T_{b^*}S$ ,  $b^*$  is also an exponential barycentre of  $P_\lambda$  and, since  $P_\lambda$  has a unique barycentre by assumption, it is also the barycentre of  $P_\lambda$ .

Now let  $x, b \in \text{supp}(P)$  and denote  $x^\lambda := \gamma_x^+(1 + \lambda)$ . Remark that on  $T_xS$  (through the identification by  $\log_x$ ) we have  $x^\lambda = x + \lambda(x - b^*)$ . Now take  $t = 1/(1 + \lambda)$ . Then, on  $T_xS$ ,

$$x = tx^\lambda + (1-t)b^*. \quad (11)$$

As  $S$  is a global NNC space,  $T_xS$  itself is a global NNC space by Proposition 31. Hence, we deduce that

$$\begin{aligned} \|x - b\|_x^2 &\geq (1-t)\|b - b^*\|_x^2 + t\|x^\lambda - b\|_x^2 - t(1-t)\|x^\lambda - b^*\|_x^2 \\ &= \frac{\lambda}{1+\lambda}\|b - b^*\|_x^2 + \frac{1}{1+\lambda}\|x^\lambda - b\|_x^2 - \frac{\lambda}{(1+\lambda)^2}\|x^\lambda - b^*\|_x^2, \end{aligned}$$

which, using the fact  $\|x^\lambda - b^\star\|_x^2 = (1 + \lambda)^2 \|x - b^\star\|_x^2$ , we can rewrite as

$$\begin{aligned} \frac{\lambda}{1 + \lambda} \|b - b^\star\|_x^2 &\leq d^2(x, b) - \frac{1}{1 + \lambda} \|x^\lambda - b\|_x^2 + \lambda d^2(x, b^\star) \\ &= d^2(x, b) - d^2(x, b^\star) - \frac{1}{1 + \lambda} \|x^\lambda - b\|_x^2 + (1 + \lambda) d^2(x, b^\star). \end{aligned}$$

Now, letting

$$c(\lambda) = 1 + \lambda - \frac{(\int \|x^\lambda - b\|_x^2 dP(x))^{1/2}}{(\int d(x, b^\star)^2 dP(x))^{1/2}},$$

it follows that

$$\int_S \left( -\frac{1}{1 + \lambda} \|x^\lambda - b\|_x^2 + (1 + \lambda) d^2(x, b^\star) \right) dP(x) = \left( 2c(\lambda) - \frac{c(\lambda)^2}{1 + \lambda} \right) \int_S d^2(x, b^\star) dP(x).$$

Hence, since  $\|b - b^\star\|_x \geq d(b, b^\star)$  and  $\|b - x\|_x = d(b, x)$ , we deduce from the previous computations that

$$\frac{\lambda}{1 + \lambda} d(b, b^\star)^2 \leq \int_S (d(x, b)^2 - d(x, b^\star)^2) dP(x) + 2c(\lambda) \int_S d(x, b^\star)^2 dP(x).$$

Finally, we observe that since  $b^\star$  is the unique barycentre of  $P_\lambda$ , we have

$$c(\lambda) \leq 1 + \lambda - \frac{(\int d^2(x^\lambda, b^\star) dP(x))^{1/2}}{(\int d(x, b^\star)^2 dP(x))^{1/2}} \leq 0,$$

which concludes the proof.

## A Bound in expectation

This section provides, for completeness, a bound in expectation derived from Theorem 1.

**Theorem 22.** *Suppose that the support of  $P$  and the set  $B$  have finite diameter. Then, under Assumptions (A1) and (A2), and for all  $n \geq 1$ ,*

$$\mathbf{E} \int_S (d(b_n, x)^2 - d(\bar{b}, x)^2) dP(x) \leq \begin{cases} (A_1 + \alpha A_2) n^{-\frac{1}{2-\beta}} + \frac{\alpha A_3}{n} & \text{if } D < 2, \\ A'_1 (\log n)^2 n^{-\frac{1}{2-\beta}} + \alpha A_2 n^{-\frac{1}{2-\beta}} + \frac{\alpha A_3}{n} & \text{if } D = 2, \\ A''_1 n^{-\frac{2}{D(2-\beta)}} + \alpha A_2 n^{-\frac{1}{2-\beta}} + \frac{\alpha A_3}{n} & \text{if } D > 2, \end{cases}$$

where  $A_1, A'_1, A''_1, A_2, A_3$  are as in Statements 19, 20 and 21 and where finally  $\alpha = 2$  if  $\beta = 1$  and  $\alpha = 1 + 4/e(2 - \beta)$  if  $\beta \in (0, 1)$ .

