

Wasserstein barycenters on metric graphs

Jérôme Bertrand^{*1} and Jianyu Ma ^{†1}

¹Univ Toulouse, INSA Toulouse, CNRS, IMT, Toulouse, France.

Abstract

In this paper we solve the barycenter problem on the quadratic Wasserstein space over a possibly infinite metric graph relative to a probability measure \mathbb{P} . Assuming that \mathbb{P} gives mass to absolutely continuous measures with respect to the 1d-Hausdorff measure on the metric graph, we prove that such a barycenter is absolutely continuous outside the set of vertices where an atomic part can exist. This result is optimal. We also show that Wasserstein barycenters are not unique in general.

Contents

1	Introduction	1
2	Preliminaries	4
2.1	Metric (measure) graphs	4
2.2	Wasserstein spaces	8
2.2.1	Measurable selection	9
2.2.2	Barycenter	11
2.3	Wasserstein space over the real line	11
2.3.1	Quantile functions	11
2.3.2	Integral formula for Wasserstein barycenters on the real line	13
3	Fine properties of Optimal Transport over a metric graph	15
3.1	Restriction property	15
3.2	Subsets of geodesics in a metric graph and induced properties	18
4	Branched covering & quasi regularity	20
4.1	A branched covering map from $\mathcal{W}_2(G)$ to $\mathcal{W}_2(\mathbb{R})$	20
4.2	Proof of the main results	25
4.2.1	When the barycenter is concentrated in an edge	25
4.2.2	Proof of Theorem 4.5	27
4.2.3	Proof of Proposition 1.3	28
	Bibliography	32

^{*}jerome.bertrand@math.univ-toulouse.fr

[†]jianyu.ma@math.univ-toulouse.fr

1 Introduction

This paper is devoted to the study of Wasserstein barycenters over a metric graph (G, d) . By metric graph, we mean a finite or countable set of vertices connected by edges viewed as actual paths homeomorphic to compact intervals of the real line. We always assume this graph to be connected and to have finite degree at each point. We equip this graph with a length distance and its associated 1d-Hausdorff measure which turns G into a proper, complete, and separable geodesic space. We refer to Section 2 for precise definitions.

Given a probability measure \mathbb{P} on a metric space (X, d) , a barycenter of \mathbb{P} is a minimum of the functional $z \mapsto \int_X d^2(z, x) d\mathbb{P}(x)$. This condition coincides with the standard definition when (X, d) is the n -dimensional Euclidean space.

In this paper, we consider the case where the metric space is the space of probability measures with finite second order moment $(\mathcal{W}_2(G), d_W)$ with (G, d) a metric graph.

The (quadratic Kantorovitch-)Wasserstein distance on $\mathcal{W}_2(G)$ is defined by

$$d_W(\mu, \nu) := \left(\min_{\pi \in \Gamma(\mu, \nu)} \int_{X \times X} d^2(x, y) d\pi(x, y) \right)^{1/2},$$

where $\Gamma(\mu, \nu)$ is the set of probability measures on $G \times G$ whose marginals are μ and ν respectively. We refer to Section 2.2 for more details.

Existence of a barycenter on a non-proper metric space may fail, see [24] for a rather complete collection of counter-examples, but barycenters over $(\mathcal{W}_2(X), d_W)$ do exist whenever the underlying metric space (X, d) is proper, complete, and separable [22]. Such barycenters are called *Wasserstein barycenters*.

The problem of studying Wasserstein barycenters (aka the Wasserstein barycenter problem) and its regularised version have drawn a lot of interest for its many applications in statistics, image processing, and data science [26]. In statistics, Wasserstein barycenter can be used as a mean on a non-linear space, and, among other things, results analogous to the law of large numbers do exist, see for instance [22, 9]; estimating the convergence rate of empirical barycenters towards the actual barycenter is also a question of interest, the obtained rates depend on geometric properties of the space like in [2]. Wasserstein barycenters are also of primary importance in the field of metric geometry as a tool to characterise synthetic curvature bounds. Indeed to a probability measure $\mathbb{P} = (1 - \lambda)\delta_{\mu_0} + \lambda\delta_{\mu_1}$ corresponds a collection of barycenters for λ varying in $[0, 1]$ which induces a geodesic curve $(\mu_\lambda)_{\lambda \in [0, 1]}$ in the Wasserstein space, and k -convexity of Entropy-like functionals on the Wasserstein space characterises lower Ricci curvature bound, see for instance [38, 5, 36, 15] for precise statements and (much) more on this subject. More precisely, barycenters and curvature bounds are related through the so-called Jensen inequality that we now report. Given a (geodesically) convex functional F on a (geodesic) metric space (X, d) , and a measure \mathbb{P} on (X, d) admitting a barycenter $z \in X$, Jensen's inequality holds if

$$F(z) \leq \int_X F(x) d\mathbb{P}(x).$$

Inequalities of this kind have been proved in various geometric setups: when (X, d) is an Aleksandrov space of curvature bounded below [32] or when (X, d) is a quadratic Wasserstein space over Euclidean space, a (compact) Riemannian manifold or a metric measure space with synthetic Ricci curvature bounded below (aka RCD spaces) [1, 21, 25, 18].

Uniqueness of the barycenter is false in general. A classical example is that of $\mathbb{P} = 1/2(\delta_x + \delta_{-x})$ where x and $-x$ are two antipodal points on a unit sphere equipped with its canonical metric $(X, d) = (\mathbb{S}^2, \text{can})$, in that situation any point of the equator relative to x and $-x$ is a barycenter of \mathbb{P} . Uniqueness of the barycenter holds if the space (X, d) is a metric space of nonpositive curvature thanks to strong convexity properties satisfied by $(x, y) \mapsto d(x, y)$ [35]. For Wasserstein barycenters, while nonnegative curvature bound is preserved from the base space (X, d) to its Wasserstein space $(\mathcal{W}_2(X), d_W)$, this is *not* the case for nonpositive curvature bound despite some nonpositive curvature like properties that do hold on $(\mathcal{W}_2(X), d_W)$ when (X, d) is a nonpositively curved metric space [38, 8]. However, another fruitful approach is available: if the probability measure \mathbb{P} gives mass to measures μ in $\mathcal{W}_2(X)$ such that for any $\nu \in \mathcal{W}_2(X)$, the distance $d_W(\mu, \nu)$ is induced by a transport map:

$$d_W(\mu, \nu)^2 = \int d^2(x, T(x)) \, d\mu(x), \quad (1)$$

where $T : X \rightarrow X$ satisfies $T_{\#}\mu = \nu$, then the Wasserstein barycenter of \mathbb{P} is *unique*, see for instance [34, 24] for a detailed proof. The property appearing in (1) is known to hold whenever the measure μ is absolutely continuous with respect to a reference measure for metric measure spaces satisfying some regularity properties. Originally this was proved by Brenier in the Euclidean space equipped with the Lebesgue measure, then Brenier's result was extended to very general settings, see for instance [28, 7, 17]. Besides, Wasserstein space $(\mathcal{W}_2(X), d_W)$ is known to admit a tangent space $T_{\mu}\mathcal{W}_2(X)$, which can be described in analytical or geometrical terms, at any absolutely continuous measure μ when (X, d) is sufficiently regular; this was used at the formal level by Otto [30], see [4] for a proof in the Euclidean case. Consequently, it is natural to seek for conditions under which a Wasserstein barycenter is absolutely continuous with respect to the reference measure, this barycenter is then said to be *regular*.

The first breakthrough regarding the regularity of Wasserstein barycenters is a paper by Agueh and Carlier who proved the result in Euclidean space [1]. This result was then generalised to compact manifolds [21], compact Aleksandrov spaces (of curvature bounded below) [20], and possibly non-compact Riemannian manifolds with Ricci curvature bounded below [25]. There is also a recent preprint [18] where the authors prove the absolute continuity of Wasserstein barycenter provided that \mathbb{P} -a.e. surely $\text{Ent}_m(\mu) < \infty$ ¹ and the base space is a metric measure space (X, d, m) with synthetic Ricci curvature bounded below ((X, d, m) is a RCD space). It is important to keep in mind that none of these results applies to metric graphs because it is a branching space and/or the curvature at each vertex is $-\infty$ (in other terms, a metric graph is *not* an Aleksandrov space of curvature bounded below not even a metric measure space where the Entropy functional is k -convex (aka $CD(k, \infty)$ spaces) as proved in [14]). Indeed, in the case of a metric (measure) graph, actually even on a metric tree, it is known that the absolute continuity of the barycenter with respect to the 1d-Hausdorff measure cannot hold in general. We provide an example from [24] inspired by [19]. Consider the tripod in Figure 1, formed by three copies of the unit interval $[0, 1]$ joined at a common origin o . Let $\mathbb{P} := \frac{1}{3} \sum_{i=1}^3 \delta_{\nu_i}$ be a probability measure on the Wasserstein space over the tripod, where each ν_i is an absolutely continuous measure supported on the outer half $[\frac{1}{2}, 1]$ of a distinct branch. The unique Wasserstein barycenter of \mathbb{P} is the Dirac measure $\mu_{\mathbb{P}} = \delta_o$ at the central vertex.

More in general, the absence of synthetic lower curvature bounds on metric (measure) graphs forces one to tackle each standard question in optimal mass transport with a specific approach.

¹this condition is probably not necessary for the barycenter being a.c.

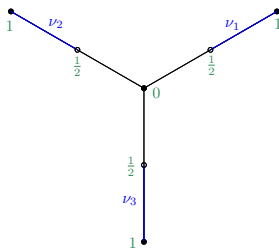


Figure 1: $\mathbb{P} = \sum_{i=1}^3 \frac{1}{3} \delta_{\nu_i}$ on the tripod

The Monge problem on metric graphs (where $d_W(\mu, \nu) := \min_{\pi \in \Gamma(\mu, \nu)} \int_{X \times X} d(x, y) d\pi(x, y)$ is used instead of the distance we mentioned above) has been studied in [27]. In [14], the authors prove that the Wasserstein distance on a finite metric graph can be described in a more fluid dynamics framework known as Benamou-Brenier formula. We refer to [6, 3] for more on the Benamou-Brenier formula when the base space is the Euclidean space.

Our main result concerns the quasi-regularity of Wasserstein barycenters over a metric graph. This result is optimal in view of the example described in Figure 1.

Theorem 1.1. *Let (G, d, \mathcal{H}) be a metric measure graph. Consider a measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(G))$ that gives mass to the set $\mathcal{W}_2^{ac}(G)$ of absolutely continuous measures on (G, d) with respect to \mathcal{H} , namely $\mathbb{P}(\mathcal{W}_2^{ac}(G)) > 0$. If $\mu_{\mathbb{P}}$ is a barycenter of \mathbb{P} then the restriction of $\mu_{\mathbb{P}}$ to the interior of any given edge is absolutely continuous. Therefore, if $\mu_{\mathbb{P}}$ is not absolutely continuous then its singular part is a sum of Dirac measures located at vertices of (G, d) .*

Remark 1.2. A similar result for metric trees was obtained earlier by the second author in his PhD [24].

One can ask whether the barycenter of \mathbb{P} is, to some extent, unique. This is not the case in general.

Proposition 1.3. *In general, given a metric graph (G, d) and $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(G))$, neither the atomic part nor the absolutely continuous part of a Wasserstein barycenter $\mu_{\mathbb{P}}$ is unique.*

Organisation of the paper: In Section 2, we first define the metric graphs to be considered in this paper. We also introduce the 1d-Hausdorff measure on such a graph as the canonical reference measure. We briefly recall and/or prove the main properties of metric graph. In the second part, we recall classical properties of Wasserstein spaces, putting the emphasis on the measurable selection properties available on this space. Finally, we describe extra properties in the case of the Wasserstein space over the real line. In this very specific situation, the Wasserstein space $\mathcal{W}_2(\mathbb{R})$ is *flat*, namely isometric to a convex subset of a Hilbert space. Through this isometry, we prove the folklore result that the barycenter of a measure is the standard average of the measures with respect to the given probability measure \mathbb{P} . We refer to Theorem 2.26 for a statement.

