

A BRENIER–STRASSEN THEOREM ON $\text{CAT}(\kappa)$ SPACES

NATHAEL GOZLAN, HUGO MALAMUT, AND SHIN-ICHI OHTA

ABSTRACT. We extend the Brenier–Strassen theorem about projections in convex order to non-flat spaces with curvature bounded from above. Precisely, for probability measures μ, ν of finite second moment on a complete separable $\text{CAT}(0)$ space, we prove that μ admits a unique W_2 -projection $\bar{\mu}$ to the set of probability measures dominated by ν in convex order. Moreover, the unique optimal coupling from μ to $\bar{\mu}$ is induced by a 1-Lipschitz map, without any absolute-continuity assumption on μ . Our proof identifies the projection problem with a weak optimal transport problem whose cost is the squared distance to the set of convex means. We also establish a localized version on $\text{CAT}(\kappa)$ spaces with $\kappa > 0$, where the optimal map is $1/2$ -Hölder continuous. Finally, we give a Strassen-type characterization of one-step barycentric martingales on proper $\text{CAT}(0)$ spaces.

Introduction

The aim of this article is to generalize a result about convex order projections in the Wasserstein space obtained in [24] to a non-Euclidean framework.

For a complete separable metric space (E, d) , we will denote by $\mathcal{P}_p(E)$, $p \geq 1$, the set of Borel probability measures μ such that $\int_E d^p(x_0, x) d\mu(x) < +\infty$ for some (and thus all) $x_0 \in E$. For $\mu, \nu \in \mathcal{P}_2(E)$, the quadratic Monge–Kantorovich distance, also known as Wasserstein distance, denoted by W_2 , is defined as

$$(0.1) \quad W_2^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{E \times E} d^2(x, y) \pi(dxdy), \quad \forall \mu, \nu \in \mathcal{P}_2(E),$$

where $\Pi(\mu, \nu)$ is the set of all transport plans from μ to ν , that is, the set of all probability measures π on $E \times E$ such that $\pi(A \times E) = \mu(A)$ and $\pi(E \times A) = \nu(A)$ for all Borel sets $A \subset E$. As is well known, there always exist transport plans realizing the infimum in (0.1); such plans are called optimal transport plans. The distance W_1 is defined analogously on $\mathcal{P}_1(E)$, with d in place of d^2 .

When $E = \mathbb{R}^d$ is equipped with the standard Euclidean distance d , the transport problem above is very well understood. According to a classical result of Brenier [9, 10], whenever μ is absolutely continuous, there is a unique optimal transport plan π^* for the problem (0.1) which is of the form $\pi^* = \text{Law}(X, \nabla\phi(X))$ with X a random variable of law μ and $\phi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ a convex function. The map $T = \nabla\phi$ is well defined μ -almost everywhere and is called the Brenier transport map from μ to ν . Regularity properties of the Brenier map have attracted a lot of attention; see in particular [12, 13]. As concerns existence of the Brenier map, the absolute continuity assumption on μ can be relaxed a bit; see [31, 22, 23]. It is however easy to construct examples where there is no transport

Date: July 30, 2026.

2020 Mathematics Subject Classification. Primary 49Q22; Secondary 60E15, 53C23, 60G48.

Key words and phrases. $\text{CAT}(\kappa)$ spaces; optimal transport; convex order; barycenter; Jensen inequality.

The first and second named author are supported by a grant of the Agence nationale de la recherche (ANR), Grant ANR-23-CE40-0017 (Project SOCOT). The third named author is supported by the JSPS Grant-in-Aid for Scientific Research (KAKENHI) 22H04942, 24K00523, 24K21511, 26H01996, and by the JST CREST JPMJCR25Q2. He is also grateful to Universität Bonn for its hospitality during his visit in Summer 2026, a part of this work was carried out there. This research has been conducted within the FP2M federation (CNRS FR 2036).

map, or examples where there are transport maps, but no optimal ones. There are also well known examples where the Brenier map exists but is not continuous.

In [24], a simple geometric condition has been identified for the existence and the Lipschitz continuity of an optimal transport map. To recall this result, let us introduce some additional notation. Two probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ are said to be in convex order, denoted by $\mu \leq_{\text{cvx}} \nu$, if $\int_{\mathbb{R}^d} f d\mu \leq \int_{\mathbb{R}^d} f d\nu$ for all convex functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$. If $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, there exists a unique probability measure $\bar{\mu} \leq_{\text{cvx}} \nu$ such that

$$(0.2) \quad W_2^2(\mu, \bar{\mu}) = \inf_{\eta \leq_{\text{cvx}} \nu} W_2^2(\mu, \eta).$$

In other words, $\bar{\mu}$ is the unique metric projection (also called the nearest point projection) of μ to the convex set $\{\eta \in \mathcal{P}_2(\mathbb{R}^d) : \eta \leq_{\text{cvx}} \nu\}$ of all probability measures dominated by ν in convex order. Existence and uniqueness of $\bar{\mu}$ were obtained in [24] or [1]. This convex order projection has been further studied by several authors [29, 5, 2]. It is naturally related to the weak optimal transport problem with a quadratic barycentric cost [26, 3], and admits interesting applications in terms of concentration of measure [25], sampling [1], and extrapolation of geodesics [21]. More importantly for the purpose of this article, it was shown in [24] that the optimal transport between μ and its projection $\bar{\mu}$ is always deterministic (that is, given by a transport map) and regular. Precisely, there exists a continuously differentiable convex function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ such that the map $T = \nabla \phi$ pushes μ forward to $\bar{\mu}$ and is 1-Lipschitz. The converse is also true: if ϕ is a differentiable convex function transporting a probability measure μ to another probability measure ν , and if $\nabla \phi$ is 1-Lipschitz, then $\bar{\mu} = \nu$. See [20] for a proof of Caffarelli's contraction theorem based on this idea.

We shall generalize this construction to $\text{CAT}(\kappa)$ spaces with $\kappa \geq 0$. Precise definitions will be recalled in Section 1. The class of $\text{CAT}(0)$ spaces contains Hilbert spaces and their closed convex subsets, metric trees, Euclidean buildings, and complete simply-connected Riemannian manifolds with nonpositive sectional curvature (in particular, hyperbolic spaces). We refer to [36, 11, 7] for an overview of the subject.

The case $\kappa > 0$ is treated in Theorems 2.1, 3.4 below. In the case $\kappa = 0$, we obtain the following result.

Theorem 1. *Suppose that (E, d) is a complete separable $\text{CAT}(0)$ space. For any $\mu, \nu \in \mathcal{P}_2(E)$, the following hold.*

- (1) *There exists a unique $\bar{\mu} \leq_{\text{cvx}} \nu$ such that $\inf_{\eta \leq_{\text{cvx}} \nu} W_2^2(\mu, \eta) = W_2^2(\mu, \bar{\mu})$.*
- (2) *There exists a unique W_2 -optimal transport plan from μ to $\bar{\mu}$ given by a 1-Lipschitz transport map T defined on the support of μ .*

Theorem 1 is deduced from the following structure theorem, which recasts the projection problem (0.2) as a weak optimal transport problem and provides a full description of its solutions. Below $C(q)$ denotes the set of all convex means of a probability measure q ; that is, the set of points $x \in E$ such that $f(x) \leq \int_E f dq$ for all lower semicontinuous convex functions $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ (see Section 1.3.4), and $\mu p = \nu$ means that a probability kernel $p = (p_x)_{x \in E}$ satisfies $\int_E p_x(dy) \mu(dx) = \nu(dy)$.

Theorem 2. *Let (E, d) be a complete separable $\text{CAT}(0)$ space and $\mu, \nu \in \mathcal{P}_2(E)$. Then*

$$\min_{\eta \leq_{\text{cvx}} \nu} W_2^2(\mu, \eta) = \min_{p: \mu p = \nu} \int_E d^2(x, C(p_x)) \mu(dx) = \max_{f: \text{convex}} \left\{ \int_E Qf d\mu - \int_E f d\nu \right\},$$

where the second minimum is taken over all probability kernels p such that $\mu p = \nu$, the maximum over all ν -integrable, lower semicontinuous, convex functions $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$, and $Qf(x) := \inf_{y \in E} \{f(y) + d^2(x, y)\}$. Solutions $\bar{\mu}, p, f$ of these problems are related by a unique map T defined μ -almost everywhere as follows:

- (1) T is the W_2 -optimal transport map from μ to $\bar{\mu}$;
- (2) For μ -almost all x , $T(x)$ is the metric projection of x to $C(p_x)$;
- (3) T is the proximal operator¹ associated with f in the sense that, for μ -almost all x ,

$$T(x) = \arg \min_{y \in E} \left\{ f(y) + d^2(x, y) \right\}.$$

The optimal transport map T from μ to $\bar{\mu}$ describes the first step of a coupling of μ and ν provided by an optimal weak transport plan. As in the Euclidean setting, the second step is a coupling between $\bar{\mu}$ and ν which satisfies a martingale property in the sense of Émery and Mokobodzki [19], but in general not in the (stronger) sense of Sturm [35] (see Remark 3.2 below for a precise statement).

The paper is organized as follows. In Section 1, we recall the definition of $\text{CAT}(\kappa)$ spaces with $\kappa \geq 0$, of convex order between probability measures, and the notions of barycenter and convex mean. In Section 2, we study the weak optimal transport problem associated with the cost function $c(x, p) = d^2(x, C(p))$. The first part of the section is devoted to continuity properties of this cost with respect to the product topology on $E \times \mathcal{P}_2(E)$, and the second part to the proof of a general version of Theorem 2, valid on $\text{CAT}(\kappa)$ spaces (Theorem 2.1). In Section 3, we study the regularity of the optimal transport map T from μ to its projection $\bar{\mu}$ and complete the proof of Theorem 1 as well as its generalization to $\text{CAT}(\kappa)$ spaces with $\kappa > 0$ (Theorem 3.4). Finally, in Section 4, we apply similar ideas to give a Strassen type characterization for barycentric martingales in proper $\text{CAT}(0)$ spaces.

1. $\text{CAT}(\kappa)$ spaces and convexity

Throughout the section, (E, d) will always be a complete separable geodesic space. We recall that the product space $E \times E$ equipped with the ℓ_2 distance

$$d_2((a, b), (c, d)) = \sqrt{d^2(a, c) + d^2(b, d)}, \quad (a, b), (c, d) \in E \times E,$$

is then also a complete separable geodesic space.

1.1. $\text{CAT}(\kappa)$ spaces

1.1.1. Definition. Let $M_0^2 := \mathbb{R}^2$ and, for $\kappa > 0$, let M_κ^2 denote the 2-sphere of radius $1/\sqrt{\kappa}$. Define d_κ as the corresponding geodesic distance on M_κ^2 . We set

$$D_\kappa := \begin{cases} +\infty, & \kappa = 0, \\ \pi/\sqrt{\kappa}, & \kappa > 0. \end{cases}$$

Definition 1.1. A geodesic metric space (E, d) is called a $\text{CAT}(\kappa)$ space if every geodesic triangle² of perimeter $< 2D_\kappa$ is no thicker than its comparison triangle in M_κ^2 . More precisely, if $\Delta(x, y, z) \subset E$ is a geodesic triangle with comparison triangle³ $\Delta(\bar{x}, \bar{y}, \bar{z}) \subset M_\kappa^2$, and $\bar{p}, \bar{q}$ are the comparison points⁴ of $p, q \in \Delta(x, y, z)$, then

$$d(p, q) \leq d_\kappa(\bar{p}, \bar{q}).$$

For broader background on spaces with curvature bounded above we refer to [11].

¹Our definition of proximal operator differs from the usual one by a factor 1/2.