*Proof.* We only outline the proof for  $D < 1$  as the other cases are exactly the same. As in the proof of Theorem 1, let

$$\delta_n := \int_S (d(b_n, x)^2 - d(\bar{b}, x)^2) dP(x).$$

Abbreviate

$$a_1 := A_1 \left( \frac{1}{n} \right)^{\frac{1}{2-\beta}}; a_2 := A_2 \left( \frac{1}{n} \right)^{\frac{1}{2-\beta}} \quad \text{and} \quad a_3 := \frac{A_3}{n},$$

for simplicity, where  $A_1, A_2$  and  $A_3$  are the constants involved in Statement 19. Then, it follows that

$$\begin{aligned}\mathbf{E}\delta_n &= \int_0^{+\infty} \mathbf{P}(\delta_n > t) dt \\ &\leq a_1 + \int_0^{+\infty} \mathbf{P}(\delta_n > a_1 + t) dt \\ &= a_1 + (a_2 + a_3) \int_0^{+\infty} \mathbf{P}(\delta_n > a_1 + (a_2 + a_3)t) dt.\end{aligned}$$

If  $\beta = 1$ , we deduce from Statement 19 that

$$\mathbf{E}\delta_n \leq a_1 + (a_2 + a_3) \int_0^{+\infty} 2e^{-t} dt = a_1 + 2(a_2 + a_3),$$

which proves the statement. Now if  $\beta \in (0, 1)$ , and using the lines above, we write

$$\begin{aligned}\mathbf{E}\delta_n &\leq a_1 + (a_2 + a_3) \int_0^{+\infty} \mathbf{P}(\delta_n > a_1 + (a_2 + a_3)t) dt \\ &\leq a_1 + (a_2 + a_3) + (a_2 + a_3) \int_1^{+\infty} \mathbf{P}(\delta_n > a_1 + (a_2 + a_3)t) dt \\ &= a_1 + (a_2 + a_3) + \frac{2(a_2 + a_3)}{2 - \beta} \int_1^{+\infty} t^{\frac{2}{2-\beta}-1} \mathbf{P}(\delta_n > a_1 + (a_2 + a_3)t^{\frac{2}{2-\beta}}) dt \\ &\leq a_1 + (a_2 + a_3) + \frac{2(a_2 + a_3)}{2 - \beta} \int_1^{+\infty} \mathbf{P}(\delta_n > a_1 + a_2 t^{\frac{2}{2-\beta}} + a_3 t) dt \\ &\leq a_1 + (a_2 + a_3) + \frac{2(a_2 + a_3)}{2 - \beta} \int_1^{+\infty} 2e^{-t} dt \\ &\leq a_1 + (a_2 + a_3) + \frac{4(a_2 + a_3)}{e(2 - \beta)},\end{aligned}$$

where the last inequality follows again from Statement 19. This completes the proof.  $\square$

## B Background in metric geometry

This appendix gathers material used throughout the paper. General references, for topics discussed below can be found for instance in [Burago et al. \(2001\)](#) or [Alexander et al. \(2017\)](#).

### B.1 Geodesic spaces

Let  $(S, d)$  be a metric space. We call path in  $S$  a continuous map  $\gamma : I \rightarrow S$  defined on a nonempty interval  $I \subset \mathbf{R}$ . Define the length  $L(\gamma) \in [0, +\infty]$  of a path  $\gamma : I \rightarrow S$  by

$$L(\gamma) := \sup_{n \geq 1} \sup_{t_1 < \dots < t_n \in I} \sum_{i=1}^n d(\gamma(t_i), \gamma(t_{i+1})).$$