In Section 3, taking advantage of the tools introduced in Section 2, we first prove the restriction property for barycenters. Generalising the fact that the restriction of an optimal plan remains optimal, Proposition 3.4 is given in a broad set-up since we believe it can be useful in other contexts. This statement will allow us to reduce the proof of our main theorem to the “mesoscopic” scale where a Wasserstein barycenter is assumed to be supported in a edge. In the last part, we also

define our branched covering map and then prove its main properties that will allow us to somehow transfer the Wasserstein barycenter problem over (G, d) to one such problem over the real line. It is to be emphasized that the presence of cycles in (G, d) does not prevent the branched covering map to have these powerful features.

Finally, in Section 4.2 we prove our main results. The main theorem proof is in two steps as mentioned above. First, we prove the result in the case when the Wasserstein barycenter is supported in an edge, then building on the restriction property, we prove our main result holds without the restriction on the support. In the last paragraph, we discuss the non-uniqueness of the Wasserstein barycenter on a metric graph.

2 Preliminaries

2.1 Metric (measure) graphs

In this subsection, we present a constructive definition of a metric graph in terms of length functions defined on its edges. This construction induces a canonical measure on the metric graph, which coincides with the Lebesgue measure when restricted to each edge. A metric graph equipped with this canonical measure is referred to as a *metric measure graph*, a basic concept that underlies much of the subsequent development in this chapter.

Definition 2.1 (Graph). A *graph* is a pair (V, \mathcal{E}) where V denotes the set of *vertices* and $\mathcal{E} \subset \{\{x, y\}; x, y \in V, x \neq y\}$ the set of (non-oriented) edges. We assume that V is at most countable, hence so is \mathcal{E} . According to this definition, there is at most *one* edge between two given distinct vertices and *no* edge between a single vertex (such an edge is often called loop in graph theory). We shall use the notation e or $\{x, y\}$ (with $x, y \in V, x \neq y$) for an edge. When $e = \{x, y\} \in \mathcal{E}$, x and y are the *endpoints* of e while e is an edge at x (or y). The *degree* of a vertex is the number of edges at the vertex.

The set of *oriented edges* is denoted by $E \subset (V \times V) \setminus \{(v, v), v \in V\}$, an oriented edge is denoted either by \vec{e} or (x, y) . A *path* p in a graph is a finite sequence $p = g_0 \cdots g_k$ where $k \geq 1$ and for each $i \in \{0, \dots, k-1\}$, $\{g_i, g_{i+1}\} \in \mathcal{E}$; an edge is a particular path and we call g_0 and g_k the *ends* of the path p . A cycle is a path whose endpoints coincide. A graph is said to be *connected* if there exists a path between any pair of distinct vertices.

We shall assume that the degree at each vertex is finite and at least 1.

Such a graph is said to be locally finite.

We now introduce metric features on a graph through the notion of length structure on the set of edges.

Definition 2.2 (Metric graph). A *length function* on a graph (V, \mathcal{E}) is a function $l : \mathcal{E} \rightarrow [0, \infty)$ uniformly bounded from below by a strictly positive number, i.e.

$$c := \inf_{\alpha \in \mathcal{E}} l(\alpha) > 0. \tag{2}$$

We identify an oriented edge $\vec{e} \in E$ with the interval $]0, l(e)[= \{x \in \mathbb{R}; 0 < x < l(e)\}$ ². When $(\vec{e}_0, \vec{e}_1) = \vec{e} \in E$, the vertices \vec{e}_0 and \vec{e}_1 are identified with the points 0 and $l(e)$ of $[0, l(e)]$.

²We use french notation to avoid confusion with an element of \mathbb{R}^2 .

We get a *metric graph* by considering the quotient space:

$$G := \left(\prod_{\vec{e} \in E} [0, l(e)] \right) / \sim,$$

where points in G corresponding to the same vertex of (V, \mathcal{E}) are identified.

Last, we define $\mathring{e} :=]0, l(e)[$ the interior of the edge e .

Our next goal is to introduce a (length) metric on G as follows: we first define the length of a path $p = g_0 \cdots g_k$ as

$$l(p) = \sum_{i=0}^{k-1} l(\{g_i, g_{i+1}\}),$$

so that this length coincides with the length function introduced above for paths which are edges (*i.e.* when $k = 1$). Note also the bound

$$l(p) \geq k c > 0.$$

According to [12, §3.2.2], we need to define the *length* of a path in G with arbitrary endpoints. Let $x \neq y \in G$ and assume that $x \notin V$ then $x \in]0, l(e)[$ for a unique oriented edge \vec{e} . Similarly, if $y \notin V$ then $y \in]0, l(f)[$ for a unique $f \in E$. For $x, y \in e$ (equivalently $e = f$), we define a path p in e from x to y as the affine path $t \mapsto (1-t)x + ty$ and its length is $l(p) = |x - y|$. In the remaining cases, a path from x to y is a finite sequence $p = x\vec{e}_i g_1 \cdots g_{k-1} \vec{f}_j y$ where $i, j \in \{0, 1\}$, $\vec{e}_i g_1 \cdots g_{k-1} \vec{f}_j$ is a path between two vertices of (V, \mathcal{E}) , and

$$l(p) = |i \times l(e) - x| + \sum_{i=0}^{k-1} l(\{g_i, g_{i+1}\}) + |j \times l(f) - y|,$$

where $g_0 = \vec{e}_i$ and $g_k = \vec{f}_j$. For short, we write

$$p : x \curvearrowright y$$

for a path from x to y .

Note that the length function coincides with the previous definition when $x = \vec{e}_i$ and $y = \vec{f}_j$.

Definition 2.3 (length distance on G). The distance $d : G \times G \rightarrow [0, \infty]$ is defined by

$$d(x, y) = \begin{cases} 0 & \text{if } x = y \\ \infty & \text{if there is no path from } x \text{ to } y \\ \inf_{p: x \curvearrowright y} l(p) & \text{otherwise} \end{cases} .$$

Proposition 2.4. *Let G be a connected, locally finite metric graph. Then the function d turns G into a metric space.*

We refer to [12, Chapter 2] or [24] for details. We only mention that the metric graph is originally equipped with the quotient topology, the triangle inequality follows the additivity of the length function applied to the concatenation of two paths while the positivity of d is a consequence of (2).

In what follows, we always assume the graph to be connected and locally finite, and we simply denote the metric graph by (G, d) .

Basic metric properties

We first fix notation for general metric space (X, d) .

Definition 2.5 (Various properties of metric spaces). A metric space (X, d) is said to be

- *Polish* if it is complete and separable,
- *proper* if any closed ball in (X, d) is compact, the closed ball of radius R centered at $x \in X$ is denoted by $\overline{B}(x, R)$,
- *geodesic* if for any pair of distinct points $x \neq y \in X$, there exists a *geodesic* σ from x to y . A geodesic σ in X is a map $\sigma : [0, 1] \rightarrow X$ such that there exists a number $l(\sigma) \geq 0$ called its *length*, and, for any $s \leq t \in [0, 1]$,

$$d(\sigma(s), \sigma(t)) = l(\sigma)(t - s).$$

In particular $d(\sigma(0), \sigma(1)) = l(\sigma)$.³

We first recall that a metric graph is a particular instance of (proper) Polish space. This feature makes available most of the classical tools from the optimal mass transport theory.

Theorem 2.6. *Metric graphs are proper, complete, and geodesic metric spaces.*

Proof. Let us show that each closed ball $\overline{B}(v, R)$ is compact. Indeed according to (2), a path starting at v whose length is at most, say, $R + 1$, is contained in the union of at most $(R + 1)/c$ edges. The degree of each vertex being finite, only finitely many vertices belong to that ball. Finally, $\overline{B}(v, R)$ is contained in (the image of) finitely many segments $[0, l(e_i)]$ for $i \in \{1, \dots, N\}$, thus it is compact. The rest of the proof is classical, we refer to [12] and the references therein for more details. \square

Our next goal is to define a canonical measure on a metric graph. Before that, we emphasize some features of the distance on G when restricted to an edge.

Lemma 2.7. *Given (G, d) a metric graph, let $\vec{e} \in E$ and $x, y \in [0, l(e)]$. Then,*

- $d(x, y) \leq |x - y|$,
- $d(x, y) = |x - y|$ if x and y are close enough,
- if $d(\vec{e}_0, \vec{e}_1) = l(e)$ the edge e is said to be *minimising*. When it is so, for any $s, t \in [0, l(e)]$,

$$d(s, t) = |s - t|.$$

Proof. The proof of the first inequality follows from the definition since the path from x to y contained in e is the affine path of length $|x - y|$. For x, y close enough, *i.e.* such that $|x - y| < c$ (where c is as in (2)), any other path σ from x to y has to leave e and go through another edge of length at least c before reentering e , such a path cannot be of minimal length. Note that without loss of generality, we can assume that σ reenters \vec{e} through the other end, otherwise it is clearly not a path of minimal length. If we further assume \vec{e} is minimising then any path σ from x to y as above can be restricted to a subinterval for which its endpoints are \vec{e}_0 and \vec{e}_1 . Thus the length of σ is greater than $d(\vec{e}_0, \vec{e}_1) = l(e) > |x - y|$ whenever $\{x, y\} \neq \{0, l(e)\}$. \square

³a geodesic according to this definition is called “minimising geodesic” in Riemannian geometry.

The canonical measure \mathcal{H} on (G, d) can be defined using its edges and vertices, it is indeed the one-dimensional Hausdorff measure on (G, d) [12, §1.7].

Definition 2.8 (Canonical measures on metric graphs). Let $G = (V, E, d)$ be a metric graph. The canonical measure \mathcal{H} is the 1d-Hausdorff measure induced by the distance d on G . Thanks to Lemma 2.7, the measure \mathcal{H} gives no mass to the set of vertices V , and for each oriented edge $\vec{e} \in E$, the restriction $\mathcal{H} \llcorner_{\vec{e}}$ to \vec{e} is the Lebesgue measure on $[0, l(\vec{e})]$.

Proof. By Lemma 2.7, for any $\vec{e} \in E$, any $x \leq y \in [0, l(\vec{e})]$, $\mathcal{H}^1([x, y]) \leq |x - y|$ and equality holds whenever $d(x, y) = |x - y|$. Since the index at each vertex is finite, this implies $\mathcal{H}(\{v\}) = 0$ for any $v \in V$. Last, since the 1d-Hausdorff measure with respect to $|\cdot|$ on $[0, l(\vec{e})]$ coincides with the Lebesgue measure (up to scaling) [16], the second item of Lemma 2.7 together with simple covering arguments taking advantage of the compactness of $[0, l(\vec{e})]$, lead to the fact that $\mathcal{H} \llcorner_{\vec{e}}$ coincides with the Lebesgue measure on $[0, l(\vec{e})]$. \square

In the course of proving our main result, we shall need to single out some special points which are, in some weak sense, in the Cut Locus of the graph. We define the exact meaning of this in the next definition.

Definition 2.9. Let (G, d) be a metric graph. For $x \neq y \in G, x \notin V$, there exist $e, f \in E$ such that $x \in \mathring{e}$ and $y \in f$.

If there are at least two distinct geodesics γ_1, γ_2 between x and y such that γ_1 and γ_2 exit e by different endpoints, the pair (x, y) is said to belong to the *oriented* cut locus $\overrightarrow{Cut}(G)$. In particular this forces $e \neq f$.

With that definition in hand, one can easily generalise part of a classical fact in Riemannian geometry to our metric graph set-up.

Lemma 2.10. Let (G, d) be a metric graph and $x \neq y$ such that $(x, y) \in (G \setminus V \times G) \setminus \overrightarrow{Cut}(G)$. Then, any geodesic γ from y to x can be locally extended to a geodesic beyond x .