²A geodesic triangle $\Delta(x, y, z)$ is the union $[x, y] \cup [y, z] \cup [z, x]$ of three geodesic segments; its perimeter is the sum of the lengths of the segments.

³A comparison triangle $\Delta(\bar{x}, \bar{y}, \bar{z}) \subset M_\kappa^2$ is a geodesic triangle in the model surface such that $d(x, y) = d_\kappa(\bar{x}, \bar{y})$, $d(y, z) = d_\kappa(\bar{y}, \bar{z})$, $d(z, x) = d_\kappa(\bar{z}, \bar{x})$.

⁴If $p \in [x, y]$, for instance, the comparison point $\bar{p} \in [\bar{x}, \bar{y}]$ is chosen so that $d(x, p) = d_\kappa(\bar{x}, \bar{p})$.

1.1.2. Examples. When $\kappa = 0$, one recovers the class of nonpositively curved spaces in the sense of Alexandrov, also called global NPC spaces or Hadamard spaces (when complete). In that case, an equivalent definition is the following property of the squared distance: for any geodesic $(x_t)_{t \in [0,1]}$ and any point z , we have

$$(1.1) \quad d^2(z, x_t) \leq (1-t) d^2(z, x_0) + t d^2(z, x_1) - t(1-t) d^2(x_0, x_1), \quad \forall t \in [0, 1].$$

Typical examples are Hilbert spaces, metric trees, Euclidean buildings, Hadamard manifolds (complete simply connected Riemannian manifolds with nonpositive sectional curvature), as well as products and closed convex subsets of such spaces. For the properties of such spaces, we refer to [11, 36].

For $\kappa > 0$, the basic model example is the round sphere $S_{1/\sqrt{\kappa}}^n$, and more generally geodesically convex subsets of the sphere and spherical buildings provide natural examples. We refer again to [11] for details and examples.

1.2. Convex functions

We recall that a subset $D \subset E$ is said to be (*geodesically*) *convex* if, for all $x_0, x_1 \in D$ and every geodesic $(x_t)_{t \in [0,1]}$ joining x_0 to x_1 , we have $x_t \in D$ for all $t \in [0, 1]$. A function $f : D \rightarrow \mathbb{R} \cup \{+\infty\}$ defined on a convex subset D of a geodesic metric space (E, d) is *convex* if for all geodesics $(x_t)_{t \in [0,1]}$ valued in D , it holds

$$(1.2) \quad f(x_t) \leq (1-t)f(x_0) + tf(x_1), \quad \forall t \in (0, 1).$$

One can slightly relax these conditions by requiring merely the existence of a geodesic satisfying the above properties (sometimes called weak convexity), however, it makes no difference in what follows since we consider the case where geodesics in question are unique.

If $D \subset E$ is a closed convex set, then a convex function $f : D \rightarrow \mathbb{R} \cup \{+\infty\}$ canonically extends to a convex function defined on E by simply taking value $+\infty$ outside of D ; we might also call this function f in the following. The same applies to lower semicontinuous (l.s.c. in short) functions.

The convexity is strict if the inequality (1.2) is strict for $t \in (0, 1)$ and $x_0 \neq x_1$. For $k \in \mathbb{R}$, f is called *k-convex* if we have

$$f(x_t) \leq (1-t)f(x_0) + tf(x_1) - \frac{k}{2} t(1-t) d^2(x_0, x_1), \quad \forall t \in (0, 1),$$

instead of (1.2). The simplest function to exhibit k -convexity is the squared distance: inequality (1.1) expresses precisely that, in CAT(0) spaces, the function $d^2(z, \cdot)$ is 2-convex on E , for every $z \in E$. When $\kappa > 0$, there is a local analog (see for example [32, Proposition 3.1(i)]):

Lemma 1.2. *Let (E, d) be a CAT(κ) space with $\kappa > 0$. Recall $D_\kappa = \pi/\sqrt{\kappa}$. Fix $\varepsilon \in (0, D_\kappa/2)$ and $z \in E$. Then the function $d^2(z, \cdot)$ is k -convex on the closed ball $\overline{B}(z, \varepsilon)$, where*

$$k := \frac{2\sqrt{\kappa}\varepsilon}{\tan(\sqrt{\kappa}\varepsilon)}.$$

That is to say, for every $x_0, x_1 \in \overline{B}(z, \varepsilon)$ and the geodesic $(x_t)_{t \in [0,1]}$ joining x_0 to x_1 , we have

$$d^2(z, x_t) \leq (1-t) d^2(z, x_0) + t d^2(z, x_1) - \frac{k}{2} t(1-t) d^2(x_0, x_1), \quad \forall t \in [0, 1].$$

1.3. Convex order

1.3.1. Definition. For $\eta, \nu \in \mathcal{P}_1(E)$, we will say that η is dominated by ν in *convex order*, denoted by $\eta \leq_{\text{cvx}} \nu$, if

$$(1.3) \quad \int_E f d\eta \leq \int_E f d\nu$$

for all l.s.c convex functions $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ such that $\int_E [f]_- d\eta < +\infty$ and $\int_E [f]_- d\nu < +\infty$, where $[f]_- := \max(-f, 0)$. If $\eta \leq_{\text{cvx}} \nu$, then a direct consequence of the definition of convex order is that for any closed convex set D , $\nu(D) = 1$ implies $\eta(D) = 1$.

1.3.2. Test functions. The following lemma shows that, under reasonable assumptions, the functions used to check the convex order according to (1.3) can be assumed to be continuous.

Lemma 1.3. *Let (E, d) be a complete separable $\text{CAT}(\kappa)$ space with $\kappa \geq 0$ and let $\eta, \nu \in \mathcal{P}_1(E)$.*

- (1) *If $\kappa = 0$, then $\eta \leq_{\text{cvx}} \nu$ if and only if (1.3) holds for all Lipschitz convex functions f bounded from below.*
- (2) *If $\kappa > 0$ and if η and ν are concentrated in a closed convex subset $D \subset E$ with $D \subset \overline{B}(x_0, \varepsilon)$ for some $x_0 \in E$ and $\varepsilon < D_\kappa/2$, then $\eta \leq_{\text{cvx}} \nu$ if and only if (1.3) holds for all convex, bounded, uniformly continuous functions $f : \overline{B}(x_0, \varepsilon) \rightarrow \mathbb{R}$. The same conclusion holds if $\eta, \nu \in \mathcal{P}_1(E \times E)$ are concentrated in $D \times D$.*

Proof. (1) Assume that the bound (1.3) holds for Lipschitz convex functions $f : E \rightarrow \mathbb{R}$ bounded from below. Let $\varphi : E \rightarrow \mathbb{R} \cup \{+\infty\}$ be a l.s.c. convex function bounded from below. Consider the sequence of functions $f_n : E \rightarrow \mathbb{R}$, $n \geq 1$, defined by $f_n(x) := \inf_{y \in E} \{\varphi(y) + nd(x, y)\}$, $x \in E$, $n \geq 1$. It is well known that f_n is n -Lipschitz continuous on E and converges pointwise monotonically to φ from below, as $n \rightarrow \infty$. Moreover, f_n is also convex on E . Indeed, the distance d being jointly convex (in other words, convex in the sense of Busemann; see e.g. [36, Corollary 2.5]), for two geodesics $(x_t)_{t \in [0, 1]}$ and $(y_t)_{t \in [0, 1]}$, we get

$$f_n(x_t) \leq \varphi(y_t) + nd(x_t, y_t) \leq (1-t)[\varphi(y_0) + nd(x_0, y_0)] + t[\varphi(y_1) + nd(x_1, y_1)]$$

and so, taking the infimum over $y_0, y_1 \in E$, gives

$$f_n(x_t) \leq (1-t)f_n(x_0) + tf_n(x_1).$$

By assumption, we have

$$\int_E f_n d\eta \leq \int_E f_n d\nu,$$

thus, by monotone convergence, letting $n \rightarrow \infty$ yields

$$\int_E \varphi d\eta \leq \int_E \varphi d\nu.$$

The assumption that φ is bounded from below can be removed by considering $\max(\varphi, -k)$ (which is still convex) and letting $k \rightarrow \infty$. So we conclude that it is enough to have (1.3) for all Lipschitz convex functions f which are bounded from below to ensure that $\eta \leq_{\text{cvx}} \nu$.

(2) The proof is similar, so we only point out main differences and we deal with the case of measures $\eta, \nu \in \mathcal{P}_1(E \times E)$ which is more general.

Set $B := \overline{B}(x_0, \varepsilon)$ with $\varepsilon < D_\kappa/2$. We use an approximation procedure from [38]. Given a l.s.c. convex function $\varphi : D \times D \rightarrow \mathbb{R}$ bounded from below, the sequence f_n is now defined by

$$f_n(x, y) := \inf_{(u, v) \in D \times D} \{\varphi(u, v) + n\Phi(x, u) + n\Phi(y, v)\}, \quad (x, y) \in B^2, \quad n \geq 1,$$

where $\Phi : B \times B \rightarrow \mathbb{R}$ is a jointly convex function satisfying

$$(1.4) \quad c_0 d^m(x, y) \leq \Phi(x, y) \leq C_0 d^m(x, y), \quad \forall x, y \in B,$$

for some $c_0, C_0 > 0$ and $m > 1$. The existence of such a function Φ is granted by [38, Theorem A] (see also (18) in [38]), which extends to $\text{CAT}(\kappa)$ spaces with $\kappa > 0$ a construction going back to [27, 28]. Reasoning exactly as in the first part of the proof, we see that f_n is convex on $B \times B$. Moreover, it can be checked on the explicit formula given in [38] that Φ is uniformly continuous on $B \times B$, so that f_n is uniformly continuous on $B \times B$ as well. It is also easy to check that f_n converges to φ pointwise monotonically. Then we conclude as in the first part of the proof. $\square$

1.3.3. Barycenters. Let (E, d) be a complete $\text{CAT}(0)$ space. We recall that, given $p \in \mathcal{P}_1(E)$, the *barycenter* of p is the unique minimizer of

$$y \mapsto \int_E \{d^2(y, x) - d^2(o, x)\} p(dx),$$

where $o \in E$ is an arbitrary fixed point. We refer to [36, Proposition 4.3] for existence and uniqueness. The barycenter of p does not depend on o , and will be denoted by $b(p)$ throughout what follows. According to e.g. [36, Theorem 6.2], any l.s.c. and convex function $f : E \rightarrow \mathbb{R}$ satisfies *Jensen's inequality*

$$(1.5) \quad \int_E f dp \geq f(b(p)).$$

Note that the integral on the left hand side is well defined; see [7, Lemma 2.3.7]. If $p \in \mathcal{P}_2(E)$, then $b(p)$ admits the more classical characterization of being the unique minimizer of

$$y \mapsto \int_E d^2(y, x) p(dx).$$

The barycenter map satisfies the following important contraction property (see [36, Theorem 6.3]):

$$(1.6) \quad d(b(p), b(q)) \leq W_1(p, q), \quad \forall p, q \in \mathcal{P}_1(E).$$

For $\kappa > 0$ there is also a local analog, due to Kuwae [30] and Yokota [38, Theorems B, 25].