A path  $\gamma : [a, b] \rightarrow S$  is said to connect  $x$  to  $y$  if  $\gamma(a) = x$  and  $\gamma(b) = y$ . The space  $S$  is called a length space if, for all  $x, y \in S$ ,

$$d(x, y) = \inf_{\gamma} L(\gamma), \tag{12}$$

where the infimum is taken over all paths  $\gamma$  that connect  $x$  to  $y$ . A length space is said to be geodesic if, for all  $x, y \in S$ , the infimum on the right hand side of (12) is attained. A path, realizing this infimum is called a shortest path (or geodesic) between  $x$  and  $y$  and needs not be unique. In a geodesic space, a constant speed shortest path between  $x$  and  $y$  is a path  $\gamma : [0, 1] \rightarrow S$  such that  $\gamma(0) = x$ ,  $\gamma(1) = y$  and such that, for all  $0 \leq s \leq t \leq 1$ ,

$$L(\gamma_{[s,t]}) = (t - s)d(x, y),$$

where  $\gamma_{[s,t]}$  denotes the restriction of  $\gamma$  to  $[s, t]$ . A unit speed shortest path between  $x$  and  $y$  is a path  $\gamma : [0, T] \rightarrow S$  such that  $\gamma(0) = x$ ,  $\gamma(T) = y$  and such that, for all  $0 \leq s \leq t \leq 1$ ,

$$L(\gamma_{[s,t]}) = (t - s).$$

In this situation, note in particular that  $T = d(x, y)$ . The following characterization of geodesic spaces is helpful.

**Proposition 23.** *Let  $(S, d)$  be a metric space.*

- (1) *If  $S$  is a geodesic space, then any two points  $x, y \in S$  admit a midpoint, i.e. a point  $z \in S$  such that*

$$d(x, z) = d(y, z) = \frac{1}{2}d(x, y).$$

- (2) *Conversely, if  $S$  is complete and if any two points in  $S$  admit a midpoint, then  $S$  is a geodesic space.*

## B.2 Curvature

Geodesic spaces of interest are those with bounded curvature in the sense of Alexandrov. In this setting, curvature bounds are defined by comparison arguments involving model surfaces defined as follows. For  $\kappa \in \mathbf{R}$ , we'll denote  $(M_\kappa^2, d_\kappa)$  the 2-dimensional model surface (also referred to as the  $\kappa$ -plane) defined by

- The Hyperbolic plane (modeled as the Poincaré disc for instance), with metric multiplied by  $1/\sqrt{-\kappa}$ , if  $\kappa < 0$ ,
- The Euclidean plane equipped with its Euclidean metric, if  $\kappa = 0$ ,
- The Euclidean sphere in  $\mathbf{R}^3$  of radius  $1/\sqrt{\kappa}$  equipped with the angular metric, if  $\kappa > 0$ .

In addition to these model surfaces, the definition of curvature bounds given below requires to define the notion of comparison triangles. For any triangle in  $S$ , i.e. any three points set  $\{p, x, y\}$  in  $S$ , a comparison triangle in  $M_\kappa^2$  is defined as a set  $\{\bar{p}, \bar{x}, \bar{y}\}$  in  $M_\kappa^2$  with the property that  $d_\kappa(\bar{p}, \bar{x}) = d(p, x)$ ,  $d_\kappa(\bar{p}, \bar{y}) = d(p, y)$  and  $d_\kappa(\bar{x}, \bar{y}) = d(x, y)$ .

**Definition 24.** *Let  $\kappa \in \mathbf{R}$  and  $(S, d)$  be a geodesic space.*

- *The space  $(S, d)$  is said to have curvature bounded below by  $\kappa$  if any point  $z \in S$  has a neighborhood  $U_z$  for which the following holds: for any points  $p, x, y \in U_z$ , any comparison triangle  $\bar{p}, \bar{x}, \bar{y}$  in  $M_\kappa^2$ , any constant speed shortest path  $\gamma$  joining  $x$  to  $y$  in  $S$  (whose image is contained in  $U_z$ ) and any constant speed shortest path  $\bar{\gamma}$  joining  $\bar{x}$  to  $\bar{y}$  in  $M_\kappa^2$ , we have*

$$\forall t \in [0, 1], \quad d(p, \gamma(t)) \geq d_\kappa(\bar{p}, \bar{\gamma}(t)). \quad (13)$$

- The space  $(S, d)$  is said to have curvature bounded above by  $\kappa$  if the above definition holds with opposite inequality in (13).
- The space  $(S, d)$  is said to be nonnegatively curved (NNC) if it has curvature bounded below by 0.
- The space  $(S, d)$  is said to be nonpositively curved (NPC) if it has curvature bounded above by 0.