Proof. The proof is by contradiction. Assume γ cannot be extended beyond $x \in \mathring{e}$. Identifying e with the segment $[0, l(\vec{e})]$, so that $x \in]0, l(\vec{e})[$, assuming that γ reaches x by going through \vec{e}_0 ; one gets for any $\varepsilon > 0$ small, the existence of geodesic from y to $x + \varepsilon$ whose length is smaller than $d(x, y) + \varepsilon$, this geodesic has to reach $x + \varepsilon$ through \vec{e}_1 . The lower semicontinuity of the length functional combined with an easy extraction argument easily provides us with another geodesic between x and y entering \vec{e} through \vec{e}_1 hence a contradiction. \square

Lemma 2.11. Let (G, d) be a metric graph and $v \in V$. Then the set

$$\overrightarrow{Cut}_v(G) := \{y \in G; (y, v) \in \overrightarrow{Cut}(G)\}$$

is discrete, in particular it is at most countable.

Proof. Let $y \in \overrightarrow{Cut}_v(G)$. Thus, there are two geodesics γ_1 and γ_2 connecting v and y crossing different endpoints of e . As a consequence, the cardinality of $\overrightarrow{Cut}_v(G) \cap \mathring{e}$ is at most 1. Therefore, since the number of edges is at most countable and the degree of each vertex is finite, we infer the claim. \square

2.2 Wasserstein spaces

In this paper we consider the quadratic Wasserstein space over a metric space (X, d) , namely the space of Borel probability measures μ with finite second order moment:

$$\int_X d^2(x_0, x) \, d\mu(x),$$

where x_0 is an arbitrary fixed point of X .

The set of such probability measures is denoted by $\mathcal{W}_2(X)$. The space $\mathcal{W}_2(X)$ is equipped with the Wasserstein distance whose definition is recalled below. To define this metric, let us introduce for $\mu_1, \mu_2 \in \mathcal{W}_2(X)$,

$$\Gamma(\mu_1, \mu_2) := \{\pi \in \mathcal{P}(X \times X); p_{i\sharp}\pi = \mu_i, i \in \{1, 2\}\}$$

the set of transport plans between μ_1 and μ_2 . In this definition, the maps $p_i : X \times X \rightarrow X$ are the canonical projections defined by $p_1(x, y) = x$ and $p_2(x, y) = y$.

The Wasserstein distance between two measures $\mu, \nu \in \mathcal{W}_2(X)$ is defined as

$$d_W(\mu, \nu) := \left(\min_{\pi \in \Gamma(\mu, \nu)} \int_{X \times X} d^2(x, y) \, d\pi(x, y) \right)^{1/2}.$$

The subset of *optimal* plans between μ and ν is denoted by $\Gamma^o(\mu, \nu)$; it is made of plans $\pi^o \in \Gamma(\mu, \nu)$ for which

$$d_W(\mu, \nu)^2 = \int_{X \times X} d^2(x, y) \, d\pi^o(x, y).$$

Let us recall three classical results on Wasserstein spaces:

Theorem 2.12. *Let (X, d) be a metric space. Then,*

1. $(\mathcal{W}_2(X), d_W)$ is Polish whenever (X, d) is so,
2. $(\mathcal{W}_2(X), d_W)$ is compact whenever (X, d) is so, otherwise $(\mathcal{W}_2(X), d_W)$ is not proper,
3. $(\mathcal{W}_2(X), d_W)$ is Polish and geodesic⁴ whenever (X, d) is so.

We refer to [34, 3, 38] for proofs and a complete review of the standard topological and metric properties of the quadratic Wasserstein space.

Remark 2.13. Combining Theorem 2.6 together with the previous theorem yields that in the case of a metric graph (G, d) , $(\mathcal{W}_2(G), d_W)$ is a geodesic Polish space, non-proper except if (G, d) is compact (then it is actually compact).

In the next part, we recall some results on measurable selection that are often useful to compare Wasserstein spaces over distinct base spaces.

⁴see Definition 2.5 for the definition

2.2.1 Measurable selection

In order to study the Wasserstein barycenter, we shall need to apply several *measurable* selections of family of measures -like transport plans with a variable marginal for instance.

In the result below, the product space X^2 is equipped with the product metric

$$d((x_1, x_2), (y_1, y_2))^2 = d(x_1, y_1)^2 + d(x_2, y_2)^2.$$

Theorem 2.14. *Let (X, d) be a Polish space. There exists a measurable map*

$$\begin{aligned} \pi : \mathcal{W}_2(X) \times \mathcal{W}_2(X) &\longrightarrow \mathcal{W}_2(X^2), \\ (\mu, \nu) &\longmapsto \pi_\mu^\nu \end{aligned} \quad (3)$$

such that $\pi_\mu^\nu \in \Gamma^o(\mu, \nu)$ for all μ, ν .

We refer to [38, Corollary 5.22] for a proof of this result. Regarding the base space, we also have the following selection result.

Proposition 2.15. *Let (X, d) be a proper geodesic Polish metric space and let $\mathcal{G}eo(X)$ be the space of geodesics in (X, d) . There exists a measurable selection map:*

$$\begin{aligned} \gamma_{\cdot, \cdot} : X \times X &\longrightarrow \mathcal{G}eo(X), \\ (x, y) &\longmapsto \gamma_{x, y} \end{aligned} \quad (4)$$

where $\gamma_{x, y}(0) = x$ and $\gamma_{x, y}(1) = y$.

The proof of this result is a rather straightforward consequence of the Arzela-Ascoli theorem. We refer to [23] for a proof of it in the general setting of Polish spaces.

Definition 2.16 (Evaluation map). For $t \in [0, 1]$, we define

$$\begin{aligned} \cdot_t : \mathcal{G}eo(X) &\longrightarrow X \\ \gamma &\longmapsto \gamma_t := \gamma(t) \end{aligned} \quad (5)$$

By combining the two previous results, we get the existence of optimal *dynamical* plans. A measurable selection of such plans does exist.

Theorem 2.17. *Let (X, d) be a proper geodesic Polish space. There exists a measurable map*

$$\begin{aligned} \Pi : \mathcal{W}_2(X) \times \mathcal{W}_2(X) &\longrightarrow \mathcal{P}(\mathcal{G}eo(X)), \\ (\mu, \nu) &\longmapsto \Pi_\mu^\nu \end{aligned} \quad (6)$$

such that

$(\cdot_0, \cdot_1)_\# \Pi_\mu^\nu \in \Gamma^o(\mu, \nu)$ for all μ, ν . Besides, the following equality holds

$$d_W(\mu, \nu)^2 = \int_{\mathcal{G}eo(X)} s^2(\gamma) \, d\Pi_\mu^\nu(\gamma), \quad (7)$$

where $s(\gamma) = d(\gamma(0), \gamma(1))$ is the speed (or length) of γ .

Proof. The set $\mathcal{Geo}(X)$ equipped with the uniform distance

$$d_\infty(\gamma, \tilde{\gamma}) = \max_{t \in [0,1]} d(\gamma(t), \tilde{\gamma}(t)),$$

is a σ -compact metric space hence complete and separable. We then denote the quadratic Wasserstein space over $(\mathcal{Geo}(X), d_\infty)$ by $(\mathcal{W}_2(\mathcal{Geo}(X)), d_W)$. Observe that for a probability measure Π over $\mathcal{Geo}(X)$, $\Pi \in \mathcal{W}_2(\mathcal{Geo}(X))$ amounts to

$$\int s^2(\gamma) d\Pi(\gamma) < \infty,$$

thanks to the bound, say for a constant geodesic $\gamma_0 \equiv x_0$,

$$d_\infty^2(\gamma_0, \gamma) \leq 2(d^2(x_0, \gamma(0)) + s^2(\gamma)).$$

The proof then follows from the Kuratowski and Ryll-Nardzewski measurable selection theorem [11]. We consider the set-valued mapping:

$$F : \mathcal{W}_2(X) \times \mathcal{W}_2(X) \rightrightarrows \mathcal{W}_2(\mathcal{Geo}(X)),$$

where $F(\mu, \nu) = \Gamma_{dyn}^o(\mu, \nu)$ is the set of optimal dynamical plans whose ends are μ and ν (in this order). We have to check that $F(\mu, \nu)$ is a non-empty closed subset of $\mathcal{W}_2(\mathcal{Geo}(X))$ (we shall actually prove the *compactness* of $F(\mu, \nu)$) and that for any closed subset K of $\mathcal{W}_2(\mathcal{Geo}(X))$, the set

$$\mathcal{S}_K := \{(\mu, \nu) \in \mathcal{W}_2(X)^2; F(\mu, \nu) \cap K \neq \emptyset\}$$

is a measurable subset of $\mathcal{W}_2(X)^2$ ⁵.

We have already checked the existence of optimal dynamical plans between μ and ν . The compactness of $F(\mu, \nu)$ is a consequence of Prokhorov's theorem. Indeed, given $R > 0$ such that $\max(\mu(X \setminus \overline{B}(x_0, R)), \nu(X \setminus \overline{B}(x_0, R))) < \varepsilon$ then we claim that the set

$$\mathcal{G}_R := \{\gamma \in \mathcal{Geo}(X); \gamma(0), \gamma(1) \in \overline{B}(x_0, R)\}$$

is compact. First, the set \mathcal{G}_R is clearly closed.

To prove the claim, note that for $\gamma \in \mathcal{G}_R$, $s(\gamma) = d(\gamma(0), \gamma(1)) \leq 2R$ thus $d(x_0, \gamma(t)) \leq R + 1/2s(\gamma) \leq 2R$ for all $t \in [0, 1]$. Consequently, \mathcal{G}_R is contained in the subset \mathcal{S} of Lipschitz functions $f : [0, 1] \rightarrow X$ whose range satisfies $f([0, 1]) \subset \overline{B}(x_0, 2R)$ and whose Lipschitz constant is at most $2R$. The space (\mathcal{S}, d_∞) is compact thanks to the Ascoli-Arzelà theorem.

Now, let us prove that \mathcal{S}_K is a closed subset of $\mathcal{W}_2(X)^2$. Let

$$(\mu_n, \nu_n) \xrightarrow{n \rightarrow \infty} (\mu_\infty, \nu_\infty)$$

be a converging sequence where the (μ_n, ν_n) belong to \mathcal{S}_K . For each n , let $\Pi_n \in F(\mu_n, \nu_n) \cap K$. Since $\mu_n \rightarrow \mu_\infty$ and $\nu_n \rightarrow \nu_\infty$ in $\mathcal{W}_2(X)$, the collection $\{(\mu_n, \nu_n); n \in \mathbb{N}\} \cup \{(\mu_\infty, \nu_\infty)\}$ is a compact set and the proof of the claim above leads to the compactness of $\bigcup_{n \in \mathbb{N}} F(\mu_n, \nu_n)$ (this is classical for optimal plans). Finally, the continuity of the evaluation mapping $\Pi \mapsto (e_0, e_1)_\# \Pi$ implies that any accumulation point of $(\Pi_n)_{n \in \mathbb{N}}$ belongs to $F(\mu_\infty, \nu_\infty) \cap K$ and the result is proven. \square

In the case of a metric graph, we shall take advantage of the graph structure in order to further study dynamical plans. This will be done in Section 3.

⁵the standard statement of this measurable theorem requires to check this property for K an open subset of $\mathcal{W}_2(X)^2$, this property follows from the one we state by elementary calculations noting that any open subset of a metric space is a countable union of closed sets

2.2.2 Barycenter

We end this introductory part on Wasserstein spaces with our main object of interest: Wasserstein barycenter. We start with the definition of a barycenter on an arbitrary metric space before briefly discussing the basic properties of barycenters over Wasserstein space.

Definition 2.18 (Barycenter). Let (X, d) be a metric space and let μ be a Borel probability measure on X such that $\int_X d(x_0, y)^2 d\mu(y) < \infty$ for some point $x_0 \in X$. We call $z_\mu \in X$ a barycenter of μ if

$$\int_X d(z_\mu, y)^2 d\mu(y) = \min_{x \in X} \int_X d(x, y)^2 d\mu(y).$$

Barycenters always exist in proper spaces since a minimizing sequence is bounded and thus pre-compact. We refer to Ohta [29] for more details and some other properties of barycenters in a proper space.