Theorem 1.4. *Let (E, d) be a complete $\text{CAT}(\kappa)$ space for $\kappa > 0$. Let $p \in \mathcal{P}(E)$ and suppose that there exist $x_0 \in E$ and $\varepsilon \in (0, D_\kappa/2)$ such that $\text{supp } p \subset \overline{B}(x_0, \varepsilon)$. Then, the function $y \mapsto \int_E d^2(y, x) p(dx)$ admits a unique minimizer, denoted by $b(p) \in E$, which belongs to $\overline{B}(x_0, \varepsilon)$ and is called the barycenter of p . Moreover, for any l.s.c. convex function $f : \overline{B}(x_0, \varepsilon) \rightarrow \mathbb{R}$, Jensen's inequality (1.5) is satisfied.*

1.3.4. Convex means. Slightly extending the definition of [36], if (E, d) is a geodesic space and $p \in \mathcal{P}_1(E)$, we will say that a point $x \in E$ is a *convex mean* of p if $\delta_x \leq_{\text{cvx}} p$, that is,

$$f(x) \leq \int_E f dp$$

for any l.s.c. convex function $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ such that $\int_E [f]_- dp < +\infty$. If (E, d) is a $\text{CAT}(\kappa)$ space with $\kappa \geq 0$, then the class of convex test functions can be reduced, under the assumptions of Lemma 1.3. The set of all convex means of p will be denoted by $C(p)$ in the sequel. Whenever the barycenter $b(p)$ is available, Jensen's inequality (together with Lemma 1.3 to allow functions with values in $\mathbb{R} \cup \{+\infty\}$) gives $b(p) \in C(p)$. In particular, we know that $C(p)$ is nonempty for $p \in \mathcal{P}_1(E)$ in the $\kappa = 0$ case or in the setting of Theorem 1.4 in the $\kappa > 0$ case. The set $C(p)$ is always a closed convex subset of E . In the $\kappa = 0$ case, $C(p)$ is bounded for every $p \in \mathcal{P}_1(E)$, since for any $x \in C(p)$, it holds

$$d(x_0, x) \leq \int_E d(x_0, y) p(dy),$$

where x_0 is an arbitrary point in E , by the convexity of the function $d(x_0, \cdot)$. In the $\kappa > 0$ case, $C(p)$ is also obviously bounded, whenever p satisfies the assumptions of Theorem 1.4. If $D \subset E$ is a closed convex set such that $p(D) = 1$, then $C(p) \subset D$.

Lemma 1.5 (Stability along geodesics). *Let p_0, p_1 be two probability measures on E . Consider a geodesic $(x_t)_{0 \leq t \leq 1}$. If x_0 is a convex mean of p_0 and x_1 a convex mean of p_1 , then x_t is a convex mean of $(1-t)p_0 + tp_1$.*

Proof. Let f be a l.s.c. convex function on E . By the convexity of f and the assumptions on x_0 and x_1 ,

$$f(x_t) \leq (1-t)f(x_0) + tf(x_1) \leq (1-t) \int_E f dp_0 + t \int_E f dp_1 = \int_E f d((1-t)p_0 + tp_1). \quad \square$$

2. A weak optimal transport formulation

2.1. Definition and main result

Let (E, d) be a complete separable $\text{CAT}(\kappa)$ space with $\kappa \geq 0$. We define the weak cost function $c : E \times \mathcal{P}_2(E) \rightarrow [0, +\infty]$ by

$$c(x, p) := d^2(x, C(p)),$$

where $d(x, B) := \inf_{y \in B} d(x, y)$ denotes the usual distance from a point to a set and $C(p)$ is the set of convex means of p introduced in Section 1.3.4. If $C(p)$ is empty, we set $d(x, C(p)) = +\infty$. This section is devoted to studying the weak optimal transport problem

$$\mathcal{T}_c(\mu, \nu) := \inf_{p: \mu p = \nu} \int_E c(x, p_x) \mu(dx),$$

where the infimum runs over the set of probability kernels $p = (p_x)_{x \in E}$ such that $\int_E p_x(dy) \mu(dx) = \nu(dy)$. This class of transport problems was introduced in [26] (see also [3]); we refer to [6] for a survey and to [26, 3, 4, 8] for theoretical results (existence of optimizers, Kantorovich-type duality, cyclical monotonicity) available in this framework.

In the $\kappa > 0$ case, we will always assume for $\mu, \nu \in \mathcal{P}_2(E)$ that the support of ν is contained in a closed convex set D such that

$$(2.1) \quad \varepsilon_{\mu, \nu} := \sup_{x \in \text{supp } \mu} \sup_{y \in D} d(x, y) < \frac{D_\kappa}{2}.$$

To unify the cases, when $\kappa = 0$, we will always use the convention $D = E$.

The goal of this section is to prove the following theorem, which contains Theorem 2 in the case $\kappa = 0$.

Theorem 2.1. *Let (E, d) be a complete separable $\text{CAT}(\kappa)$ space with $\kappa \geq 0$, and let $\mu, \nu \in \mathcal{P}_2(E)$. In the $\kappa > 0$ case, we assume (2.1). Then*

$$\min_{\eta \leq_{\text{cvx}} \nu} W_2^2(\mu, \eta) = \min_{p: \mu p = \nu} \int_E d^2(x, C(p_x)) \mu(dx) = \max_{f: \text{convex}} \left\{ \int_E Qf d\mu - \int_E f d\nu \right\},$$

where the maximum is taken over all lower semicontinuous, ν -integrable, convex functions $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ and $Qf(x) := \inf_{y \in D} \{f(y) + d^2(x, y)\}$. The solution $\bar{\mu} \leq_{\text{cvx}} \nu$ of the former minimizing problem is unique. Moreover, solutions $\bar{\mu}, p, f$ of these three problems are related by a map T defined μ -almost everywhere as follows:

- (1) The plan $(\text{id}, T)_\# \mu$ is the unique optimal transport plan from μ to $\bar{\mu}$;
- (2) For μ -almost all x , $T(x)$ is the metric projection of x to $C(p_x)$;

(3) T is the proximal operator associated with the function f , i.e., for μ -almost all x ,

$$T(x) = \arg \min_{y \in D} \left\{ f(y) + d^2(x, y) \right\}.$$

Note that our definition of proximal operator differs from the usual one by a factor $1/2$ (in other words, T is the operator at time $1/2$; cf. [7]). Note also that in the case of the flat Euclidean space $E = \mathbb{R}^d$, the set of convex means of a measure p only contains its barycenter, and the cost c is simply the barycentric cost $|x - \int_{\mathbb{R}^d} y p(dy)|^2$; hence Theorem 2.1 generalizes [24, Theorem 1.2]. See also [29, 33] for similar projection problems involving general stochastic orders.

In the first part of the section, we study in detail the cost function $c : (x, p) \mapsto d^2(x, C(p))$: we show that it is continuous with respect to the product topology on $E \times \mathcal{P}_2(E)$ (with the appropriate localization when $\kappa > 0$). The second part of the section is devoted to the proof of Theorem 2.1.

2.2. Study of the cost function

Recall that c denotes the cost function defined in the preceding subsection.

Proposition 2.2. *Let (E, d) be a complete separable $\text{CAT}(\kappa)$ space with $\kappa \geq 0$.*

- (1) *If $\kappa = 0$, then c is continuous on $E \times \mathcal{P}_2(E)$ for the product topology, and is convex in the second variable with respect to convex interpolations.*
- (2) *When $\kappa > 0$, let D be a closed convex set included in a closed ball of radius $\varepsilon < D_\kappa/2$ and set $\tilde{D} := \{x \in E : \sup_{y \in D} d(x, y) \leq \varepsilon\}$. Then c is finite valued, continuous for the product topology, and convex in the second variable on $\tilde{D} \times \mathcal{P}_2(D)$ with respect to convex interpolations.*

Note that, under (2.1), $D \subset \overline{B}(x_0, \varepsilon_{\mu, \nu})$ holds for any $x_0 \in \text{supp } \mu$, and $\text{supp } \mu \subset \tilde{D}$ with $\varepsilon = \varepsilon_{\mu, \nu}$. We first record a variant of Cantor's intersection theorem for complete $\text{CAT}(\kappa)$ spaces which will be used in the proof; see e.g. [7, Proposition 2.1.16] and [14, Lemma 5.1].

Lemma 2.3 (Finite-intersection property). *Let (E, d) be a complete $\text{CAT}(\kappa)$ space with $\kappa \geq 0$. Let $(K_i)_{i \in I}$ be a family of nonempty closed convex subsets of E , such that every finite subfamily has nonempty intersection and one member of the family is included in a ball of radius $< D_\kappa/2$. Then $\bigcap_{i \in I} K_i \neq \emptyset$.*

For nonempty bounded closed sets $A, B \subset E$, let

$$d_H(A, B) := \max \left\{ \sup_{x \in A} d(x, B), \sup_{y \in B} d(y, A) \right\}$$

denote the Hausdorff distance between A and B . The proof of the continuity of c is based on the Lipschitz stability of the set of convex means with respect to the 1-Wasserstein distance.

Proposition 2.4. *Let (E, d) be a complete separable $\text{CAT}(\kappa)$ space with $\kappa \geq 0$. Let $p, q \in \mathcal{P}_1(E)$ be two probability measures supported in a convex closed subset $D \subset E$. When $\kappa > 0$, suppose that D is included in a closed ball B of radius $\varepsilon < D_\kappa/2$. Then the following hold.*

- (1) *For any coupling $\pi \in \Pi(p, q)$, we have*

$$C(p) = \text{pr}_1(C(\pi)), \quad C(q) = \text{pr}_2(C(\pi)),$$

where pr_1, pr_2 are the coordinate projections from $E \times E$ to E .

- (2) *As a consequence, we have*

$$\begin{aligned} d_H(C(p), C(q)) &\leq W_1(p, q) && \text{if } \kappa = 0, \\ d_H(C(p), C(q)) &\leq CW_1^\alpha(p, q) && \text{if } \kappa > 0, \end{aligned}$$

with constants $C > 0$ and $\alpha \in (0, 1)$ depending on κ and ε .

Note that (2) above extends in particular the Wasserstein contraction property of barycenters (1.6) to the set-valued context of convex means.

Proof of Proposition 2.4. Proof of item 1. By symmetry, it is enough to prove the statement for the first marginal, namely $C(p) = \text{pr}_1(C(\pi))$. The inclusion $\text{pr}_1(C(\pi)) \subset C(p)$ is immediate: if $(x, y) \in C(\pi)$ and $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ is a l.s.c. convex function, then $f \circ \text{pr}_1 : E \times E \rightarrow \mathbb{R} \cup \{+\infty\}$ is also l.s.c. convex and thus

$$f(x) \leq \int_{E \times E} f(u) \pi(du dv) = \int_E f dp,$$

provided that the integral makes sense.

Let us prove the converse inclusion. In the $\kappa = 0$ case, we set $D = B = E$ in order to unify the reasoning. Fix $x_0 \in C(p)$ and an arbitrary point $o \in E$. Let Φ_0 be the function defined by $\Phi_0(x, y) := d(o, y)$ in the case $\kappa = 0$, and $\Phi_0 := 0$ if $\kappa > 0$. Let $\Phi_1, \dots, \Phi_m$ be convex, uniformly continuous, π -integrable functions bounded from below on $B \times B$. Put

$$c_i := \int_{B \times B} \Phi_i d\pi, \quad 0 \leq i \leq m.$$

Note that $c_0 < \infty$ since $q \in \mathcal{P}_1(E)$. Define

$$L := \left\{ r \in \mathbb{R}^{m+1} : \exists y \in D \text{ such that } \Phi_i(x_0, y) \leq r_i, 0 \leq i \leq m \right\}.$$

Then L is convex and satisfies $L + \mathbb{R}_+^{m+1} = L$. Moreover, L is closed. Indeed, let $r^n \in L$ be such that $r^n \rightarrow r$, and y_n witnesses r^n . Let us show that $r \in L$. First, (y_n) is bounded. This is obvious in the $\kappa > 0$ case, since D is bounded. In the $\kappa = 0$ case, it holds $d(o, y_n) = \Phi_0(x_0, y_n) \leq r_0^n$, which implies the boundedness of (y_n) . By the finite-intersection property (Lemma 2.3) applied to the decreasing closed convex hulls $\overline{\text{co}}\{y_n : n \geq k\}$, there exists y in their intersection. For each k , $y \in \overline{\text{co}}\{y_n : n \geq k\}$, thus by the convexity and continuity of each $\Phi_i(x_0, \cdot)$,

$$\Phi_i(x_0, y) \leq \sup_{n \geq k} \Phi_i(x_0, y_n).$$

Hence, by letting $k \rightarrow \infty$,

$$\Phi_i(x_0, y) \leq \limsup_{n \rightarrow \infty} \Phi_i(x_0, y_n) \leq r_i,$$

and so $r \in L$.