The previous definition has a natural geometric interpretation: in a space  $S$  with curvature lower (resp. upper) bounded by  $\kappa$ , a (small enough) triangle (meaning here three points joined by shortest paths) looks fatter (resp. thinner) than a corresponding comparison triangle in  $M_\kappa^2$ . In the context of NNC and NPC spaces, the above definition may be given an alternative form of practical interest.

**Proposition 25.** *A geodesic space  $(S, d)$  is NNC if, and only if, any point  $z \in S$  has a neighbourhood  $U_z$  for which the following holds: for any points  $p, x, y \in U_z$  and any constant speed shortest path  $\gamma$  joining  $x$  to  $y$ , whose image is contained in  $U_z$ , we have*

$$\forall t \in [0, 1], \quad d(p, \gamma(t))^2 \geq (1-t)d(p, x)^2 + td(p, y)^2 - t(1-t)d(x, y)^2.$$

*A space  $(S, d)$  is NPC if, and only if, the same statement holds with opposite inequality.*

The proof follows immediately from Definition 24 by exploiting the geometry of the flat Euclidean plane. For  $\kappa \neq 0$ , an equivalent formulation of Definition 24, given only in terms of the ambient metric  $d$ , is given in the next subsection using the notion of angle. It is worth noting that the previous definitions are of local nature as they require inequalities to be valid in the neighborhood of a point. Some spaces have the additional property that the comparison inequalities of Definition 24 are valid without restriction on the points  $p, x$  and  $y$ . Such spaces are said to have curvature bounded from below (or above) by  $\kappa$  in the large (or globally). Note that  $(M_\kappa^2, d_\kappa)$  has diameter  $\pi/\sqrt{\kappa}$  when  $\kappa > 0$ . Hence, for  $\kappa > 0$ , one can find a comparison triangle for  $\{p, x, y\}$  provided the perimeter  $\text{peri}\{p, x, y\} := d(p, x) + d(p, y) + d(x, y) \leq 2\pi/\sqrt{\kappa}$ . The next definition accounts for this observation.

**Definition 26.** *Let  $(S, d)$  be a geodesic space.*

- *If  $\kappa > 0$ , then  $(S, d)$  is said to have curvature bounded from above (or below) by  $\kappa$  in the large, if the comparison inequality of Definition 24 (from above or below) holds for all points  $p, x, y \in S$  such that  $\text{peri}\{p, x, y\} \leq 2\pi/\sqrt{\kappa}$ .*
- *If  $\kappa \leq 0$ , then  $(S, d)$  is said to have curvature bounded from above (or below) by  $\kappa$  in the large, if the comparison inequality of Definition 24 (from above or below) holds for all points  $p, x, y \in S$ .*

A geodesic space with curvature upper bounded by some  $\kappa \geq 0$ , in the large, is sometimes referred to as a  $\text{CAT}(\kappa)$  space in the literature in reference to the contributions of E. Cartan, A.D. Alexandrov and V.A. Topogonov. A  $\text{CAT}(0)$  space is also referred to as a global NPC space. Similarly, a space with curvature lower bounded by 0 in the large will be called a global NNC space. Note that a  $\text{CAT}(\kappa)$  space is also a  $\text{CAT}(\kappa')$  space for all  $\kappa' > \kappa$ . Riemannian manifolds with sectional curvature at most  $\kappa$  and injectivity radius not less than  $\pi/\sqrt{\kappa}$  are typical examples of  $\text{CAT}(\kappa)$  spaces. An important fact is that in a  $\text{CAT}(\kappa)$  space, two points  $x, y$  with  $d(x, y) < \pi/\sqrt{\kappa}$  have a unique shortest path connecting them (up to reparametrization). The following results are known as globalization theorems.

**Theorem 27** (Globalization). *Let  $(S, d)$  be a geodesic space.*

- *Suppose  $(S, d)$  is complete, simply connected and has curvature bounded from above by some  $\kappa \leq 0$ . Then  $(S, d)$  has curvature bounded from above by  $\kappa$  in the large (i.e. is a  $\text{CAT}(\kappa)$  space).*
- *Suppose  $(S, d)$  is complete and has curvature bounded from below by some  $\kappa \in \mathbf{R}$ . Then  $(S, d)$  has curvature bounded from below by  $\kappa$  in the large.*

### B.3 Angles, spaces of directions and tangent cones

Metric spaces considered so far have a priori no differentiable structure. In this context, an analog of a tangent space is provided by the notion of a tangent cone. This section shortly reviews this notion. We first need the notion of angles and spaces of directions.