However, simple counter-examples show the lack of uniqueness for barycenters.

Example 2.19. Let \mathbb{S}^2 be the two-dimensional sphere and fix two antipodal points $x, -x$ of \mathbb{S}^2 . Consider the measure $\mu = \frac{1}{2}\delta_x + \frac{1}{2}\delta_{-x}$ on \mathbb{S}^2 . Then all points in the equator, i.e., the set of all points with equal distances to x and $-x$, are barycenters of μ .

Barycenters on non-proper metric spaces may not exist, see [24] for counter-examples. Recall that $(\mathcal{W}_2(X), d_W)$ is *not* proper unless (X, d) is compact. The existence of Wasserstein barycenter is therefore not straightforward. Nevertheless, a general existence result is proved in [22].

Theorem 2.20. *If (X, d) is a locally compact, complete, geodesic space then any measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(X))$ has a barycenter.*

We shall need finer properties of the Wasserstein space when the base space is $(\mathbb{R}, |\cdot|)$ that are explained in the next paragraph.

2.3 Wasserstein space over the real line

Most results in this section are based on the fact that $\mathcal{W}_2(\mathbb{R})$ is a flat space, a precise statement is Theorem 2.24. The main object to express the extra properties of $\mathcal{W}_2(\mathbb{R})$ is the quantile function of $\mu \in \mathcal{W}_2(\mathbb{R})$ which is discussed in the next part.

2.3.1 Quantile functions

In the literature, there exist different definitions of distribution functions and quantile functions, we thus make our definition explicit.

Definition 2.21 (Distribution functions and quantile functions). Let μ be a probability measure on \mathbb{R} . Its distribution function $f_\mu : \mathbb{R} \rightarrow [0, 1]$ is defined by $f_\mu(x) := \mu((-\infty, x])$, and its quantile function $f_\mu^{-1} : [0, 1] \rightarrow \overline{\mathbb{R}}$ is defined by

$$f_\mu^{-1}(t) := \inf_x \{x \in \mathbb{R}; f_\mu(x) > t\} \text{ for } 0 < t < 1$$

and $f_\mu^{-1}(0) := \lim_{t \downarrow 0} f_\mu^{-1}(t), \quad f_\mu^{-1}(1) := \lim_{t \uparrow 1} f_\mu^{-1}(t),$

where $\overline{\mathbb{R}} = [-\infty, \infty]$.

With this definition, the quantile function is fully characterised by its values on $(0, 1)$. While the values at the boundary of $[0, 1]$ can be infinite, the quantile function is real-valued as a function on $(0, 1)$. Moreover, as the distribution function f_μ , the quantile function f_μ^{-1} is a non-decreasing, right continuous function on $(0, 1)$. Besides, any $g : (0, 1) \rightarrow \mathbb{R}$ non-decreasing, right continuous function is the quantile function of some probability $\mu \in \mathcal{P}(\mathbb{R})$. We refer to [10, 37] and the references therein for more on this properties.

In the next lemma, to be used in the proof of our main result, the size of the support of μ is related to the range of its quantile function.

Lemma 2.22. *Let μ be a probability measure on \mathbb{R} . The infimum and supremum of the support of μ are related to its quantile function as follows,*

$$f_\mu^{-1}(0) = \inf \text{supp}(\mu) \quad \text{and} \quad f_\mu^{-1}(1) = \sup \text{supp}(\mu).$$

In particular, μ has compact support if and only if f_μ^{-1} is finite on the whole unit interval $[0, 1]$.

Proof. We first prove the following two inequalities,

$$f_\mu^{-1}(0) \leq \inf \text{supp}(\mu) \quad \text{and} \quad f_\mu^{-1}(1) \geq \sup \text{supp}(\mu).$$

The case that $f_\mu^{-1}(0) = -\infty$ or $f_\mu^{-1}(1) = +\infty$ is trivial, we are left to consider the case where they are finite. If $y < f_\mu^{-1}(0)$, then for any $t \in (0, 1)$, $\frac{1}{2}y + \frac{1}{2}f_\mu^{-1}(0) < f_\mu^{-1}(t)$ and thus $\frac{1}{2}y + \frac{1}{2}f_\mu^{-1}(0) \notin \{x \in \mathbb{R} \mid f_\mu(x) > t\}$. It follows that $f_\mu(\frac{1}{2}y + \frac{1}{2}f_\mu^{-1}(0)) = 0$, and hence the point y , strictly smaller than $\frac{1}{2}y + \frac{1}{2}f_\mu^{-1}(0)$, is not in the support of μ . As y is arbitrarily chosen, we have $(-\infty, f_\mu^{-1}(0)) \cap \text{supp}(\mu) = \emptyset$ and thus $f_\mu^{-1}(0) \leq \inf \text{supp}(\mu)$. If $z > f_\mu^{-1}(1)$, then for any $t \in (0, 1)$, $\frac{1}{2}z + \frac{1}{2}f_\mu^{-1}(1) > f_\mu^{-1}(t)$ and thus $\frac{1}{2}z + \frac{1}{2}f_\mu^{-1}(1) \in \{x \in \mathbb{R} \mid f_\mu(x) > t\}$. It follows that $f_\mu(\frac{1}{2}z + \frac{1}{2}f_\mu^{-1}(1)) = 1$, and hence the point z , strictly bigger than $\frac{1}{2}z + \frac{1}{2}f_\mu^{-1}(1)$, is not in the support of μ . As z is arbitrarily chosen, we have $(f_\mu^{-1}(1), +\infty) \cap \text{supp}(\mu) = \emptyset$ and thus $f_\mu^{-1}(1) \geq \sup \text{supp}(\mu)$.

We now prove the inequalities,

$$f_\mu^{-1}(0) \geq \inf \text{supp}(\mu) \quad \text{and} \quad f_\mu^{-1}(1) \leq \sup \text{supp}(\mu).$$

The case that $\inf \text{supp}(\mu) = -\infty$ or $\sup \text{supp}(\mu) = +\infty$ is trivial, we are left to consider the case where they are finite. If $y < \inf \text{supp}(\mu)$, then $f_\mu(y) = 0$ and thus $y \leq f_\mu^{-1}(t)$ for all $t \in (0, 1)$, which implies $y \leq f_\mu^{-1}(0)$. As y is arbitrarily chosen, we must have $f_\mu^{-1}(0) \geq \inf \text{supp}(\mu)$. If $z > \sup \text{supp}(\mu)$, then $f_\mu(z) = 1$ and thus $z \geq f_\mu^{-1}(t)$ for all $t \in (0, 1)$, which implies $z \geq f_\mu^{-1}(1)$. As z is arbitrarily chosen, we must have $f_\mu^{-1}(1) \leq \sup \text{supp}(\mu)$.

Since μ has compact support if and only if both $\inf \text{supp}(\mu)$ and $\sup \text{supp}(\mu)$ are finite, our last statement in the lemma follows. \square

In Subsection 2.3.2, we describe the barycenter $\mu_{\mathbb{P}}$ of a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$ as the integral of the map $\mu \mapsto f_\mu^{-1}(t)$ with respect to \mathbb{P} . The goal of the next lemma is to make sure that this map is measurable.

It is a classical result that weak convergence of probability measures $\mu_n \rightarrow \mu_\infty$ can be equivalently characterised as the pointwise convergence of the corresponding distribution functions $f_{\mu_n}(t)$ towards $f_{\mu_\infty}(t)$ at every continuity point t of f_{μ_∞} . The same characterisation holds when quantile functions are used instead, also at continuity points of $f_{\mu_\infty}^{-1}$; we refer to [13, Proposition 5.7 of Chapter III] for a proof. The discontinuity points of a nondecreasing function $g : (0, 1) \rightarrow \mathbb{R}$ being at most countable, one infer the following property:

Lemma 2.23. *Let $\{\mu_n\}_{n \in \mathbb{N}}$ be a sequence of probability measures on \mathbb{R} converging weakly to μ_∞ . Then for any $t \in (0, 1)$, it holds*

$$\limsup_{n \rightarrow +\infty} f_{\mu_n}^{-1}(t) \leq f_{\mu_\infty}^{-1}(t), \quad \forall t \in [0, 1] \quad \text{and} \quad \liminf_{n \rightarrow +\infty} f_{\mu_n}^{-1}(1) \geq f_{\mu_\infty}^{-1}(1). \quad (8)$$

Proof. For $t \in [0, 1)$, there is sequence of decreasing and positive numbers $\{\varepsilon_k\}_{k \in \mathbb{N}}$ such that $t + \varepsilon_k \in (0, 1)$ and f_μ^{-1} is continuous at $t + \varepsilon_k$. Since quantile functions are non-decreasing, we have

$$\limsup_{n \rightarrow +\infty} f_{\mu_n}^{-1}(t) \leq \limsup_{n \rightarrow +\infty} f_{\mu_n}^{-1}(t + \varepsilon_k) = \lim_{n \rightarrow +\infty} f_{\mu_n}^{-1}(t + \varepsilon_k) = f_\mu^{-1}(t + \varepsilon_k),$$

which implies $\limsup_{n \rightarrow +\infty} f_{\mu_n}^{-1}(t) \leq f_\mu^{-1}(t)$ by the right-continuity of f_μ^{-1} at t . The proof for $t = 1$ is similar and left to the reader. \square

We end this part by showing that $\mathcal{W}_2(\mathbb{R})$ is isometric to a closed convex subset of a Hilbert space.

In what follows, $[g]$ stands for the equivalence class in $L^2([0, 1])$ of a function $g : (0, 1) \rightarrow \mathbb{R}$. Let us define

$$Q := \{[g] \in L^2([0, 1]); g : (0, 1) \rightarrow \mathbb{R} \text{ nondecreasing, right continuous } \}.$$

Theorem 2.24. *The map*

$$\begin{aligned} \mathcal{W}_2(\mathbb{R}) &\longrightarrow Q \subset L^2([0, 1]) \\ \mu &\longmapsto f_\mu^{-1} \end{aligned}$$

is a surjective isometry onto the closed convex subset Q of $L^2([0, 1])$:

$$d_W(\mu, \nu)^2 = \int_0^1 [f_\mu^{-1}(t) - f_\nu^{-1}(t)]^2 dt. \quad (9)$$

We refer to [37, Theorem 2.18] and [34] for a proof.

Remark 2.25. Note that if $\nu = \delta_0$ then $f_\nu^{-1}(t) \equiv 0$ on $(0, 1)$. Therefore, for such a choice (9) reads

$$\int_{\mathbb{R}} |x|^2 d\mu(x) = \int_0^1 [f_\mu^{-1}(t)]^2 dt.$$

2.3.2 Integral formula for Wasserstein barycenters on the real line

In this part, we take advantage of the linear structure on $L^2([0, 1])$ to prove a more standard expression of a Wasserstein barycenter on \mathbb{R} through its quantile function. In the case of a finitely supported probability measure $\mathbb{P} = \sum_{i=1}^n \lambda_i \delta_{\nu_i}$, the quantile function of its barycenter $\mu_{\mathbb{P}}$ is given by the formula [31, §3.1.4],

$$f_{\mu_{\mathbb{P}}}^{-1}(t) = \sum_{i=1}^n \lambda_i f_{\nu_i}^{-1}(t) = \int_{\mathcal{W}_2(\mathbb{R})} f_\nu^{-1}(t) d\mathbb{P}(\nu), \quad \forall t \in [0, 1].$$

For general $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$, we have

Theorem 2.26 (Wasserstein barycenters on the real line). *Let $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$ be a probability measure on the Wasserstein space $(\mathcal{W}_2(\mathbb{R}), d_W)$. Then \mathbb{P} has a unique Wasserstein barycenter $\mu_{\mathbb{P}} \in \mathcal{W}_2(\mathbb{R})$, whose quantile functions satisfies*

$$f_{\mu_{\mathbb{P}}}^{-1}(t) = \int_{\mathcal{W}_2(\mathbb{R})} f_{\nu}^{-1}(t) d\mathbb{P}(\nu), \quad \forall t \in [0, 1]. \quad (10)$$

In particular, the integral in (10) is finite for $t \in (0, 1)$.