We claim that $c = (c_0, \dots, c_m) \in L$. If not, since L is a closed convex upper set, the hyperplane separation theorem between L and $\{c\}$ gives some nonzero $a \in \mathbb{R}_+^{m+1}$ such that

$$\inf_{y \in D} \sum_{i=0}^m a_i \Phi_i(x_0, y) > \sum_{i=0}^m a_i c_i.$$

Set $g(x) := \inf_{y \in D} \sum_{i=0}^m a_i \Phi_i(x, y)$, $x \in B$, which is clearly bounded from below. As the infimum in one variable of a jointly convex uniformly continuous function, the function g is convex and uniformly continuous on B . Since $x_0 \in C(p)$,

$$g(x_0) \leq \int_B g dp = \int_{B \times B} g(x) \pi(dx dy) \leq \sum_{i=0}^m a_i c_i,$$

contradicting the above separation inequality. Thus $c \in L$.

Therefore, for every finite family $\Phi_1, \dots, \Phi_m$, there exists $y \in D$ such that

$$\Phi_i(x_0, y) \leq \int_{B \times B} \Phi_i d\pi, \quad 0 \leq i \leq m.$$

Equivalently, the closed convex sets

$$K_\Phi := \left\{ y \in D : \Phi(x_0, y) \leq \int_{B \times B} \Phi d\pi \right\},$$

indexed by convex, uniformly continuous, π -integrable and lower bounded functions Φ on $B \times B$, have the finite-intersection property, and K_{Φ_0} is bounded. By Lemma 2.3, there exists $y_0 \in \bigcap_\Phi K_\Phi$, which, according to Lemma 1.3, exactly means $(x_0, y_0) \in C(\pi)$. Thus $C(p) \subset \text{pr}_1(C(\pi))$.

Proof of item 2; case $\kappa = 0$. Let $\pi \in \Pi(p, q)$ be an optimal W_1 -coupling, and fix $x \in C(p)$. By item 1, there exists $y \in E$ such that $(x, y) \in C(\pi)$. Applying the defining inequality for $C(\pi)$ to the convex function $\Phi(u, v) := d(u, v)$ gives

$$d(x, y) \leq \int_{E \times E} d(u, v) \pi(du dv) = W_1(p, q).$$

Since $y \in C(q)$ again by item 1, we get $d(x, C(q)) \leq W_1(p, q)$. Taking the supremum over $x \in C(p)$, and arguing symmetrically on the second coordinate, we obtain

$$d_H(C(p), C(q)) \leq W_1(p, q).$$

Proof of item 2; case $\kappa > 0$. Let $\Phi : B \times B \rightarrow \mathbb{R}_+$ be the uniformly continuous jointly convex function used in the proof of Lemma 1.3. Reasoning as in the $\kappa = 0$ case yields

$$\max \left(\sup_{x \in C(p)} \inf_{y \in C(q)} \Phi(x, y), \sup_{y \in C(q)} \inf_{x \in C(p)} \Phi(x, y) \right) \leq \inf_{\pi \in \Pi(p, q)} \int_{D \times D} \Phi(u, v) \pi(du dv).$$

Then, using (1.4) and the boundedness of D , we conclude that

$$d_H(C(p), C(q)) \leq C W_1^\alpha(p, q)$$

for $C = (C_0/c_0)^{1/m} \text{diam}(D)^{(m-1)/m}$ and $\alpha = 1/m$ (for $m > 1$ in (1.4)). $\square$

Proof of Proposition 2.2. We give the proof in the $\kappa > 0$ case, the other case is simpler. By the triangle inequality, for all $x, y \in \tilde{D}$ and $p, q \in \mathcal{P}_2(D)$,

$$|d(x, C(p)) - d(y, C(q))| \leq d(x, y) + d_H(C(p), C(q)).$$

Together with the Hölder estimate of Proposition 2.4, it shows that the cost function c is continuous for the product topology with the 1-Wasserstein topology and hence with the 2-Wasserstein topology. Now, let us prove the convexity with respect to the second variable. Fix $x \in \tilde{D}$. By Lemma 1.2, the function $d^2(x, \cdot)$ is convex on the closed ball of center x and radius ε , and thus on D (by the definition of $\tilde{D}$). Let $p_0, p_1 \in \mathcal{P}_2(D)$ and put $p_t := (1-t)p_0 + tp_1 \in \mathcal{P}_2(D)$, $t \in [0, 1]$. According to Lemma 1.5, if $(y_t)_{t \in [0, 1]}$ is a geodesic connecting $y_0 \in C(p_0)$ to $y_1 \in C(p_1)$, then y_t belongs to $C(p_t)$. Hence

$$c(x, p_t) \leq d^2(x, y_t) \leq (1-t)d^2(x, y_0) + td^2(x, y_1).$$

Taking the infimum over y_0, y_1 then yields

$$c(x, p_t) \leq (1-t)c(x, p_0) + tc(x, p_1),$$

which completes the proof. $\square$

2.3. Convex and lower semicontinuous envelopes

Before turning to the proof of Theorem 2.1, we need to introduce the notions of convex and l.s.c. envelopes of a function. In this subsection, D is a convex subset of a geodesic space E .

Let $f : D \rightarrow \mathbb{R} \cup \{+\infty\}$ be a measurable function satisfying the lower bound $f \geq -a(1 + d^2(x_0, \cdot))$ for some $a \geq 0$ and $x_0 \in E$. The *convex envelope* of f is the function $\bar{f} : D \rightarrow \mathbb{R} \cup \{\pm\infty\}$ defined by

$$\bar{f}(x) := \inf \left\{ \int_D f \, dp : p \in \mathcal{P}_2(D), x \in C(p) \right\}.$$

Since $x \in C(\delta_x)$, the infimum is taken over a nonempty set, and $\bar{f}(x) \leq f(x)$ for every $x \in D$.

Lemma 2.5. *Let $f : D \rightarrow \mathbb{R} \cup \{+\infty\}$ be such that $\bar{f}$ does not take the value $-\infty$. Then the convex envelope $\bar{f}$ is convex. Moreover, if f is l.s.c. and convex, then $\bar{f} = f$.*

Proof. Let $x_0, x_1 \in D$ and $(x_t)_{t \in [0,1]}$ be a geodesic in D , and fix $p_0, p_1 \in \mathcal{P}_2(D)$ such that $x_i \in C(p_i)$ for $i = 0, 1$. By Lemma 1.5, $x_t \in C((1-t)p_0 + tp_1)$, thereby

$$\bar{f}(x_t) \leq \int_D f \, d((1-t)p_0 + tp_1) = (1-t) \int_D f \, dp_0 + t \int_D f \, dp_1.$$

Taking the infimum over all admissible p_0, p_1 gives

$$\bar{f}(x_t) \leq (1-t)\bar{f}(x_0) + t\bar{f}(x_1).$$

Now assume that $f : D \rightarrow \mathbb{R} \cup \{+\infty\}$ is l.s.c. and convex. We already know that $\bar{f} \leq f$. To see the converse, let $x \in D$. Since f is l.s.c. and convex, by the definition of convex mean, for all p with $x \in C(p)$ it holds

$$f(x) \leq \int_D f \, dp.$$

Taking the infimum over p in the right hand side yields $f(x) \leq \bar{f}(x)$. $\square$

Next, given a function $f : D \rightarrow \mathbb{R} \cup \{+\infty\}$, its *lower semicontinuous envelope* $\tilde{f}$ is defined as the greatest l.s.c. function below f . It is in fact given by

$$(2.2) \quad \tilde{f}(x) = \liminf_{y \rightarrow x} f(y) = \sup_{r > 0} \inf_{y \in B(x,r) \cap D} f(y), \quad \forall x \in D.$$

We will use the following result showing that the l.s.c. envelope of a convex function is still convex.

Lemma 2.6. *Let (E, d) be a complete $\text{CAT}(\kappa)$ space with $\kappa \geq 0$ and $D \subset E$ be a convex closed subset. When $\kappa > 0$, suppose that D is included in a ball of radius $\varepsilon < D_\kappa/2$. If $f : D \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex function such that $\tilde{f}$ does not take the value $-\infty$, then $\tilde{f} : D \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex as well.*

Proof. Fix $x_0, x_1 \in D$ and denote by γ the unique constant speed geodesic connecting x_0 to x_1 . Fix $t \in (0, 1)$ and let us prove that

$$(2.3) \quad \tilde{f}(\gamma(t)) \leq (1-t)\tilde{f}(x_0) + t\tilde{f}(x_1).$$

If $\tilde{f}(x_0) = +\infty$ or $\tilde{f}(x_1) = +\infty$, there is nothing to prove. So let us assume that both endpoint values are finite. By the sequential characterization (2.2) of the l.s.c. envelope, for each $i \in \{0, 1\}$ there is a sequence $(x_i^n)_n$ in D such that

$$x_i^n \longrightarrow x_i \quad \text{and} \quad f(x_i^n) \longrightarrow \tilde{f}(x_i).$$

Let $\gamma^n : [0, 1] \rightarrow D$ be the geodesic connecting x_0^n to x_1^n . According to [11, Proposition II.1.4(1)], geodesic segments in a $\text{CAT}(\kappa)$ space joining endpoints at distance less than D_κ are unique and depend (uniform) continuously on their endpoints. Hence $\gamma^n(t) \rightarrow \gamma(t)$ and, by (2.2),

$$\tilde{f}(\gamma(t)) \leq \liminf_{n \rightarrow \infty} f(\gamma^n(t)).$$

On the other hand, the convexity of f gives

$$f(\gamma^n(t)) \leq (1-t)f(x_0^n) + tf(x_1^n).$$

Taking the limits as $n \rightarrow \infty$ and using the convergence of the endpoint values, we obtain (2.3). $\square$

We will also need the following lemma showing that the infimum convolution operator Q (recall Theorem 2.1) does not see the difference between a function and its l.s.c. envelope.

Lemma 2.7. *For any $f : D \rightarrow \mathbb{R} \cup \{+\infty\}$, it holds $Qf = Q\tilde{f}$.*

Proof. By construction $\tilde{f} \leq f$, therefore

$$Q\tilde{f}(x) = \inf_{y \in D} \{\tilde{f}(y) + d^2(x, y)\} \leq \inf_{y \in D} \{f(y) + d^2(x, y)\} = Qf(x).$$

On the other hand, for all $x \in E$, it holds

$$Qf(x) \leq f(z) + d^2(x, z)$$

for all $z \in D$. Therefore, taking the $\liminf$ as $z \rightarrow y$ gives

$$Qf(x) \leq \tilde{f}(y) + d^2(x, y).$$

Taking the infimum in $y \in D$ yields $Qf(x) \leq Q\tilde{f}(x)$ and completes the proof. $\square$

2.4. Proof of the main result

Proof of Theorem 2.1. In the sequel, D denotes the closed convex set satisfying (2.1) and let $\varepsilon = \varepsilon_{\mu, \nu}$. Set $\tilde{D} := \{x \in E : \sup_{y \in D} d(x, y) \leq \varepsilon\}$ with the convention $D = \tilde{D} = E$ in the case $\kappa = 0$. Recall that, under (2.1), we have $\mu \in \mathcal{P}_2(\tilde{D})$ and $\nu \in \mathcal{P}_2(D)$.