For  $\kappa \in \mathbf{R}$  and a triangle  $\{p, x, y\}$  in  $S$ , such that  $\text{peri}\{p, x, y\} < 2\pi/\sqrt{\kappa}$  if  $\kappa > 0$ , we define the comparison angle  $\angle_\kappa(x, p, y) \in [0, \pi]$  at  $p$  by

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

where  $C_\kappa := S'_\kappa$  and

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

The notion of angle allows to define bounded curvature in several equivalent manners.

**Proposition 28.** *For  $\kappa \in \mathbf{R}$ , a geodesic space  $(S, d)$  has curvature lower (resp. upper) bounded by  $\kappa$  in the sense of Definition 24 if one of the two following equivalent statements holds.*

- (Quadruple comparison) *Any point in  $S$  has a neighbourhood  $U$  in which the following holds: for any four distinct points set  $\{p, x, y, z\} \subset U$ , we have*

$$\angle_\kappa(x, p, y) + \angle_\kappa(x, p, z) + \angle_\kappa(y, p, z) \leq (\text{resp. } \geq) 2\pi.$$

- (Angle monotonicity) *Any point in  $S$  has a neighbourhood  $U$  in which the following holds: for any any three distinct points set  $\{p, x, y\} \subset U$  and any constant speed shortest path  $\gamma$  from  $p$  to  $y$ , we have*

$$0 \leq s \leq t \leq 1 \quad \Rightarrow \quad \angle_\kappa(x, p, \gamma(s)) \geq (\text{resp. } \leq) \angle_\kappa(x, p, \gamma(t)).$$

Now for a point  $p$  in  $(S, d)$  geodesic, and two unit speed shortest paths  $\alpha : [0, \varepsilon] \rightarrow S$  and  $\beta : [0, \tau] \rightarrow S$  such that  $\alpha(0) = \beta(0) = p$ , the angle  $\angle_\kappa(\alpha, \beta)$  between  $\alpha$  and  $\beta$  is defined by

$$\angle_\kappa(\alpha, \beta) := \lim_{s,t \rightarrow 0} \angle_\kappa(\alpha(s), p, \beta(t)),$$

when it exists. If  $S$  is geodesic with upper or lower bounded curvature, the angle monotonicity guarantees that  $\angle_\kappa(\alpha, \beta)$  always exists. Next, we abbreviate  $\angle_0$  by  $\angle$ . Given a third unit speed shortest path  $\gamma : [0, \eta] \rightarrow S$  such that  $\gamma(0) = p$ , we have the triangular inequality

$$\angle(\alpha, \beta) \leq \angle(\alpha, \gamma) + \angle(\gamma, \beta). \quad (14)$$

Now for fixed  $p \in S$ , let  $\Sigma'_p$  be the set of all unit speed shortest paths emanating from  $p$ , two of them being identified if the angle between them is 0. On the set  $\Sigma'_p$ , the angle  $\angle$  defines a metric by (14) and we call space of directions at  $p$  the metric completion  $(\Sigma_p, \angle)$  of  $(\Sigma'_p, \angle)$  (with a slight abuse of notation, the metric in  $\Sigma_p$  is still denoted  $\angle$ ).

**Definition 29** (Tangent cone). *Let  $(S, d)$  be a geodesic space, with upper or lower bounded curvature in the sense of Definition 24. Let  $p \in S$ . The tangent cone  $T_p S$  at  $p$  is the Euclidean cone over the space of directions  $(\Sigma_p, \angle)$ . In other words,  $T_p S$  is the metric space:*

- Whose underlying set consists in equivalent classes in  $\Sigma_p \times [0, +\infty)$  for the equivalence relation  $\sim$  defined by

$$(\alpha, s) \sim (\beta, t) \Leftrightarrow ((s = 0 \text{ and } t = 0) \text{ or } (s = t \text{ and } \alpha = \beta)).$$