This theorem was proved under stronger assumptions in [9].

Proof. By Lemma 2.23, the map $\nu \mapsto f_{\nu}^{-1}(t)$ is upper semi-continuous for $t \in [0, 1)$ and lower semi-continuous for $t = 1$, hence it is a measurable map for any parameter $t \in [0, 1]$. Besides, it follows from Theorem 2.24 and Fubini's theorem that

$$\int_0^1 \int_{\mathcal{W}_2(\mathbb{R})} [f_{\nu}^{-1}(t)]^2 d\mathbb{P}(\nu) dt = \int_{\mathcal{W}_2(\mathbb{R})} \int_0^1 [f_{\nu}^{-1}(t)]^2 dt d\mathbb{P}(\nu) = \int_{\mathcal{W}_2(\mathbb{R})} d_W(\delta_0, \nu)^2 d\mathbb{P}(\nu) < +\infty.$$

Consequently we infer from the Cauchy-Schwarz inequality that the function $g : [0, 1] \rightarrow \overline{\mathbb{R}}$ defined by $g(t) := \int_{\mathcal{W}_2(\mathbb{R})} f_{\nu}^{-1}(t) d\mathbb{P}(\nu)$ for $t \in [0, 1]$ is an element in $L^2([0, 1])$.

We claim that g is a non-decreasing and right-continuous function on $(0, 1)$ with $g(0) = \lim_{t \downarrow 0} g(t)$ and $g(1) = \lim_{t \uparrow 1} g(t)$. We first show that g must be finite on $(0, 1)$. Indeed, if $g(t) = +\infty$ for some $t \in (0, 1)$, then $g(s) = +\infty$ for any $s \in [t, 1]$ since any quantile function f_{ν}^{-1} is non-decreasing, which contradicts the fact that $g \in L^2([0, 1])$. Due to the same reason, we cannot have $g(t) = -\infty$ for some $t \in (0, 1)$. Hence, g is finite and non-decreasing on $(0, 1)$. Fix $t \in [0, 1)$, we show that g is right-continuous at t . Let $\{t_n\}_{n \geq 1} \subset (t, \frac{1+t}{2})$ be a decreasing sequence smaller than $\frac{1+t}{2}$ that converges to t . Applying the monotone convergence theorem with measure \mathbb{P} to the non-increasing sequence of non-positive functions $\nu \mapsto f_{\nu}^{-1}(t_n) - f_{\nu}^{-1}(\frac{1+t}{2})$, we obtain $\lim_{n \rightarrow \infty} g(t_n) - g(\frac{1+t}{2}) = g(t) - g(\frac{1+t}{2})$, which shows that g is right-continuous at t since the decreasing sequence $\{t_n\}_{n \geq 1}$ is arbitrarily chosen and g is non-decreasing on $(0, 1)$. By a similar application of the monotone convergence theorem, we see that $g(1) = \lim_{t \uparrow 1} g(t)$. Hence, our claim on g is proven.

Since $g|_{(0,1)}$ is non-decreasing and right-continuous, there exists a unique measure $\mu_{\mathbb{P}} \in \mathcal{W}_2(\mathbb{R})$ such that $f_{\mu_{\mathbb{P}}}^{-1} = g$. It remains to show that $\mu_{\mathbb{P}}$ is the unique barycenter of \mathbb{P} . For any measure $\eta \in \mathcal{W}_2(\mathbb{R})$, by Theorem 2.24 and Fubini's theorem, we have

$$\begin{aligned} \int_{\mathcal{W}_2(\mathbb{R})} d_W(\eta, \nu)^2 d\mathbb{P}(\nu) &= \int_0^1 \int_{\mathcal{W}_2(\mathbb{R})} [f_{\eta}^{-1}(t) - f_{\nu}^{-1}(t)]^2 d\mathbb{P}(\nu) dt \\ &= \int_0^1 \left[f_{\eta}^{-1}(t) - \int_{\mathcal{W}_2(\mathbb{R})} f_{\nu}^{-1}(t) d\mathbb{P}(\nu) \right]^2 dt + I(\mathbb{P}), \end{aligned} \quad (11)$$

where the abbreviated term $I(\mathbb{P})$ is independent of η :

$$I(\mathbb{P}) := \int_0^1 \int_{\mathcal{W}_2(\mathbb{R})} [f_{\nu}^{-1}(t)]^2 d\mathbb{P}(\nu) - \left(\int_{\mathcal{W}_2(\mathbb{R})} f_{\nu}^{-1}(t) d\mathbb{P}(\nu) \right)^2 dt.$$

It follows from (11) that the infimum $\inf_{\eta \in \mathcal{W}_2(\mathbb{R})} \int_{\mathcal{W}_2(\mathbb{R})} d_W(\eta, \nu)^2 d\mathbb{P}(\nu)$ is reached by η if and only if $f_{\eta}^{-1}(t) = \int_{\mathcal{W}_2(\mathbb{R})} f_{\nu}^{-1}(t) d\mathbb{P}(\nu) = g(t) = f_{\mu_{\mathbb{P}}}^{-1}(t)$ for almost every $t \in [0, 1]$. Therefore, $\mu_{\mathbb{P}}$ is a barycenter of \mathbb{P} and its uniqueness follows from the injectivity of the embedding in Theorem 2.24. \square

Let us also mention the following result which is a particular case of the main result in [25]:

Theorem 2.27. *Let $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$ be a probability measure on $(\mathcal{W}_2(\mathbb{R}), d_W)$. Assume that $\mathbb{P}(\mathcal{W}_2^{ac}(\mathbb{R})) > 0$ where $\mathcal{W}_2^{ac}(\mathbb{R})$ is the set of probability measures in $\mathcal{W}_2(\mathbb{R})$ that are absolutely continuous with respect to the Lebesgue measure on \mathbb{R} . Then the unique barycenter $\mu_{\mathbb{P}}$ also belongs to $\mathcal{W}_2^{ac}(\mathbb{R})$.*

3 Fine properties of Optimal Transport over a metric graph

3.1 Restriction property

It is a classical result that the restriction $\tilde{\pi}$ of an optimal plan π (namely $0 \neq \tilde{\pi} \leq \pi$) is still optimal, see for instance [38, Theorem 4.6] for a proof. This section aims at generalising this property to barycenters of measures, see Proposition 3.4. While we shall make use of this result in the case of a metric graph, the natural set-up for this property is that of Polish spaces. We believe this result can be useful in other contexts so we state and prove it at this level of generality.

Throughout this section, (X, d) is a *Polish space*. Let us recall the existence of a *measurable* mapping (see Theorem 2.14)

$$\begin{aligned} \pi : \mathcal{W}_2(X) \times \mathcal{W}_2(X) &\longrightarrow \mathcal{W}_2(X \times X), \\ (\mu, \nu) &\longmapsto \pi_{\mu}^{\nu} \end{aligned}$$

such that $\pi_{\mu}^{\nu} \in \Gamma^o(\mu, \nu)$ for all μ, ν .

We fix $\mu \in \mathcal{W}_2(X)$. Let us consider $0 \neq \tilde{\mu}^1 \not\leq \mu$ and $\mu^1 := \frac{1}{\tilde{\mu}^1(X)} \tilde{\mu}^1$ its associated probability measure. Observe that

- The measure $\tilde{\mu}^1$ induced a decomposition of μ into a convex combination of probability measures ($\lambda \in (0, 1)$):

$$\mu = \lambda \mu^1 + (1 - \lambda) \mu^2. \quad (12)$$

- There exist nonnegative and bounded measurable functions g_1 and g_2 such that

$$\mu^i = g_i \mu,$$

for $i \in \{1, 2\}$.

- For any $\nu \in \mathcal{W}_2(X)$, the map g_i induces an optimal plan

$$g_i \pi_{\mu}^{\nu} \in \Gamma^o(\mu^i, \nu^i),$$

where $\nu^i = p_{2\sharp}(g_i \pi_{\mu}^{\nu})$.

Remark 3.1. Note that $\lambda g_1 + (1 - \lambda) g_2 = 1$ μ -a.e. The last item is a particular instance of the inheritance of optimality through the restriction of optimal plans as recalled at the beginning of this section.

In order to state the restriction property for barycenters of measures, we first need to define

Definition 3.2 (Restriction mappings). For each $i \in \{1, 2\}$, we define the *restriction mapping* associated to (12) by

$$\begin{aligned} \mathcal{R}^i : \mathcal{W}_2(X) &\longrightarrow \mathcal{W}_2(X) \\ \nu &\longrightarrow \nu^i \end{aligned}$$

where $\nu^i = p_{2\#}(g_i \pi_\mu^\nu)$.

Let us prove

Lemma 3.3. *The restriction mappings \mathcal{R}^i , $i \in \{1, 2\}$, are measurable.*

Proof. Given $\pi_1, \pi_2 \in \mathcal{W}_2(X \times X)$, if $\sigma \in \mathcal{W}_2(X^2 \times X^2)$ is an optimal transport plan between π_1 and π_2 , then

$$\begin{aligned} d_W(\pi_1, \pi_2)^2 &= \int_{X^2 \times X^2} d(x_1, x_2)^2 + d(y_1, y_2)^2 \, d\sigma((x_1, y_1), (x_2, y_2)) \\ &\geq \int_{X \times X} d(y_1, y_2)^2 \, d[p_2 \times p_2]_{\#} \sigma(y_1, y_2) \geq d_W(p_{2\#}\pi_1, p_{2\#}\pi_2)^2, \end{aligned} \quad (13)$$

which implies the continuity of the push-forward map $p_{2\#}$.

Now recall that $\nu^i = p_{2\#}(g_i \pi_\mu^\nu)$. Therefore, we are done if we can prove the Borel measurability of $\pi \mapsto g_i \pi$ for $\pi \in \mathcal{W}_2(X \times X)$ such that $p_{1\#}\pi = \mu$. Let us first assume that $g_i \in C^b(X)$. We shall use the characterisation of the Wasserstein convergence on $\mathcal{W}_2(X \times X)$ as a weak convergence for continuous functions $\phi : X \times X \rightarrow \mathbb{R}$ such that

$$\exists C > 0, \exists (x_0, y_0) \in X^2; \forall (x, y) \in X^2 \quad |\phi(x, y)| \leq C(1 + d^2(x_0, x) + d^2(y_0, y)), \quad (14)$$

we refer to [38] for a proof of this characterisation.

Given a sequence $(\pi_n)_{n \in \mathbb{N}}$ such that $\pi_n \rightarrow \pi_\infty$ in $\mathcal{W}_2(X \times X)$, and ϕ a continuous function as in (14), noting that

$$|\phi(x, y)g_i(x)| \leq \|g\|_\infty C(1 + d^2(x_0, x) + d^2(y_0, y)),$$

we infer

$$\int_{X \times X} \phi g_i \, d\pi_n \longrightarrow \int_{X \times X} \phi g_i \, d\pi$$

when $n \rightarrow \infty$. This proves the continuity of the mapping $\pi \mapsto g_i \pi$ when $g_i \in C^b(X)$.

To prove the general case, namely when $g_i \geq 0$ belongs to $L^\infty(\mu)$, we make use of Lusin's theorem as proved in [11, Theorem 7.1.13]. Using that (X, d) is a Polish space, for any $\varepsilon > 0$, we get from this theorem the existence of a continuous function $f : X \rightarrow [0, \infty)$ such that $\|f\|_\infty \leq \|g_i\|_\infty$ and $\mu(\{x \in X; f(x) \neq g_i(x)\}) < \varepsilon$. For each positive integer n , let us set f_n a continuous function satisfying Lusin's statement for $\varepsilon = 1/n^2$.