By Proposition 2.2, the cost function $c : \tilde{D} \times \mathcal{P}_2(D) \rightarrow \mathbb{R}_+$ is continuous for the product topology and convex in its second variable. In particular, the condition (C) of [8] is satisfied. Moreover, for all $x \in \tilde{D}$ and $p \in \mathcal{P}_2(D)$, the triangle inequality and the convexity of $d^2(x_0, \cdot)$ yield

$$c(x, p) \leq 2d^2(x, x_0) + 2d^2(x_0, C(p)) \leq 2d^2(x, x_0) + 2 \int_D d^2(x_0, y) p(dy),$$

for any $x_0 \in \tilde{D}$. Thus, the condition (B) of [8] is also satisfied. Therefore, the fundamental theorem of weak optimal transport [8, Theorem 1.2] (see also [4, Theorems 1.2 and 1.3]) ensures that

$$(2.4) \quad \mathcal{T}_c(\mu, \nu) = \min_{p: \mu p = \nu} \left\{ \int_{\tilde{D}} c(x, p_x) \mu(dx) \right\} = \max_f \left\{ \int_{\tilde{D}} Q_c f d\mu - \int_D f d\nu \right\},$$

where the maximum is taken over ν -integrable functions $f : D \rightarrow \mathbb{R} \cup \{+\infty\}$, and

$$Q_c f(x) := \inf_{\rho \in \mathcal{P}_2(D)} \left\{ \int_D f d\rho + c(x, \rho) \right\}.$$

Note that $\int_{\tilde{D}} Q_c f d\mu$ always makes sense in $\mathbb{R} \cup \{-\infty\}$, and that the maximum can be further restricted to those f such that $\int_{\tilde{D}} Q_c f d\mu > -\infty$. Let f be such a function. Then, for all $y \in D$, taking $\rho = \delta_y$ in the definition of $Q_c f(x_0)$ implies

$$f(y) \geq Q_c f(x_0) - d^2(x_0, y),$$

where x_0 is any point in $\tilde{D}$ such that $Q_c f(x_0) > -\infty$. Let $\bar{f}$ be the convex envelope introduced in Section 2.3. Writing $c(x, \rho)$ as $\inf_{y \in C(\rho)} d^2(x, y)$ and exchanging the two infima, we get

$$\begin{aligned} Q_c f(x) &= \inf_{\rho \in \mathcal{P}_2(D)} \inf_{y \in C(\rho)} \left\{ \int_D f d\rho + d^2(x, y) \right\} = \inf_{y \in D} \inf_{\rho: y \in C(\rho)} \left\{ \int_D f d\rho + d^2(x, y) \right\} \\ &= \inf_{y \in D} \{ \bar{f}(y) + d^2(x, y) \}, \end{aligned}$$

that is, $Q_c f = Q\bar{f}$. This shows also that $\bar{f}$ does not take the value $-\infty$, otherwise $Q_c f$ would be identically $-\infty$. Denoting by $\tilde{f}$ the l.s.c. envelope of the convex envelope $\bar{f}$, we deduce from Lemma 2.7 that

$$Q_c f = Q\bar{f} = Q\tilde{f}.$$

Consequently,

$$\tilde{f}(y) \geq Q_c f(x_0) - d^2(x_0, y), \quad \forall y \in D,$$

so the negative part of $\tilde{f}$ is finite valued and ν -integrable. Moreover, since $\tilde{f} \leq \bar{f} \leq f$, the positive part of $\tilde{f}$ is also ν -integrable. Therefore $\tilde{f} \in L^1(\nu)$ and one gets

$$(2.5) \quad \int_{\tilde{D}} Q_c f d\mu - \int_D f d\nu \leq \int_{\tilde{D}} Q\tilde{f} d\mu - \int_D \tilde{f} d\nu.$$

Moreover, according to Lemma 2.6, $\tilde{f}$ is convex and l.s.c. Conversely, if $g : D \rightarrow \mathbb{R} \cup \{+\infty\}$ is a ν -integrable, convex and l.s.c. function, then Lemma 2.5 implies $\bar{g} = g$, so that $\tilde{g} = g$. Together with (2.5), it means that the maximum in (2.4) can be restricted to this class of functions:

$$(2.6) \quad \mathcal{T}_c(\mu, \nu) = \max_f \left\{ \int_{\tilde{D}} Qf d\mu - \int_D f d\nu \right\},$$

where the maximum is taken over all ν -integrable, convex and l.s.c. functions $f : D \rightarrow \mathbb{R} \cup \{+\infty\}$.

Let f be a ν -integrable, convex and l.s.c. function on D , and let $\eta \leq_{\text{cvx}} \nu$. The convex order gives $\int_D f d\eta \leq \int_D f d\nu$. Hence

$$\int_{\tilde{D}} Qf d\mu - \int_D f d\nu \leq \int_{\tilde{D}} Qf d\mu - \int_D f d\eta \leq W_2^2(\mu, \eta),$$

where the latter inequality follows from the definition of Qf . Taking the supremum over f on the left and the infimum over η on the right, we obtain

$$(2.7) \quad \mathcal{T}_c(\mu, \nu) \leq \inf_{\eta \leq_{\text{cvx}} \nu} W_2^2(\mu, \eta).$$

Let p be an optimal kernel for $\mathcal{T}_c(\mu, \nu)$. For μ -almost every x , the set $C(p_x)$ is a nonempty closed convex subset of D , and the metric projection $T(x)$ of x to $C(p_x)$ is well defined (see e.g. [36, Proposition 2.6] for $\kappa = 0$ and [30, Lemma 2.8] for $\kappa > 0$). Using Proposition 2.4 together with Lemma A.1 in Appendix A, it is easy to see that T is measurable, details are left to the reader. Set $\bar{\mu} := T_{\#}\mu$. Then

$$\mathcal{T}_c(\mu, \nu) = \int_{\tilde{D}} d^2(x, T(x)) \mu(dx) \geq W_2^2(\mu, \bar{\mu}).$$

Moreover, for every convex l.s.c. function g , since $T(x) \in C(p_x)$,

$$\int_D g d\bar{\mu} = \int_{\tilde{D}} g(T(x)) \mu(dx) \leq \int_{\tilde{D}} \int_D g(y) p_x(dy) \mu(dx) = \int_D g d\nu,$$

so that $\bar{\mu} \leq_{\text{cvx}} \nu$, and therefore

$$\mathcal{T}_c(\mu, \nu) \geq W_2^2(\mu, \bar{\mu}) \geq \inf_{\eta \leq_{\text{cvx}} \nu} W_2^2(\mu, \eta).$$

The opposite inequality (2.7) shows that equality holds and that $\bar{\mu}$ is a minimizer for the right hand side, thereby

$$\mathcal{T}_c(\mu, \nu) = \min_{\eta \leq_{\text{cvx}} \nu} W_2^2(\mu, \eta).$$

This yields the first assertion of Theorem 2.1. As for the optimality conditions, note that (1), (2) of Theorem 2.1 describe how we constructed T . Now it remains to prove the uniqueness of T and that T is the proximal operator for f as in (3).

Uniqueness of $\bar{\mu}$ and of the map T . Put $\eta_0 := \bar{\mu}$ and take $\eta_1 \leq_{\text{cvx}} \nu$ such that

$$W_2^2(\mu, \eta_1) = \mathcal{T}_c(\mu, \nu).$$

It follows from the convex order that $\eta_0(D) = \eta_1(D) = 1$. Let (Z, X_0, X_1) be random variables constructed on the same probability space such that (Z, X_0) and (Z, X_1) are optimal couplings for (μ, η_0) and (μ, η_1) , respectively (such random variables are easily obtained from any optimal couplings by a gluing argument, we omit details). We use the generalized geodesic between η_0 and η_1 with base μ : let $X_{1/2} := [X_0, X_1]_{1/2}$ be the midpoint of X_0 and X_1 , and let $\eta_{1/2}$ denote its law. For any l.s.c. convex function $g : D \rightarrow \mathbb{R}$, convexity gives $g(X_{1/2}) \leq \frac{1}{2}g(X_0) + \frac{1}{2}g(X_1)$, so that

$$\int_D g d\eta_{1/2} \leq \frac{1}{2} \int_D g d\eta_0 + \frac{1}{2} \int_D g d\eta_1 \leq \int_D g d\nu.$$

Hence $\eta_{1/2} \leq_{\text{cvx}} \nu$, which ensures

$$\mathbb{E} d^2(Z, X_{1/2}) \geq W_2^2(\mu, \eta_{1/2}) \geq \mathcal{T}_c(\mu, \nu).$$

On the other hand, the function $d^2(z, \cdot)$ is k -convex on D for every z in the support of μ , with $k = 2$ if $\kappa = 0$, and, by Lemma 1.2 together with condition (2.1), with $k = 2\sqrt{\kappa}\varepsilon / \tan(\sqrt{\kappa}\varepsilon)$ if $\kappa > 0$. Hence

$$d^2(Z, X_{1/2}) \leq \frac{1}{2}d^2(Z, X_0) + \frac{1}{2}d^2(Z, X_1) - \frac{k}{8}d^2(X_0, X_1).$$

Taking expectations and combining the two previous inequalities yields $\mathbb{E} d^2(X_0, X_1) \leq 0$, thus $X_0 = X_1$ almost surely. Therefore $\eta_1 = \bar{\mu}$ and the optimal coupling between μ and the projection $\bar{\mu}$ is unique (since the optimal couplings between μ and η_0, η_1 could be chosen arbitrarily), which is the one given by T .

T is the proximal operator associated with f . Let f be an optimal convex function in (2.6). Since $\bar{\mu} \leq_{\text{cvx}} \nu$, we have $\int_D f d\bar{\mu} \leq \int_D f d\nu$, and, for all $x \in \tilde{D}$, we have $Qf(x) \leq f(T(x)) + d^2(x, T(x))$ by the definition of Qf . Therefore

$$\begin{aligned} \mathcal{T}_c(\mu, \nu) &= \int_{\tilde{D}} Qf d\mu - \int_D f d\nu \leq \int_{\tilde{D}} Qf d\mu - \int_D f d\bar{\mu} = \int_{\tilde{D}} \{Qf(x) - f(T(x))\} \mu(dx) \\ &\leq \int_{\tilde{D}} d^2(x, T(x)) \mu(dx) = \mathcal{T}_c(\mu, \nu), \end{aligned}$$

so that all these inequalities become equality. In particular, for μ -almost every x ,

$$Qf(x) = f(T(x)) + d^2(x, T(x)),$$

that is, $T(x)$ attains the infimum defining $Qf(x)$. Moreover, the function $y \mapsto f(y) + d^2(x, y)$ is the sum of a convex function and a k -convex function with $k > 0$, thus it is k -convex and admits at most one minimizer. Consequently, for μ -almost every x , $T(x)$ is the unique minimizer of $y \mapsto f(y) + d^2(x, y)$, which is the desired proximal characterization as in (3). $\square$

2.5. Consequences of Theorem 2.1

The strong duality established in Theorem 2.1 — the existence of an optimal convex function f playing the role of a Lagrange multiplier — has two useful consequences, recorded below: a weak cyclical monotonicity property satisfied by optimal kernels, and a version of Strassen’s theorem adapted to the convex order.