*With a slight abuse of notation, we'll consider that a point in  $T_p S$  is either the tip of the cone  $o_p$ , i.e. the class  $\Sigma_p \times \{0\}$ , or a couple  $(\alpha, s) \in \Sigma_p \times (0, +\infty)$  (hence identified to the class  $\{(x, s)\}$ ).*

- Whose metric  $d_p$  is defined by

$$d_p((\alpha, s), (\beta, t)) := \sqrt{s^2 + t^2 - 2st \cos \angle(\alpha, \beta)}.$$

For  $u = (\alpha, s)$  and  $v = (\beta, t) \in T_p S$ , we sometimes denote  $\|u - v\|_p := d_p(u, v)$ ,

$$\|u\|_p := d_p(o_p, u) = s \quad \text{and} \quad \langle u, v \rangle_p := st \cos \angle(\alpha, \beta) = (\|u\|_p^2 + \|v\|_p^2 - \|u - v\|_p^2)/2.$$

Next, we define the notion of logarithmic and exponential maps that play an important role in the sequel.

**Definition 30** (Logarithmic and exponential maps). *Let  $(S, d)$  be a geodesic space with curvature bounded from above or below in the sense of Definition 24 and fix  $p \in S$ . Let  $C(p)$  be the set of points  $x \in S$  such that there exists strictly more than one shortest path joining  $p$  to  $x$ . Then we denote*

$$\log_p : S \setminus C(p) \ni x \mapsto (\gamma_{px}, d(p, x)) \in T_p S,$$

*where  $\gamma_{px}$  denotes the unique unit speed shortest path from  $p$  to  $x$ . In particular, for any unit speed  $\gamma : [0, T] \rightarrow S$  emanating from  $p$  and any  $t \in [0, T]$ ,  $\log_p(\gamma(t)) = (\gamma, t)$ . Also, for a neighbourhood  $U \subset T_p S$  of  $o_p$ , we denote by  $\exp_p$  any function  $U \rightarrow S$  such that*

$$\log_p \circ \exp_p = Id_U$$

In the previous definition, note that  $C(p)$  is empty for spaces of curvature bounded from above. We end by mentioning a connection used in the paper between the curvature properties of  $S$  and that of the tangent cone.

**Proposition 31.** *Let  $(S, d)$  be a geodesic space and  $p \in S$  be fixed. If  $S$  is has curvature bounded from below in the large by some  $\kappa \in \mathbf{R}$ , then the tangent cone  $T_p S$  is a global NNC space. In addition, for any  $x, y \in S$ ,*

$$d(x, y) \leq \| \log_p(x) - \log_p(y) \|_p,$$

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

The first statement of the previous Proposition 31 is a special case of Proposition 28 in Yokota (2012). The second statement follows by combining the monotonicity property of angles in spaces with lower curvature bound (see Theorem 4.3.5 in Burago et al., 2001) with the Lang-Schroeder-Sturm characterisation of spaces of curvature bounded from below (see Sturm, 1999). We use several times in the proofs the following result (Theorem A in Yokota, 2012).

**Theorem 32.** *Let  $b^*$  be an exponential barycentre of  $P \in \mathcal{P}_2(S)$ . Denote  $l_P$  the image of the support of  $P$  through  $\log_{b^*}$ . Then  $l_P \subset T_{b^*}S$  can be isometrically embedded in a Hilbert space. The same goes with the convex hull of  $l_P$ .*

The Hilbert structure of the tangent cone can be seen as follows. For a point  $x = (\alpha, t)$  and  $\lambda \geq 0$ , we define  $\lambda x := (\alpha, \lambda t)$ . The sum between two points  $x$  and  $y$  is defined as the mid-point of  $2x$  and  $2y$ .