Since a pointwise limit of measurable maps is measurable, we are left with proving that for any $\pi \in \mathcal{W}_2(X \times X)$ such that $p_{1\#}\pi = \mu$, $f_n \pi \rightarrow g_i \pi$ in $\mathcal{W}_2(X \times X)$ when $n \rightarrow \infty$.

To this aim, let ϕ be a continuous function as in (14); observe that

$$\begin{aligned} \left| \int_{X \times X} \phi f_n \, d\pi - \int_{X \times X} \phi g \, d\pi \right| &\leq \int_{X \times X} |\phi| |f_n - g| \, d\pi \\ &\leq C \int_{X \times X} (1 + d^2(x_0, x) + d^2(y_0, y)) |f_n(x) - g_i(x)| \, d\pi. \end{aligned}$$

By construction and the Borel-Cantelli lemma,

$$\lim_{n \rightarrow \infty} (1 + d^2(x_0, x) + d^2(y_0, y)) |f_n(x) - g_i(x)| = 0$$

for μ -a.e x and every y thus π -almost everywhere. For $\pi \in \mathcal{W}_2(X \times X)$, the function

$$(x, y) \mapsto (1 + d^2(x_0, x) + d^2(y_0, y))$$

belongs to $L^1(\pi)$ the Lebesgue dominated convergence theorem yields

$$\lim_{n \rightarrow \infty} \left| \int_{X \times X} \phi f_n \, d\pi - \int_{X \times X} \phi g \, d\pi \right| = 0$$

and the proof is complete. \square

The following proposition details our method to construct new Wasserstein barycenters by restricting an existing one.

Proposition 3.4 (Restriction property of Wasserstein barycenters). *Let (X, d) be a proper Polish space and $\mu \in \mathcal{W}_2(X)$ be a fixed probability measure on X .*

We consider a decomposition as in (12):

$$\mu = \lambda \mu^1 + (1 - \lambda) \mu^2$$

and $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(X))$.

Then, for $i \in \{1, 2\}$, if μ is a barycenter of \mathbb{P} then μ^i is a barycenter of $\mathbb{Q}^i := \mathcal{R}^i \# \mathbb{P}$, where \mathcal{R}^i is the corresponding restriction mapping.

Proof. Assume that $\mu = \mu_{\mathbb{P}}$ is a barycenter of \mathbb{P} and observe that

$$\begin{aligned} & \int_{\mathcal{W}_2(X)} d_W(\mu_{\mathbb{P}}, \nu)^2 \, d\mathbb{P}(\nu) = \int_{\mathcal{W}_2(X)} \int_{X \times X} d(x, y)^2 \, d\pi_{\mu}^{\nu}(x, y) \, d\mathbb{P}(\nu) \\ &= \int_{\mathcal{W}_2(X)} \int_{X \times X} d(x, y)^2 \left[\lambda \, d\pi_{\mu^1}^{\nu^1}(x, y) + (1 - \lambda) \, d\pi_{\mu^2}^{\nu^2}(x, y) \right] \, d\mathbb{P}(\nu) \\ &= \int_{\mathcal{W}_2(X)} \left[\lambda \, d_W(\mu^1, \nu^1)^2 + (1 - \lambda) \, d_W(\mu^2, \nu^2)^2 \right] \, d\mathbb{P}(\nu) \\ &\geq \lambda \min_{\eta \in \mathcal{W}_2(X)} \int_{\mathcal{W}_2(X)} d_W(\eta, \nu)^2 \, d\mathbb{Q}^1(\nu) + (1 - \lambda) \min_{\eta \in \mathcal{W}_2(X)} \int_{\mathcal{W}_2(X)} d_W(\eta, \nu)^2 \, d\mathbb{Q}^2(\nu), \end{aligned}$$

where we used the fact that $g_i \pi_{\mu}^{\nu}$ is an optimal transport plan between μ^i and ν^i . It follows that $\mathbb{Q}^i \in \mathcal{W}_2(\mathcal{W}_2(X))$ for $i = 1, 2$. We claim that the last inequality above must be an equality. Consider the measure $\bar{\mu} := \lambda \mu_{\mathbb{Q}^1} + (1 - \lambda) \mu_{\mathbb{Q}^2}$ with $\mu_{\mathbb{Q}^i}$ being a Wasserstein barycenter of \mathbb{Q}^i . According to the decomposition $\nu = \lambda \nu^1 + (1 - \lambda) \nu^2$, if θ_i ($i = 1, 2$) is a transport plan between $\mu_{\mathbb{Q}^i}$ and ν^i , then $\lambda \theta_1 + (1 - \lambda) \theta_2$ is a transport plan between $\bar{\mu}$ and ν . Hence, it follows from the definitions of $\mu_{\mathbb{Q}^i}$,

\mathbb{Q}^i , ν^i , and $d_W(\bar{\mu}, \nu)^2$ that

$$\begin{aligned}
& \lambda \min_{\eta \in \mathcal{W}_2(X)} \int_{\mathcal{W}_2(X)} d_W(\eta, \nu)^2 d\mathbb{Q}^1(\nu) + (1 - \lambda) \min_{\eta \in \mathcal{W}_2(X)} \int_{\mathcal{W}_2(X)} d_W(\eta, \nu)^2 d\mathbb{Q}^2(\nu) \\
&= \lambda \int_{\mathcal{W}_2(X)} d_W(\mu_{\mathbb{Q}^1}, \nu)^2 d\mathbb{Q}^1(\nu) + (1 - \lambda) \int_{\mathcal{W}_2(X)} d_W(\mu_{\mathbb{Q}^2}, \nu)^2 d\mathbb{Q}^2(\nu) \\
&= \int_{\mathcal{W}_2(X)} [\lambda d_W(\mu_{\mathbb{Q}^1}, \nu^1)^2 + (1 - \lambda) d_W(\mu_{\mathbb{Q}^2}, \nu^2)^2] d\mathbb{P}(\nu) \\
&\geq \int_{\mathcal{W}_2(X)} d_W(\bar{\mu}, \nu)^2 d\mathbb{P}(\nu).
\end{aligned}$$

Hence, if our claim were false, the expression $\int_{\mathcal{W}_2(X)} d_W(\cdot, \nu)^2 d\mathbb{P}(\nu)$ would admit a strictly smaller value on the measure $\bar{\mu}$ than on the barycenter measure μ . The proof is complete. \square

A key property used frequently in conjunction with Proposition 3.4 is that the map \mathcal{R}^i sends a probability measure ν to the measure ν^i that is absolutely continuous with respect to ν . The following corollary presents two direct consequences of this property.

Corollary 3.5. *Let (X, d) be a proper metric space and $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(X))$ with barycenter $\mu_{\mathbb{P}} \in \mathcal{W}_2(X)$. Assume the existence of a decomposition as in (12):*

$$\mu_{\mathbb{P}} = \lambda \mu^1 + (1 - \lambda) \mu^2$$

with $\mu^i \in \mathcal{W}_2(X)$ for $i = 1, 2$ and $\lambda \in (0, 1)$. Keeping the notation in Proposition 3.4, for any measure η on X , the following properties hold

1. if the measure \mathbb{P} gives mass to the set of probability measures that are absolutely continuous with respect to η , then so do the measures \mathbb{Q}^1 and \mathbb{Q}^2 ;
2. if the measure \mathbb{P} assigns full mass to the set of probability measures that are absolutely continuous with respect to η , then so do the measures \mathbb{Q}^1 and \mathbb{Q}^2 .

Proof. Proposition 3.4 provides us with two continuous maps $\mathcal{R}^1, \mathcal{R}^2 : \mathcal{W}_2(X) \rightarrow \mathcal{W}_2(X)$ such that

$$\lambda \mathcal{R}^1(\nu) + (1 - \lambda) \mathcal{R}^2(\nu) = \nu$$

for $\nu \in \mathcal{W}_2(X)$ and μ^i ($i = 1, 2$) is a barycenter of $\mathbb{Q}^i := \mathcal{R}^i_{\#} \mathbb{P}$. Observe that the probability measures $\mathcal{R}^1(\nu), \mathcal{R}^2(\nu)$ are absolutely continuous with respect to ν . Hence, given a measure η on X , if ν is absolutely continuous with respect to it, then so are the measures $\mathcal{R}^1(\nu)$ and $\mathcal{R}^2(\nu)$. The two assertions in the corollary then follow directly. \square

3.2 Subsets of geodesics in a metric graph and induced properties

In this part, we fix \vec{e} a *minimising* oriented edge of (G, d) . The result we obtain here will be used in the next part.

We decompose the space of geodesics in (G, d) starting at a point lying in e into three pieces. This decomposition of the space of geodesics depends on the orientation of e . We set

$$\mathcal{Geo}_{\gamma(0) \in \vec{e}}(G) = \mathcal{Geo}(\vec{e}) \sqcup \mathcal{Geo}(G)^{\vec{e}, +} \sqcup \mathcal{Geo}(G)^{\vec{e}, -}, \quad (15)$$

where

- $\mathcal{Geo}_{\gamma(0) \in \vec{e}}(G) = \{\gamma \in \mathcal{Geo}(G); \gamma(0) \in e\}$
- $\mathcal{Geo}(\vec{e}) = \{\gamma \in \mathcal{Geo}_{\gamma(0) \in \vec{e}}(G); \gamma([0, 1]) \subset \vec{e}\}$
- $\mathcal{Geo}(G)^{\vec{e},+} = \{\gamma \in \mathcal{Geo}_{\gamma(0) \in \vec{e}}(G) \setminus \mathcal{Geo}(\vec{e}); \vec{e}_1 \in \mathcal{Im}(\gamma)\}$
- $\mathcal{Geo}(G)^{\vec{e},-} = \{\gamma \in \mathcal{Geo}_{\gamma(0) \in \vec{e}}(G) \setminus \mathcal{Geo}(\vec{e}); \vec{e}_0 \in \mathcal{Im}(\gamma)\}$

Note that since \vec{e} is minimising, the set $\mathcal{Geo}(G)^{\vec{e},+} \cap \mathcal{Geo}(G)^{\vec{e},-}$ has to be empty.

Lemma 3.6. *The sets $\mathcal{Geo}(\vec{e}), \mathcal{Geo}(G)^{\vec{e},+}, \mathcal{Geo}(G)^{\vec{e},-}$ are Borel subsets of $\mathcal{Geo}(G)$ equipped with the topology of uniform convergence.*

Proof. Recall that $\mathcal{Geo}(G) \subset Lip([0, 1], G)$. It is clear that $\mathcal{Geo}(\vec{e})$ is a closed subset of $\mathcal{Geo}(G)$, hence measurable, since \vec{e} is closed. Let us now treat the case of $\mathcal{Geo}(G)^{\vec{e},+}$, the case $\mathcal{Geo}(G)^{\vec{e},-}$ can be proved by similar arguments. Observe that

$$\mathcal{Geo}(G)^{\vec{e},+} = \{\gamma \in \mathcal{Geo}_{\gamma(0) \in \vec{e}}(G); \min_{t \in [0, 1]} d(\vec{e}_1, \gamma(t)) = 0 \text{ and } \max_{t \in [0, 1]} d(e, \gamma(t)) > 0\}$$

where we use the fact that $d(\vec{e}_0, \vec{e}_1) = l(e)$ to get this equality. Consequently

$$\mathcal{Geo}(G)^{\vec{e},+} = \cup_{n \in \mathbb{N} \setminus \{0\}} \mathcal{G}_n,$$

where $\mathcal{G}_n = \{\gamma \in \mathcal{Geo}_{\gamma(0) \in \vec{e}}(G); \min_{t \in [0, 1]} d(\vec{e}_1, \gamma(t)) = 0 \text{ and } \max_{t \in [0, 1]} d(e, \gamma(t)) \geq 1/n\}$. Using the compactness of $[0, 1]$, one easily obtain that each set \mathcal{G}_n is a closed subset of $\mathcal{Geo}(G)$ hence $\mathcal{Geo}(G)^{\vec{e},+}$ is measurable. □

As a consequence of the existence of this measurable partition of $\mathcal{Geo}_{\gamma(0) \in \vec{e}}(G)$, we get a decomposition of any measure Π in $\mathcal{P}(\mathcal{Geo}(G))$ such that $\gamma(0) \in \vec{e}$ Π -almost surely. We simply consider the restriction of Π to each of the three subsets $\mathcal{Geo}(\vec{e}), \mathcal{Geo}(G)^{\vec{e},+}$, and $\mathcal{Geo}(G)^{\vec{e},-}$ above.