Corollary 2.8. *In the setting of Theorem 2.1, any kernel p optimal for the problem*

$$\min_{p: \mu p = \nu} \int d^2(x, C(p_x)) \mu(dx)$$

is c -monotone, with $c(x, p) = d^2(x, C(p))$. In particular, there exists a subset $A \subset \text{supp } \mu$ with $\mu(A) = 1$ such that for all $x, y \in A$ and all $q_1, q_2 \in \mathcal{P}_2(D)$ satisfying

$$q_1 + q_2 = p_x + p_y,$$

it holds

$$c(x, p_x) + c(y, p_y) \leq c(x, q_1) + c(y, q_2).$$

Proof. The c -monotonicity follows from [8, Corollary 2.13]. Indeed, as shown in the proof of Theorem 2.1, Assumptions (B) and (C) of [8] are granted. Thus, for any optimal kernel p , the corresponding coupling $\pi(dx dy) = p_x(dy) \mu(dx)$ is c -monotone, in the sense that there exists $\Gamma \subset \tilde{D} \times \mathcal{P}_2(D)$ such that $\mu(\{x : (x, p_x) \in \Gamma\}) = 1$ and, for every $(x_1, \rho_1), \dots, (x_n, \rho_n) \in \Gamma$ and all $\tilde{\rho}_1, \dots, \tilde{\rho}_n \in \mathcal{P}_2(D)$ with

$$\sum_{i=1}^n \rho_i = \sum_{i=1}^n \tilde{\rho}_i,$$

it holds

$$\sum_{i=1}^n c(x_i, \rho_i) \leq \sum_{i=1}^n c(x_i, \tilde{\rho}_i).$$

Letting $A := \{x \in \text{supp } \mu : (x, p_x) \in \Gamma\}$ and taking $(x_1, \rho_1) = (x, p_x)$ and $(x_2, \rho_2) = (y, p_y)$ for $x, y \in A$ shows the claim. $\square$

The following result is a version of the well known Strassen’s theorem [34] for the convex order.

Theorem 2.9 (Strassen’s theorem for the convex order). *Let (E, d) be a complete separable $\text{CAT}(\kappa)$ space with $\kappa \geq 0$, and let $\mu, \nu \in \mathcal{P}_2(E)$. Assume that (2.1) holds in the case $\kappa > 0$. Then $\mu \leq_{\text{cvx}} \nu$ if and only if there exists a probability kernel $p = (p_x)_{x \in E}$ such that $\mu p = \nu$ and $x \in C(p_x)$ for μ -almost every $x \in E$.*

Proof. The existence of $p = (p_x)_{x \in E}$ such that $\mu p = \nu$ and $x \in C(p_x)$ for μ -almost every $x \in E$ immediately implies $\mu \leq_{\text{cvx}} \nu$. Conversely, suppose that $\int_D g d\mu \leq \int_D g d\nu$ for all l.s.c. convex functions $g : E \rightarrow \mathbb{R} \cup \{+\infty\}$. Since $Qg \leq g$ on D and $\mu(D) = 1$, we also have $\int_D Qg d\mu \leq \int_D g d\nu$. Therefore, by the duality formula of Theorem 2.1, we get $\mathcal{T}_c(\mu, \nu) = 0$. Then an optimal kernel p satisfies $\int_D d^2(x, C(p_x)) \mu(dx) = 0$, and hence $x \in C(p_x)$ for μ -almost all x . $\square$

Remark 2.10. *In the $\kappa > 0$ case, the assumption (2.1) of Theorem 2.9 is not optimal. For instance, assuming E is a locally compact $\text{CAT}(\kappa)$ space with $\kappa > 0$, Theorem 2.3 of [16] together with item 2. of Proposition 1.3, shows that if $\mu \leq_{\text{cvx}} \nu$ and ν is supported on a closed convex set D included in a ball $\overline{B}(x_0, \varepsilon)$ with $\varepsilon < D_\kappa/2$, then there exists a probability kernel $p = (p_x)_{x \in E}$ such that $\mu p = \nu$ and $x \in C(p_x)$ for μ -almost every $x \in E$. We refer to [16, 17, 18] and [33] for other variants of Strassen theorem involving more general cone of functions.*

3. Regularity of the optimal map

3.1. Key convexity estimate for the regularity

The following result studies the solutions of the weak optimal transport problem given by Theorem 2.1.

Proposition 3.1. *Let (E, d) be a complete separable $\text{CAT}(\kappa)$ space with $\kappa \geq 0$, and let $\mu, \nu \in \mathcal{P}_2(E)$. In the $\kappa > 0$ case, we assume (2.1). Let T be the optimal map given by Theorem 2.1. There exists a subset $A \subset \text{supp } \mu$ with $\mu(A) = 1$ such that for all $x, y \in A$,*

$$(3.1) \quad kd^2(T(x), T(y)) \leq d^2(x, T(y)) + d^2(y, T(x)) - d^2(x, T(x)) - d^2(y, T(y)),$$

with $k = 2$ if $\kappa = 0$ and $k = 2\sqrt{\kappa}\varepsilon / \tan(\sqrt{\kappa}\varepsilon)$ if $\kappa > 0$, where $\varepsilon = \varepsilon_{\mu, \nu}$.

Proof. Let p be an optimal kernel for $\mathcal{T}_c(\mu, \nu)$. By Corollary 2.8, for any $x, y \in A$ (given in Corollary 2.8) and $q_1, q_2 \in \mathcal{P}_2(D)$ such that $q_1 + q_2 = p_x + p_y$, it holds

$$c(x, p_x) + c(y, p_y) \leq c(x, q_1) + c(y, q_2).$$

Taking $q_1 = (1-t)p_x + tp_y$ and $q_2 = (1-t)p_y + tp_x$, we get

$$d^2(x, T(x)) + d^2(y, T(y)) \leq d^2(x, C(q_1)) + d^2(y, C(q_2)).$$

Let $[T(x), T(y)]_{t \in [0,1]}$ be the geodesic joining $T(x)$ to $T(y)$. Following Lemma 1.5, $[T(x), T(y)]_t$ is a convex mean of q_1 , and $[T(y), T(x)]_t$ is a convex mean of q_2 . So we get

$$d^2(x, T(x)) + d^2(y, T(y)) \leq d^2(x, [T(x), T(y)]_t) + d^2(y, [T(y), T(x)]_t).$$

In the case $\kappa > 0$, Lemma 1.2 together with (2.1) yields that the function $d^2(x, \cdot)$ is k -convex on D for any x in the support of μ , with $k = 2\sqrt{\kappa}\varepsilon / \tan(\sqrt{\kappa}\varepsilon)$. For $\kappa = 0$, this holds with $k = 2$. Thus we have

$$\begin{aligned} & d^2(x, [T(x), T(y)]_t) + d^2(y, [T(y), T(x)]_t) \\ & \leq (1-t)d^2(x, T(x)) + td^2(x, T(y)) + (1-t)d^2(y, T(y)) + td^2(y, T(x)) - kt(1-t)d^2(T(x), T(y)). \end{aligned}$$

Putting everything together, we get

$$k(1-t)d^2(T(x), T(y)) \leq d^2(x, T(y)) + d^2(y, T(x)) - d^2(x, T(x)) - d^2(y, T(y)).$$

Letting $t \rightarrow 0$ yields the claimed inequality (3.1). $\square$

3.2. Non-positive curvature spaces

We are now ready to complete the proof of Theorem 1.

Proof of Theorem 1. The existence and uniqueness of $\bar{\mu}$ and T follow from Theorem 2.1. It remains to prove that T is 1-Lipschitz. By Proposition 3.1 with $D = E$, it holds

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

for $x, y \in A$, with $\mu(A) = 1$. According to Reshetnyak's quadruple comparison (see [36, Proposition 2.4]), we have

$$d^2(x, T(y)) + d^2(y, T(x)) - d^2(x, T(x)) - d^2(y, T(y)) \leq 2d(x, y)d(T(x), T(y)).$$

This yields

$$d(T(x), T(y)) \leq d(x, y),$$

for all $x, y \in A$. Since $\mu(A) = 1$, A is dense inside the support of μ . Therefore, T admits a unique 1-Lipschitz version defined on the support of μ which completes the proof. $\square$

Beyond the existence of the projection $\bar{\mu}$, the proof of Theorem 1 sheds light on the structure of the optimal weak transport plan from μ to ν .

Remark 3.2. *As in the Euclidean case, the optimal weak transport plan from μ to ν (see Theorem 2) decomposes into a deterministic 1-Lipschitz first step, namely the transport map T from μ to $\bar{\mu}$ given by Theorem 1, followed by a second non-deterministic step coupling $\bar{\mu}$ and ν . More precisely, since $\bar{\mu} \leq_{\text{cvx}} \nu$, by applying Theorem 2.9, one obtains a kernel $(q_y)_{y \in E}$ such that $\bar{\mu}q = \nu$ and $\delta_y \leq_{\text{cvx}} q_y$ for $\bar{\mu}$ -almost every y , which is precisely the definition of a martingale coupling in the sense of Émery and Mokobodzki [19]. Such couplings are in general not unique. Since y is only a convex mean of q_y and need not be its barycenter, this coupling is moreover not always a two time-step martingale in the (stronger) sense of Sturm [35].*

As consequences of Theorem 1, the convex order projection enjoys two properties of independent interest: a Pythagorean-type inequality and a non-expansiveness property (generalizing [2, Proposition 2.2]).

Proposition 3.3. *Let (E, d) be a complete separable $\text{CAT}(0)$ space and let $\nu \in \mathcal{P}_2(E)$. For $\mu \in \mathcal{P}_2(E)$, denote by $\bar{\mu}$ the convex order projection of μ to $\{\eta \in \mathcal{P}_2(E) : \eta \leq_{\text{cvx}} \nu\}$ given by Theorem 1. Then we have the following.*

- (1) $W_2^2(\mu, \nu) \geq W_2^2(\mu, \bar{\mu}) + W_2^2(\bar{\mu}, \nu)$ for every $\mu \in \mathcal{P}_2(E)$.
- (2) $W_2(\bar{\mu}_1, \bar{\mu}_2) \leq W_2(\mu_1, \mu_2)$ for every $\mu_1, \mu_2 \in \mathcal{P}_2(E)$.

Proof. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a non-atomic Polish probability space, and consider the space $L^2(\Omega, E)$ of square integrable E -valued random variables, equipped with the metric $D(X, Y) := \sqrt{\mathbb{E}[d^2(X, Y)]}$. According to [36, Proposition 3.10], $(L^2(\Omega, E), D)$ is itself a $\text{CAT}(0)$ space, whose geodesics are given pointwise by those of E : the geodesic from X_0 to X_1 is $X_t(\omega) = [X_0(\omega), X_1(\omega)]_t$. Since Ω is non-atomic, $W_2^2(\mu_1, \mu_2)$ coincides with $\min\{D^2(X_1, X_2) : \text{Law}(X_i) = \mu_i\}$, the minimum being attained.

The set $\tilde{K} := \{Z \in L^2(\Omega, E) : \text{Law}(Z) \leq_{\text{cvx}} \nu\}$ is convex by testing convex functions pointwise along geodesics. It is also closed, since by Lemma 1.3, the convex order can be tested against only Lipschitz functions bounded from below. Let T denote the optimal 1-Lipschitz transport map from μ to $\bar{\mu}$ provided by Theorem 1. Then for every random variable $X \in L^2(\Omega, E)$ of law μ , the random variable $\bar{X} := T \circ X$ is the metric projection of X to $\tilde{K}$. Indeed, $\bar{X} \in \tilde{K}$ with $D^2(X, \bar{X}) = W_2^2(\mu, \bar{\mu})$, while any $Z \in \tilde{K}$ satisfies $D^2(X, Z) \geq W_2^2(\mu, \text{Law}(Z)) \geq W_2^2(\mu, \bar{\mu})$.