## References

- S. Alexander, V. Kapovitch, and A. Petrunin. *Alexandrov geometry*. Book in preparation, 2017.
- J. Altschuler, J. Weed, and P. Rigollet. Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. *Advances in Neural Information Processing*, 2017.
- L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Springer, 2008.
- P. Bartlett and S. Mendelson. Empirical minimization. *Probability Theory and Related Fields*, 135:311–334, 2006.
- P. Bartlett, O. Bousquet, and S. Mendelson. Local Rademacher complexities. *The Annals of Statistics*, 33:1497–1537, 2005.
- P. Bartlett, M. Jordan, and J. McAuliffe. Convexity, classification and risk bounds. *Journal of the American Statistical Association*, 101:138–156, 2006.
- J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):1111–1138, 2015.
- R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifold (part i). *The Annals of Statistics*, 31(1):1–29, 2003.
- J. Bigot, R. Gouet, T. Klein, and A. López. Minimax convergence rate for estimating the Wasserstein barycenter of random measures on the real line. *Electronic Journal Statistics*, 2018.
- G. Blanchard, G. Lugosi, and N. Vayatis. On the rate of convergence of regularized boosting classifiers. *Journal of Machine Learning Research*, 4:861–894, 2003.
- F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137:541–593, 2007.



- S. Boucheron, O. Bousquet, and G. Lugosi. Theory of classification: A survey of some recent advances. *ESAIM Probability and Statistics*, 9:323–375, 2005.
- O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *Comptes Rendus de l'Académie des Sciences de Paris*, 334:495–500, 2002.
- D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*. American Mathematical Society, 2001.
- M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems*, page 2292–2300, 2013.
- P. Dvurechensky, A. Gasnikov, and A. Kroshnin. Computational optimal transport: Complexity by accelerated gradient descent is better than by Sinkhorn’s algorithm. *arxiv:1802.04367*, 2018.
- M. Émery and G. Mokobodzki. Sur le barycentre d’une probabilité dans une variété. In *Séminaire de probabilités de Strasbourg*, volume 25, pages 220–233. Springer, 1991.
- A. Genevay, M. Cuturi, G. Peyré, and F. Bach. Stochastic optimization for large-scale optimal transport. *Advances in Neural Information Processing*, page 3440–3448, 2016.
- E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press, 2015.
- W.S. Kendall and H. Le. Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables. *Brazilian Journal of Probability and Statistics*, 25(3):323–352, 2011.
- V. Koltchinskii. Local Rademacher complexities and oracle inequalities in risk minimization. *The Annals of Statistics*, 34:2593–2656, 2006.
- V. Koltchinskii. *Oracle inequalities in empirical risk minimization and sparse recovery problems*. Springer, 2011.
- T. Le Gouic and J.-M. Loubes. Existence and consistency of Wasserstein barycenters. *Probability Theory and Related Fields*, 168:901–917, 2017.
- M. Ledoux and M. Talagrand. *Probability in Banach Spaces*. Springer, 1991.
- J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, 169:903–991, 2009.
- E. Mammen and A. Tsybakov. Smooth discriminant analysis. *The Annals of Statistics*, 27:1808–1829, 1999.
- P. Massart. Some applications of concentration inequalities to statistics. *Annales de la Faculté des Sciences de Toulouse*, 9:245–303, 2000.
- S. Mendelson. Improving the sample complexity using global data. *IEEE Transactions on Information Theory*, 48:1977–1991, 2002.
- S-I. Ohta. Convexities of metric spaces. *Geometriae Dedicata*, 125(1):225–250, 2007.

- S-I. Ohta. Barycenters in Alexandrov spaces of curvature bounded below. *Advances in geometry*, 14:571–587, 2012.
- K-T. Sturm. Probability measures on metric spaces of nonpositive curvature. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces: Lecture Notes from a Quarter Program on Heat Kernels, Random Walks, and Analysis on Manifolds and Graphs: April 16-July 13, 2002, Emile Borel Centre of the Henri Poincaré Institute, Paris, France*, 338:357, 2003.
- K.T. Sturm. Metric spaces of lower bounded curvature. *Expositiones Mathematicae*, 17(1):35–47, 1999.
- M. Talagrand. New concentration inequalities in product spaces. *Inventiones Mathematicae*, 126:505–563, 1996.
- R. van Handel. *Probability in high dimension*. Princeton Lecture Notes, 2016.
- Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.
- T. Yokota. A rigidity theorem in Alexandrov spaces with lower curvature bound. *Mathematische Annalen*, 353(2):305–331, 2012.
- T. Yokota. Convex functions and barycenter on CAT(1)-spaces of small radii. *Journal of the Mathematical Society of Japan*, 68(3):1297–1323, 2016.
- T. Yokota. Convex functions and  $p$ -barycenter on CAT(1)-spaces of small radii. *Tsukuba Journal of Mathematics*, 41(1):43–80, 2017.