Definition 3.7 (Decomposition of a dynamical plan). Let $\Pi \in \mathcal{P}(\mathcal{Geo}(G))$ and $\vec{e} \in \vec{E}$. Assume that $\gamma(0) \in \vec{e}$ for Π -a.e γ . Then, Π has a unique decomposition in $\mathcal{P}(\mathcal{Geo}(G))$.

$$\Pi = \Pi^{\vec{e}} + \Pi^+ + \Pi^- . \tag{16}$$

This decomposition depends on \vec{e} .

We now introduced a collection of three Lipschitz maps, each of which will be used in conjunction with one of the above subsets of $\mathcal{Geo}(G)$.

Definition 3.8. We define the following maps

- $h^{\vec{e}} : G \longrightarrow \mathbb{R}$
 $x \longmapsto d(\vec{e}_0, x)$
- $h^+ : G \longrightarrow \mathbb{R}$
 $y \longmapsto l(e) + d(\vec{e}_1, y)$
- $h^- : G \longrightarrow \mathbb{R}$
 $y \longmapsto -d(\vec{e}_0, y)$

These maps are 1-Lipschitz maps hence Borel measurable.

We conclude this part with the definition of a measurable map from $\mathcal{W}_2(G)$ to $\mathcal{W}_2(\mathbb{R})$ whose properties will be proved in the next section.

Definition 3.9. For a given $\bar{\mu} \in \mathcal{W}_2(e)$, and based on the decomposition $\Pi = \Pi^{\bar{e}} + \Pi^+ + \Pi^-$ from (16), we define a map

$$\Phi : \mathcal{W}_2(G) \longrightarrow \mathcal{W}_2(\mathbb{R})$$

by the formula

$$\Phi(\nu) := h_{\sharp}^{\bar{e}} \nu^{\bar{e}} + h_{\sharp}^+ \nu^+ + h_{\sharp}^- \nu^-, \quad (17)$$

where for $i \in \{\bar{e}, +, -\}$, $\nu^i = p_{2\sharp}(\Pi_{\bar{\mu}}^i)^i$.

Remark 3.10. Note the alternative characterisation of the three terms in the definition of $\Phi(\nu)$.

$$\begin{aligned} \Phi(\nu) \llcorner_{[0, l(e)]} &= h_{\sharp}^{\bar{e}} \nu^{\bar{e}} \\ \Phi(\nu) \llcorner_{(l(e), \infty)} &= h_{\sharp}^+ \nu^+ \\ \Phi(\nu) \llcorner_{(-\infty, 0)} &= h_{\sharp}^- \nu^- \end{aligned}$$

Besides, with our convention, for $\bar{e} \sim [0, l(e)]$, note that $h^{\bar{e}}(x) = d(\bar{e}_0, x) = x$ whenever $x \in \bar{e}$. Therefore, given any $\mu \in \mathcal{W}_2(e)$, $h_{\sharp}^{\bar{e}} \mu = \mu$ and the map Φ (restricted to $\mathcal{W}_2(e)$) is *surjective onto* $\mathcal{W}_2([0, l(e)])$.

4 Branched covering & quasi regularity

In this part we study the properties of the map Φ . We call Φ a “branched covering” because of Lemma 4.2 stating that the cardinals of the preimages of the maps involved in the definition of Φ are locally constant.

Let us recall that the edge \bar{e} involved in the definition of Φ is assumed to be minimising.

4.1 A branched covering map from $\mathcal{W}_2(G)$ to $\mathcal{W}_2(\mathbb{R})$

The goal of this part is to prove

Theorem 4.1. *The map $\Phi : \mathcal{W}_2(G) \longrightarrow \mathcal{W}_2(\mathbb{R})$ satisfies the following properties:*

1. *The map Φ preserves the regularity with respect to the reference measure:*

$$\nu \ll \mathcal{H} \Leftrightarrow \Phi(\nu) \ll \mathcal{L}^1,$$

where \mathcal{L}^1 is the Lebesgue measure on \mathbb{R} .

2. $\forall \mu \in \mathcal{W}_2(e), \forall \nu \in \mathcal{W}_2(G),$

$$d_W(\Phi(\mu), \Phi(\nu)) \geq d_W(\mu, \nu),$$

3. *When $\mu = \bar{\mu}$ (where $\bar{\mu}$ is as in Definition 3.9), equality holds:*

$$d_W(\Phi(\bar{\mu}), \Phi(\nu)) = d_W(\bar{\mu}, \nu).$$

Before proving this theorem, we need to establish some intermediate results.

Lemma 4.2. *For $i \in \{\bar{e}, +, -\}$, there exists a discrete set D^i such that the cardinal of $(h^i)^{-1}(\tilde{x})$ is constant on each connected component of $\mathbb{R} \setminus D^i$.*

Proof. Let us fix $i \in \{\bar{e}, +, -\}$ and recall that the interior of an edge \bar{e} is unambiguously identified with the interval $]0, l(e)[$. Define f_i to be \bar{e}_0 if $i \in \{\bar{e}, -\}$ and $f_i = \bar{e}_1$ otherwise. Then, consider $D^i = h^i(V \cup \overrightarrow{Cut}_{f_i})$. According to Lemma 2.11, the set $V \cup \overrightarrow{Cut}_{f_i}$ is a discrete set. We infer the same property for D^i since $(h^i)^{-1}([-R, R]) \subset \bar{B}(f_i, R)$ which implies that $(h^i)^{-1}([-R, R]) \cap (V \cup \overrightarrow{Cut}_{f_i})$ is finite. Now, for $x \in (h^i)^{-1}(\tilde{x})$ with $\tilde{x} \notin D^i$, we claim that h^i is differentiable at x and $|\nabla h_i(x)| = 1$. The differentiability of h^i follows from the fact that $x \notin V \cup \overrightarrow{Cut}_{f_i}$ which implies that the distance function $d(f_i, \cdot)$ is differentiable at x , we get the result since h^i coincides up to an additive constant and switching the sign with $d(f_i, \cdot)$. The local inverse theorem then gives us the fact that $\sharp((h^i)^{-1}(\tilde{x}))$ is locally constant and the proof is complete. \square

Remark 4.3. As a consequence of the proof, note that for any interval $I_x \subset G \setminus (V \cup \overrightarrow{Cut}_{f_i})$ the inverse of $h^i \llcorner_{I_x}$ is 1-Lipschitz as well hence measurable.

We also need the following result which is a consequence of the disintegration theorem as stated in [4, 34] for instance.

Lemma 4.4. *Let (G, d) be a metric graph and $h : G \rightarrow \mathbb{R}$ be a Lipschitz map. Assume the existence of an (open) interval $E \subset \mathbb{R}$ and $S = \sqcup_{i=1}^k s_i \subset G$ such that for each $i \in \{1, \dots, k\}$, $h|_{s_i} : s_i \rightarrow E$ is a bi-Lipschitz homeomorphism onto E and $h^{-1}(E) = S$. For $\nu \in \mathcal{W}_2(G)$, set $\tilde{\nu} := h_{\sharp} \nu$.*

Then, $\tilde{\nu} \llcorner_E = h_{\sharp}(\nu \llcorner_S)$ and

$$\nu \llcorner_S = \int_E p_{\tilde{y}} d\tilde{\nu}(\tilde{y}) \quad (18)$$

where for $\tilde{\nu} \llcorner_E$ -a.e. \tilde{y} , $p_{\tilde{y}}$ is a probability measure supported in the finite set $h^{-1}(\tilde{y})$.

Moreover, if $\tilde{\theta} \in \mathcal{M}(\mathbb{R}^2)$ is a Borel measure on \mathbb{R}^2 such that its second marginal satisfies $(p_2)_{\sharp} \tilde{\theta} \ll \tilde{\nu} \llcorner_E$ then there exists a plan $\theta \in \mathcal{W}_2(\mathbb{R} \times G)$ such that

$$(Id, h)_{\sharp} \theta = \tilde{\theta}.$$

Proof. The proof of $\tilde{\nu} \llcorner_E = h_{\sharp}(\nu \llcorner_S)$ follows from $h^{-1}(E) = S$. The formula (18), is then a straightforward application of the disintegration theorem. The fact that $p_{\tilde{y}}$ is concentrated on a finite set follows from the equality $S = \sqcup_{i=1}^k s_i$.

Let us now prove the second part of the lemma. Set f the density function defined by

$$(p_2)_{\sharp} \tilde{\theta} = f \tilde{\nu} \llcorner_E.$$

Observe that

$$\tilde{\nu} \llcorner_E = \sum_{i=1}^k \tilde{\nu}^i \quad (19)$$

where $\tilde{\nu}^i = h_{\sharp}(\nu \llcorner_{s_i})$ and

$$(h|_{s_i})_{\sharp}^{-1} \tilde{\nu}^i = (h|_{s_i})_{\sharp}^{-1}(h_{\sharp}(\nu \llcorner_{s_i})) = \nu \llcorner_{s_i}.$$

Disintegrate $\tilde{\theta}$ with respect to its second marginal:

$$\tilde{\theta} = \int \tilde{\theta}_{\tilde{y}} f(\tilde{y}) \, d\tilde{\nu} \llcorner_E(\tilde{y}).$$

Then decompose $\tilde{\theta}$ as follows

$$\tilde{\theta} := \sum_{i=1}^k \tilde{\theta}^i$$

where

$$\tilde{\theta}^i = \int \tilde{\theta}_{\tilde{y}}^i f(\tilde{y}) \, d\tilde{\nu}^i(\tilde{y}).$$

Note that by assumption on the second marginal of $\tilde{\theta}$, for any $i \in \{1, \dots, k\}$,

$$\tilde{\theta}^i(\mathbb{R} \times (\mathbb{R} \setminus E)) = 0. \quad (20)$$

We now define the measure θ as follows

$$\theta = \sum_{i=1}^k \theta^i$$

where

$$\theta^i := (Id, (h|_{s_i})^{-1})_{\#} \tilde{\theta}^i$$

is well-defined thanks to (20).

Now, observe that

$$\begin{aligned} (Id, h)_{\#} \theta &= \sum_{i=1}^k (Id, h)_{\#} \theta^i \\ &= \sum_{i=1}^k (Id, h|_{s_i})_{\#} \theta^i \\ &= \sum_{i=1}^k (Id, h|_{s_i})_{\#} ((Id, (h|_{s_i})^{-1})_{\#} \tilde{\theta}^i) \\ &= \sum_{i=1}^k \tilde{\theta}^i \\ &= \tilde{\theta}. \end{aligned}$$

□

We are now in position to tackle the proof of Theorem 4.1.

Proof. (of Theorem 4.1).

1. By definition of Φ , it suffices to prove that $\nu^i \ll \mathcal{H}$ iff $h_{\#}^i \nu^i \ll \mathcal{L}^1$ for any $i \in \{\vec{e}, +, -\}$. Since each h^i is a Lipschitz map, the backward implication is a standard result (recall that \mathcal{H} is the 1-Hausdorff measure induced by the graph distance). To prove the reverse implication, we make use of Lemma 4.2. Let $I \subset \mathbb{R} \setminus D^i$ be a connected component of $\mathbb{R} \setminus D^i$. Then $(h^i)^{-1}(I) = \sum_{j=1}^k s_j$ where each s_j is an interval within an edge e_j , and e_l and e_j are distinct whenever $l \neq j$. We denote by $(h_j^i)^{-1}$ the inverse map of the bijection $h^i : s_j \rightarrow I$. As explained in the proof of Lemma 4.2, the map $(h_j^i)^{-1}$ is 1-Lipschitz hence measurable. By definition,

$$((h^i)_{\#} \nu^i) \llcorner_I = \sum_{j=1}^k (h_j^i)_{\#} (\nu^i \llcorner_{s_j}).$$

Moreover, $h^i|_{s_j}$ being a biLipschitz homeomorphism, $(h^i)_{\#} (\nu^i \llcorner_{s_j})$ is a.c. with respect to \mathcal{L}^1 if (and only if) $\nu^i \llcorner_{s_j}$ is a.c. with respect to \mathcal{H} hence the result.