Now suppose that (X, Y) gives an optimal coupling of μ and ν . The variable Y lies in $\tilde{K}$, so we can apply the Pythagorean-type inequality for metric projections in $\text{CAT}(0)$ spaces (see [7, Theorem 2.1.12(ii)]), which implies

$$W_2^2(\mu, \nu) = D^2(X, Y) \geq D^2(X, \bar{X}) + D^2(\bar{X}, Y) \geq W_2^2(\mu, \bar{\mu}) + W_2^2(\bar{\mu}, \nu).$$

This shows item (1).

Item (2) follows similarly from the non-expansiveness of metric projections in $\text{CAT}(0)$ spaces (see [7, Theorem 2.1.12(iii)]) applied to two random variables X_1, X_2 of laws μ_1, μ_2 realizing $D(X_1, X_2) = W_2(\mu_1, \mu_2)$. Namely, denoting by $\bar{X}_i$ the metric projection of X_i to $\tilde{K}$, we have

$$W_2(\bar{\mu}_1, \bar{\mu}_2) \leq D(\bar{X}_1, \bar{X}_2) \leq D(X_1, X_2). \quad \square$$

3.3. Spaces with curvature bounded from above

In the $\kappa > 0$ setting, Theorem 1 still admits a partial extension, provided that the measures are sufficiently close. One still obtains continuity of the optimal transport map, but only with 1/2-Hölder regularity instead of 1-Lipschitz regularity.

Theorem 3.4. *Let (E, d) be a complete separable $\text{CAT}(\kappa)$ space with $\kappa > 0$, and let $\mu, \nu \in \mathcal{P}_2(E)$. We assume (2.1). Then the optimal transport map T from μ to $\bar{\mu}$ given by Theorem 2.1 is Hölder continuous with exponent $1/2$. More precisely, for all x, y in the support of μ ,*

$$d(T(x), T(y)) \leq \sqrt{\frac{2 \tan(\sqrt{\kappa} \varepsilon)}{\sqrt{\kappa}}} d(x, y)^{1/2},$$

with $\varepsilon = \varepsilon_{\mu, \nu}$.

Proof. According to Proposition 3.1, for any two points $x_1, x_2 \in A$,

$$kd^2(T(x_1), T(x_2)) \leq d^2(x_1, T(x_2)) + d^2(x_2, T(x_1)) - d^2(x_1, T(x_1)) - d^2(x_2, T(x_2)),$$

with $k = 2\sqrt{\kappa}\varepsilon/\tan(\sqrt{\kappa}\varepsilon)$. Using the identity $a^2 - b^2 = (a - b)(a + b)$ on the right hand side, we obtain

$$kd^2(T(x_1), T(x_2)) \leq \sum_{i=1}^2 [d(x_{3-i}, T(x_i)) - d(x_i, T(x_i))] [d(x_{3-i}, T(x_i)) + d(x_i, T(x_i))].$$

By the triangle inequality, for both $i = 1, 2$,

$$d(x_{3-i}, T(x_i)) - d(x_i, T(x_i)) \leq d(x_1, x_2),$$

and by (2.1),

$$d(x_{3-i}, T(x_i)) + d(x_i, T(x_i)) \leq 2\varepsilon.$$

Therefore

$$kd^2(T(x_1), T(x_2)) \leq 4\varepsilon d(x_1, x_2).$$

Since $k = 2\sqrt{\kappa}\varepsilon/\tan(\sqrt{\kappa}\varepsilon)$, this yields

$$d^2(T(x_1), T(x_2)) \leq \frac{2 \tan(\sqrt{\kappa} \varepsilon)}{\sqrt{\kappa}} d(x_1, x_2).$$

Since A is dense in the support of μ , we conclude that T admits a Hölder continuous version on $\text{supp } \mu$. $\square$

4. A Strassen type result for barycentric martingales

Let (E, d) be a $\text{CAT}(0)$ space. We assume in this section that (E, d) is proper (i.e. closed balls in E are compact), thus complete and separable. Given two probability measures $\mu, \nu \in \mathcal{P}_2(E)$, it is well known that the existence of a martingale in the sense of Émery and Mokobodzki [19] is equivalent to the convex domination of μ by ν . More precisely, according to e.g. Theorem 2.9, $\mu \leq_{\text{cvx}} \nu$ if and only if there exists a pair of random variables (M_0, M_1) with $M_0 \sim \mu$ and $M_1 \sim \nu$ such that for μ -almost all $x \in E$, $x \in C(\text{Law}(M_1 | M_0 = x))$, where we recall that $C(p)$ denotes the set of convex means of a probability measure p on E .

The aim of this section is to give a similar characterization for the existence of a martingale in the sense of Sturm [35, 15]. More precisely, we would like to characterize the set of couples (μ, ν) , $\mu, \nu \in \mathcal{P}_2(E)$, such that there exists a *barycentric martingale* (M_0, M_1) with $M_0 \sim \mu$ and $M_1 \sim \nu$, that is a pair of random variables such that

$$M_0 = \mathbb{E}[M_1 | M_0] \quad \text{a.s.},$$

which means that $M_0 = b(\text{Law}(M_1 | M_0))$ a.s., where $b(p)$ denotes the barycenter of a probability measure $p \in \mathcal{P}_2(E)$ (recall Section 1.3.3).

For a measurable function $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ with $f \geq -a(1 + d^2(x_0, \cdot))$ for some $a \geq 0$ and $x_0 \in E$, let us define its *barycentric envelope* $\text{bar}(f) : E \rightarrow \mathbb{R} \cup \{\pm\infty\}$ by

$$\text{bar}(f)(x) := \inf \left\{ \int_E f \, dp : p \in \mathcal{P}_2(E), b(p) = x \right\}.$$

Remark 4.1. Note that the function $\text{bar}(f)$ is different from the convex envelope $\bar{f}$ considered in Section 2.3. It always holds $\bar{f} \leq \text{bar}(f) \leq f$. The function $\text{bar}(f)$ is not convex in general, while $\bar{f}$ always is (recall Lemma 2.5).

Remark 4.2. Let $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ be a l.s.c. function with $f \geq -a(1 + d^2(x_0, \cdot))$ for some $a \geq 0$ and $x_0 \in E$. Then f is convex if and only if $\text{bar}(f) = f$. Indeed, $\text{bar}(f) = f$ exactly means that

$$f(b(p)) \leq \int_E f \, dp.$$

Taking $p = (1-t)\delta_{x_0} + t\delta_{x_1}$ yields $f(x_t) \leq (1-t)f(x_0) + tf(x_1)$, with $(x_t)_{t \in [0,1]}$ the geodesic from x_0 to x_1 . Thus f is convex. Conversely, if f is convex, then it satisfies Jensen's inequality (1.5) which amounts to $\text{bar}(f) = f$.

Lemma 4.3. Let f be a l.s.c. function, bounded from below and with $f \leq a(1 + d^2(x_0, \cdot))$ for some $a \geq 0$ and $x_0 \in E$. Then $\text{bar}(f)$ is measurable.

Proof. Assume, without loss of generality, that $0 \leq f \leq a(1 + d^2(x_0, \cdot))$. For each $\varepsilon > 0$, define

$$\text{bar}_\varepsilon(f)(x) := \inf \left\{ \int_E (f + \varepsilon d^2(x_0, \cdot)) \, dp : p \in \mathcal{P}_2(E), b(p) = x \right\}, \quad x \in E.$$

Then, it is easily seen that $\text{bar}(f) = \inf_{k \geq 1} \text{bar}_{1/k}(f)$. Let us show that, for all $\varepsilon > 0$, $\text{bar}_\varepsilon(f)$ is l.s.c., this will imply in particular that $\text{bar}(f)$ is measurable.

Let (x_n) be a sequence converging to $x \in E$, and $p_n \in \mathcal{P}_2(E)$ be such that $b(p_n) = x_n$ and

$$\text{bar}_\varepsilon(f)(x_n) \geq \int_E (f + \varepsilon d^2(x_0, \cdot)) \, dp_n - r_n$$

for some $r_n > 0$ converging to 0. Let (n_k) be a sequence of integers such that $\text{bar}_\varepsilon(f)(x_{n_k}) \rightarrow \liminf_{n \rightarrow \infty} \text{bar}_\varepsilon(f)(x_n)$. Since

$$\text{bar}_\varepsilon(f)(x_{n_k}) \leq f(x_{n_k}) + \varepsilon d^2(x_0, x_{n_k}) \leq a(1 + d^2(x_0, x_{n_k})) + \varepsilon d^2(x_0, x_{n_k}),$$

we conclude that

$$(4.1) \quad \sup_{k \geq 1} \int_E d^2(x_0, y) p_{n_k}(dy) < +\infty.$$

Therefore, since E is proper, it follows from Prokhorov's theorem that the sequence (p_{n_k}) is pre-compact for the weak topology. Using again (4.1) and [37, Theorem 7.12], we see that any converging subsequence of (p_{n_k}) is also converging for the W_1 distance. Thus, extracting a subsequence if necessary, we can assume that $p_{n_k} \rightarrow p$ for W_1 . Then, by the continuity of the barycenter for the W_1 distance (which follows from (1.6)), we have $b(p_{n_k}) \rightarrow b(p)$, and so $b(p) = x$ and

$$\lim_{k \rightarrow \infty} \text{bar}_\varepsilon(f)(x_{n_k}) \geq \int_E (f + \varepsilon d^2(x_0, \cdot)) \, dp$$

where the inequality follows from the fact that, for any l.s.c. non-negative function g , the map $p \mapsto \int_E g \, dp$ is l.s.c. for the weak topology on $\mathcal{P}(E)$. In particular, we see that $p \in \mathcal{P}_2(E)$, and hence

$$\int_E (f + \varepsilon d^2(x_0, \cdot)) \, dp \geq \text{bar}_\varepsilon(f)(x)$$

from which follows that

$$\liminf_{n \rightarrow \infty} \text{bar}_\varepsilon(f)(x_n) \geq \text{bar}_\varepsilon(f)(x),$$

thus $\text{bar}_\varepsilon(f)$ is l.s.c. □

Theorem 4.4. *Let (E, d) be a proper CAT(0) space and $\mu, \nu \in \mathcal{P}_2(E)$. Then the following are equivalent.*

- (1) *There exists a barycentric martingale (M_0, M_1) such that $M_0 \sim \mu$ and $M_1 \sim \nu$.*
- (2) *For every l.s.c. function $f : E \rightarrow \mathbb{R}$, bounded from below and with $f \leq a(1 + d^2(x_0, \cdot))$ for some $a > 0$ and $x_0 \in E$, it holds*

$$\int_E \text{bar}(f) d\mu \leq \int_E f d\nu.$$

Proof. (1) $\Rightarrow$ (2). Let (M_0, M_1) be a barycentric martingale with marginals μ, ν . Denoting $p_x = \text{Law}(M_1 | M_0 = x)$, we get

$$\mathbb{E}[f(M_1) | M_0 = x] = \int_E f dp_x \geq \text{bar}(f)(x),$$

so $\mathbb{E}[f(M_1) | M_0] \geq \text{bar}(f)(M_0)$ a.s. and thus $\mathbb{E}[f(M_1)] \geq \mathbb{E}[\text{bar}(f)(M_0)]$.

(2) $\Rightarrow$ (1). For $p \in \mathcal{P}_2(E)$, define its *variance* as

$$V(p) := \inf_{x \in E} \int_E d^2(x, y) p(dy) = \int_E d^2(b(p), y) p(dy).$$

We consider the weak optimal transport problem between μ and ν associated to the cost function

$$C(x, p) := \int_E d^2(x, y) p(dy) - V(p) \geq 0.$$

We set

$$\mathcal{T}_C(\mu, \nu) := \inf_p \int_E C(x, p_x) \mu(dx),$$

where the infimum runs over the set of probability kernels $p = (p_x)_{x \in E}$ such that $\mu p = \nu$. For fixed $z \in E$, the map $p \mapsto \int_E d^2(z, y) p(dy)$ is linear, so V is concave (with respect to convex interpolations) as the infimum of linear functions. The cost function $C(x, p)$ is therefore convex in p , and is also continuous on $E \times \mathcal{P}_2(E)$ equipped with the product topology. Indeed, the continuity of $(x, p) \mapsto \int_E d^2(x, y) p(dy)$ is classical, and V satisfies

$$|\sqrt{V(p)} - \sqrt{V(q)}| \leq W_2(p, q),$$

by Minkowski's inequality. According to [4, Theorems 1.2 and 1.3], the infimum in $\mathcal{T}_C(\mu, \nu)$ is attained, and it holds

$$(4.2) \quad \mathcal{T}_C(\mu, \nu) = \sup_f \left\{ \int_E R_C f d\mu - \int_E f d\nu \right\},$$

where the supremum runs over continuous functions f bounded from below with $f \leq a(1 + d^2(x_0, \cdot))$ for some $a \geq 0$ and $x_0 \in E$, and

$$R_C f(x) := \inf_{p \in \mathcal{P}_2(E)} \left\{ \int_E f dp + C(x, p) \right\}.$$

For $p \in \mathcal{P}_2(E)$ with $b(p) = x$, we have $C(x, p) = 0$, thereby

$$R_C f(x) \leq \text{bar}(f)(x).$$

Combining this with the assumption (2), we get

$$\int_E R_C f d\mu \leq \int_E \text{bar}(f) d\mu \leq \int_E f d\nu,$$

and so (4.2) yields $\mathcal{T}_C(\mu, \nu) \leq 0$. Since C is non-negative, we obtain $\mathcal{T}_C(\mu, \nu) = 0$. If p is the primal optimizer, we thus get

$$(4.3) \quad \int_E C(x, p_x) \mu(dx) = 0.$$

By the uniqueness of barycenter, $C(x, p) = 0$ if and only if $b(p) = x$, thus (4.3) ensures that $b(p_x) = x$ for μ -almost every x . The coupling $\pi(dxdy) = p_x(dy)\mu(dx)$ then provides a barycentric martingale with marginals μ and ν , which completes the proof. $\square$

Appendix A. Measurability of the transport map

Lemma A.1 (Joint continuity of metric projections). *Let E be a complete $\text{CAT}(\kappa)$ space with $\kappa \geq 0$ and $D \subset E$ be a closed convex subset. When $\kappa > 0$, we take $U \subset E$ such that*

$$\sup_{x \in U} \sup_{y \in D} d(x, y) < D_\kappa/2,$$

and set $U = E$ when $\kappa = 0$. For any nonempty closed convex subsets K_n, K of D and $x_n, x \in U$, if $x_n \rightarrow x$ and $d_H(K_n, K) \rightarrow 0$, then we have $P_{K_n} x_n \rightarrow P_K x$, where $P_K x$ denotes the metric projection of x to K .

Proof. Put $z_n = P_{K_n} x_n$, $z = P_K x$, and $h_n = d_H(K_n, K)$. Choose $y_n \in K_n$ and $w_n \in K$ such that

$$d(y_n, z) \leq d(z, K_n) + n^{-1} \leq h_n + n^{-1}, \quad d(w_n, z_n) \leq d(z_n, K) + n^{-1} \leq h_n + n^{-1}.$$

By assumption, we have $y_n \rightarrow z$ and $d(w_n, z_n) \rightarrow 0$. By the choice of z_n ,

$$d(x_n, z_n) \leq d(x_n, y_n),$$

so that $\limsup_{n \rightarrow \infty} d(x_n, z_n) \leq d(x, z)$. On the other hand,

$$d(x, z) \leq d(x, w_n) \leq d(x, x_n) + d(x_n, z_n) + h_n + n^{-1},$$

and hence $d(x_n, z_n) \rightarrow d(x, z)$ as well as $d(x, w_n) \rightarrow d(x, z)$.

On the relevant common ball, by Lemma 1.2, the functions $d^2(x, \cdot)$ are uniformly k -convex for some $k > 0$. The variational inequality for the projection to K (see [30, Lemma 2.8]) gives

$$d^2(x, w_n) - d^2(x, z) \geq \frac{k}{2} d^2(w_n, z).$$

Thus $w_n \rightarrow z$, and

$$d(z_n, z) \leq d(z_n, w_n) + d(w_n, z) \rightarrow 0. \quad \square$$

Declaration on the use of generative AI

During the preparation of this work, the authors used ChatGPT 5.5 (OpenAI) in the following two ways. First, as an assistance for exploring some of the arguments of the paper. The most significant instance is Proposition 2.4: we had only conjectured the statement of item 2, and its complete proof, going through item 1, was suggested to us by ChatGPT 5.5. Second, for proofreading the manuscript. All statements and proofs obtained with this assistance were subsequently checked and written up by the authors, who reviewed and edited the content as needed and take full responsibility for the content of this article.

References

- [1] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of probability measures in the convex order by Wasserstein projection. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 56(3):1706–1729, 2020.
- [2] A. Alfonsi and B. Jourdain. Wasserstein projections in the convex order: regularity and characterization in the quadratic Gaussian case. Preprint, arXiv:2506.23981 [math.PR] (2025), 2025.
- [3] J.-J. Alibert, G. Bouchitté, and T. Champion. A new class of costs for optimal transport planning. *Eur. J. Appl. Math.*, 30(6):1229–1263, 2019.
- [4] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calc. Var. Partial Differ. Equ.*, 58(6):28, 2019. Id/No 203.
- [5] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Weak monotone rearrangement on the line. *Electron. Commun. Probab.*, 25:Paper No. 18, 16, 2020.
- [6] J. Backhoff-Veraguas and G. Pammer. Applications of weak transport theory. *Bernoulli*, 28(1):370–394, 2022.
- [7] M. Bačák. *Convex analysis and optimization in Hadamard spaces*, volume 22 of *De Gruyter Series in Nonlinear Analysis and Applications*. De Gruyter, Berlin, 2014.
- [8] M. Beiglböck, G. Pammer, L. Riess, and S. Schrott. The Fundamental Theorem of Weak Optimal Transport. Preprint, arXiv:2501.16316 [math.PR] (2025), 2025.
- [9] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. (Polar decomposition and increasing rearrangement of vector fields). *C. R. Acad. Sci., Paris, Sér. I*, 305:805–808, 1987.
- [10] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.*, 44(4):375–417, 1991.
- [11] M. R. Bridson and A. Haefliger. *Metric Spaces of Non-Positive Curvature*, volume 319 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 1999.
- [12] L. A. Caffarelli. The regularity of mappings with a convex potential. *J. Am. Math. Soc.*, 5(1):99–104, 1992.
- [13] L. A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Communications in Mathematical Physics*, 214(3):547–563, Nov. 2000.
- [14] P.-E. Capraro and A. Lytchak. At infinity of finite-dimensional CAT(0) spaces. *Math. Ann.*, 346(1):1–21, 2010.
- [15] T. Christiansen and K. T. Sturm. Expectations and martingales in metric spaces. *Stochastics*, 80(1):1–17, 2008.
- [16] K. J. Ciosmak. Applications of Strassen’s theorem and Choquet theory to optimal transport problems, to uniformly convex functions and to uniformly smooth functions. *Nonlinear Anal., Theory Methods Appl., Ser. A, Theory Methods*, 232:32, 2023. Id/No 113267.
- [17] K. J. Ciosmak. Localisation for constrained transports I: theory. *arXiv preprint arXiv:2312.12281*, 2023.
- [18] K. J. Ciosmak. Corrigendum to: “Applications of Strassen’s theorem and Choquet theory to optimal transport problems, to uniformly convex functions and to uniformly smooth functions”. *Nonlinear Anal., Theory Methods Appl., Ser. A, Theory Methods*, 244:5, 2024. Id/No 113542.
- [19] M. Émery and G. Mokobodzki. Sur le barycentre d’une probabilité dans une variété. *Séminaire de probabilités de Strasbourg*, 25:220–233, 1991.
- [20] M. Fathi, N. Gozlan, and M. Prod’homme. A proof of the Caffarelli contraction theorem via entropic regularization. *Calc. Var. Partial Differ. Equ.*, 59(3):18, 2020. Id/No 96.
- [21] T. O. Gallouët, A. Natale, and G. Todeschi. Metric extrapolation in the Wasserstein space. *Calc. Var. Partial Differ. Equ.*, 64(5):35, 2025. Id/No 147.
- [22] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.
- [23] N. Gigli. On the inverse implication of Brenier-McCann theorems and the structure of $(\mathcal{P}_2(M), W_2)$. *Methods Appl. Anal.*, 18(2):127–158, 2011.
- [24] N. Gozlan and N. Juillet. On a mixture of Brenier and Strassen theorems. *Proceedings of the London Mathematical Society*, 120(3):434–463, 2020.
- [25] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali. Characterization of a class of weak transport-entropy inequalities on the line. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 54(3):1667–1693, 2018.
- [26] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, 273(11):3327–3405, 2017.
- [27] W. Jäger and H. Kaul. Uniqueness and stability of harmonic maps and their Jacobi fields. *Manuscripta Math.*, 28(1-3):269–291, 1979.
- [28] W. S. Kendall. Convexity and the hemisphere. *J. London Math. Soc. (2)*, 43(3):567–576, 1991.
- [29] Y.-H. Kim and Y. Ruan. Backward and forward Wasserstein projections in stochastic order. *J. Funct. Anal.*, 286(2):63, 2024. Id/No 110201.
- [30] K. Kuwae. Jensen’s inequality on convex spaces. *Calculus of Variations and Partial Differential Equations*, 49(3–4):1359–1378, 2014.
- [31] R. J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.
- [32] S.-I. Ohta. Convexities of metric spaces. *Geometriae Dedicata*, 125(1):225–250, 2007.

- [33] F. Pramenković. Weak optimal transport: When is the dual potential convex? *arXiv preprint arXiv:2507.07200*, 2025.
- [34] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Stat.*, 36:423–439, 1965.
- [35] K.-T. Sturm. Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature. *The Annals of Probability*, 30(3):1195–1222, 2002.
- [36] K.-T. Sturm. Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, volume 338 of *Contemporary Mathematics*, pages 357–390. American Mathematical Society, Providence, RI, 2003.
- [37] C. Villani. *Topics in optimal transportation*, volume 58 of *Grad. Stud. Math.* Providence, RI: American Mathematical Society (AMS), 2003.
- [38] T. Yokota. Convex functions and barycenter on $\text{CAT}(1)$ -spaces of small radii. *Journal of the Mathematical Society of Japan*, 68(3):1297–1323, 2016.

UNIVERSITÉ PARIS CITÉ, CNRS, MAP5, F-75006 PARIS, FRANCE

Email address: `nathael.gozlan@u-pariscite.fr`

UNIVERSITÉ PARIS CITÉ, CNRS, MAP5, F-75006 PARIS, FRANCE

Email address: `hugo.malamut@u-pariscite.fr`

UNIVERSITY OF OSAKA, OSAKA 560-0043, JAPAN & RIKEN CENTER FOR ADVANCED INTELLIGENCE PROJECT (AIP),
1-4-1 NIHONBASHI, TOKYO 103-0027, JAPAN

Email address: `s.ohta@math.sci.osaka-u.ac.jp`