2. Let $\mu \in \mathcal{W}_2(e)$, and $\nu \in \mathcal{W}_2(G)$. Given $\tilde{\theta} \in \Gamma^o(\Phi(\mu), \Phi(\nu))$, our goal is to construct a plan $\theta \in \Gamma(\mu, \nu)$ such that

$$\int d^2(x, y) \, d\theta \leq \int |\tilde{x} - \tilde{y}|^2 \, d\tilde{\theta}.$$

Pursuing the approach initiated to prove the first item, we first construct a partition of the real line as follows. Let us set

$$D := D^+ \cap (l(e), \infty) \sqcup D^- \cap (-\infty, 0) \sqcup D^{\vec{e}} \cap [0, l(e)],$$

where the sets D^i , $i \in \{+, -, \vec{e}\}$ are discrete subsets of \mathbb{R} defined in Lemma 4.2, in particular the set D is discrete as well.

Consequently, we get an at most countable partition of the real line $\{E_b, b \in N\}$ where either $E_b = \{d\}$ for some $d \in D$ or E_b is a connected component of $\mathbb{R} \setminus D$. We denote by $A \subset N$ the subset for which $\{E_b, b \in A\}$ is a partition of $[0, l(e)]$.

The next step consists in decomposing the transport plan $\tilde{\theta}$ by making use of the partition above.

$$\tilde{\theta} = \sum_{a \in A, b \in N} \tilde{\theta} \llcorner_{E_a \times E_b}.$$

Now for a fixed $(a, b) \in A \times N$, let us set $\Xi_{a,b} := \tilde{\theta} \llcorner_{E_a \times E_b}$ and assume $\Xi_{a,b} \not\equiv 0$. Note that $E_a \subset [0, l(e)]$ and by assumption $d_{\vec{e}_0} : \vec{e} \rightarrow [0, l(e)]$, where $d_{\vec{e}_0}(x) = d(\vec{e}_0, x)$, is an isometry. Thus we only need to focus on the second marginal of $\Xi_{a,b}$. In the case when $E_b = \{d\}$ with $d \in D_i$ then $(h^i)^{-1}(d) = \{g_1, \dots, g_s\}$ ($i \in \{+, -, \vec{e}\}$) is a finite set. Thus if we set $\tilde{\nu} = h_{\#}^i \nu$, we have

$$\tilde{\nu}(\{d\}) = \sum_{i=1}^s \nu(\{g_i\}).$$

Then we simply consider the transport plan

$$\theta_{a,b} := \sum_{i=1}^s \nu(\{g_i\})(p_{1\#} \Xi_{a,b}) \otimes \delta_{g_i}.$$

Since by definition $\tilde{\theta}_{a,b}(\mathbb{R} \times (\mathbb{R} \setminus \{d\})) = 0$, we have

$$(Id, h^i)_{\#} \theta_{a,b} = \tilde{\theta}_{a,b}. \quad (21)$$

It remains to treat the case where E_b is an (open) connected component of $\mathbb{R} \setminus D$, note that by construction E_b is contained in one of these sets: $(-\infty, 0)$, $[0, l(e)]$, $(l(e), \infty)$ and thus $\tilde{\nu} \llcorner_{E_b} = ((h^i)_{\#} \nu) \llcorner_{E_b}$ for a uniquely defined $i \in \{+, -, \bar{e}\}$. Thanks to Lemma 4.4, there exists a plan $\theta_{a,b}$ such that

$$(h^i)_{\#} \theta_{a,b} = \Xi_{a,b}. \quad (22)$$

We finally define θ as

$$\theta := \sum_{a \in A, b \in N} \theta_{a,b}.$$

By definition of Φ and the partition $\{E_b, b \in N\}$, the properties (21) and (22) give us that $\theta \in \Gamma(\mu, \nu)$.

We are left with proving the inequality $\int d^2(x, y) d\theta \leq \int |\tilde{x} - \tilde{y}|^2 d\tilde{\theta}$ that gives the thesis by definition of the Wasserstein distance. As already observed above, for each E_b , $b \in B$, E_b is contained in one of these sets: $(-\infty, 0)$, $[0, l(e)]$, $(l(e), \infty)$. Let us treat the case where $E_b \subset (l(e), \infty)$, the others are proved similarly. By what precedes,

$$\begin{aligned} \int |\tilde{x} - \tilde{y}|^2 d\Xi_{a,b} &= \int |h^{\bar{e}}(x) - h^+(y)|^2 d\theta_{a,b}(x, y) \\ &= \int |d(\bar{e}_0, x) - (l(e) + d(\bar{e}_1, y))|^2 d\theta_{a,b}(x, y) \\ &= \int |-d(\bar{e}_1, x) - d(\bar{e}_1, y)|^2 d\theta_{a,b}(x, y) \\ &\geq \int d^2(x, y) d\theta_{a,b}(x, y), \end{aligned}$$

where we use the property $d(\bar{e}_0, \bar{e}_1) = l(e)$ in the second equality.

3. Let us now assume that $\mu = \bar{\mu}$ and prove $d_W(\Phi(\bar{\mu}), \Phi(\nu)) = d_W(\bar{\mu}, \nu)$. The proof is mainly based on the following observation:

$$\left. \begin{aligned} &\bullet \text{ for } \Pi^{\bar{e}}\text{-a.e geodesic } \gamma \in \mathcal{Geo}_{\gamma(0) \in \bar{e}}(G), \\ &\quad d(\gamma(0), \gamma(1)) = |d(\bar{e}_0, \gamma(0)) - d(\bar{e}_0, \gamma(1))| \\ &\bullet \text{ for } \Pi^+\text{-a.e geodesic } \gamma \in \mathcal{Geo}_{\gamma(0) \in \bar{e}}(G), \\ &\quad d(\gamma(0), \gamma(1)) = d(\bar{e}_1, \gamma(0)) + d(\bar{e}_1, \gamma(1)) \\ &\quad \quad \quad = -d(\bar{e}_0, \gamma(0)) + l(e) + d(\bar{e}_1, \gamma(1)), \\ &\bullet \text{ for } \Pi^-\text{-a.e geodesic } \gamma \in \mathcal{Geo}_{\gamma(0) \in \bar{e}}(G), \\ &\quad d(\gamma(0), \gamma(1)) = d(\bar{e}_0, \gamma(0)) + d(\bar{e}_0, \gamma(1)). \end{aligned} \right\} \quad (23)$$

We are then left with estimating $d_W(\Phi(\bar{\mu}), \Phi(\nu))$ from above. To this aim, we set $\pi^i := (\cdot_0, \cdot_1)_\# \Pi^i$ for $i \in \{+, -, \vec{e}\}$ and compute, using the definition of Φ ,

$$\begin{aligned}
d_W^2(\Phi(\bar{\mu}), \Phi(\nu)) &\leq \sum_{i \in \{+, -, \vec{e}\}} \int |h^{\vec{e}}(x) - h^i(y)|^2 d\pi^i(x, y) \\
&= \int |d(\vec{e}_0, x) - d(\vec{e}_0, x)|^2 d\pi^{\vec{e}}(x, y) \\
&\quad + \int |d(\vec{e}_0, x) - (l(e) + d(\vec{e}_1, y))|^2 d\pi^+(x, y) \\
&\quad + \int |d(\vec{e}_0, x) - (-d(\vec{e}_0, y))|^2 d\pi^-(x, y) \\
&= \sum_{i \in \{+, -, \vec{e}\}} \int d^2(x, y) d\pi^i(x, y) = d_W^2(\bar{\mu}, \nu),
\end{aligned}$$

where the penultimate equality follows from (23). □

4.2 Proof of the main results

The goal of this part is to prove our main results which we recall below:

Theorem 4.5 (Quasi regularity of Wasserstein barycenters on metric graphs). *Let (G, d, \mathcal{H}) be a metric measure graph. Consider a measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(G))$ that gives mass to absolutely continuous measures on (G, d) . If $\mu_{\mathbb{P}}$ is a barycenter of \mathbb{P} , then the restriction of $\mu_{\mathbb{P}}$ to the interior of any given edge is absolutely continuous. Therefore, if $\mu_{\mathbb{P}}$ is not absolutely continuous then its singular part has to be a sum of Dirac measures located at vertices of G .*

Then we prove the Wasserstein barycenter is not unique (Proposition 1.3).

The proof of the above theorem is in two steps. In the first one, we solve the particular case when the barycenter is concentrated within an edge of (G, d) . To do so, we build upon the branched Wasserstein covering from Definition 3.9. In the second step, we take advantage of the restriction property for Wasserstein barycenters from Section 3.9 to reduce the proof in the general case to that of the first step.

4.2.1 When the barycenter is concentrated in an edge

Lemma 4.6. *Let (G, d) be a metric graph. Fix an oriented edge \vec{e} of (G, d) and a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(G))$. Suppose that \mathbb{P} has a barycenter $\mu_{\mathbb{P}} \in \mathcal{W}_2(G)$ that is supported in the edge \vec{e} of (G, d) . Denote by $\Phi : \mathcal{W}_2(G) \rightarrow \mathcal{W}_2(\mathbb{R})$ the map introduced in Definition 3.9 with $\bar{\mu} = \mu_{\mathbb{P}}$ and define $\mathbb{Q} := \Phi_{\#} \mathbb{P}$.*

Then the quantile function of $\Phi(\mu_{\mathbb{P}})$ is determined by the quantile function of \mathbb{Q} as follows: for $t \in [0, 1]$,

$$f_{\Phi(\mu_{\mathbb{P}})}^{-1}(t) = \begin{cases} 0 & \text{if } f_{\mu_{\mathbb{Q}}}^{-1}(t) < 0 \\ f_{\mu_{\mathbb{Q}}}^{-1}(t) & \text{if } 0 \leq f_{\mu_{\mathbb{Q}}}^{-1}(t) \leq l(\vec{e}) \\ l(\vec{e}) & \text{if } f_{\mu_{\mathbb{Q}}}^{-1}(t) > l(\vec{e}). \end{cases}$$

Proof. Since $\mu_{\mathbb{P}}$ is a barycenter of \mathbb{P} that is supported in the edge \vec{e} , according to the second and third items of Theorem 4.1 we have

$$\begin{aligned} \inf_{\mu \in \mathcal{W}_2(\vec{e})} \int_{\mathcal{W}_2(G)} d_W(\Phi(\mu), \Phi(\nu))^2 d\mathbb{P}(\nu) &\geq \inf_{\mu \in \mathcal{W}_2(\vec{e})} \int_{\mathcal{W}_2(G)} d_W(\mu, \nu)^2 d\mathbb{P}(\nu) \\ \mu_{\mathbb{P}} \text{ barycenter} \int_{\mathcal{W}_2(G)} d_W(\mu_{\mathbb{P}}, \nu)^2 d\mathbb{P}(\nu) &\stackrel{\text{Item 3.}}{=} \int_{\mathcal{W}_2(G)} d_W(\Phi(\mu_{\mathbb{P}}), \Phi(\nu))^2 d\mathbb{P}(\nu). \end{aligned} \quad (24)$$

By assumption $\mu_{\mathbb{P}} \in \mathcal{W}_2(\vec{e})$ thus all the terms in (24) are equal.

Now, using the surjectivity property explained in Remark 3.10, and the definition of $\mathbb{Q} := \Phi_{\#}\mathbb{P}$, we can rephrase (24) as

$$\begin{aligned} \inf_{\mu \in \mathcal{W}_2([0, l(\vec{e})])} \int_{\mathcal{W}_2(\mathbb{R})} d_W(\mu, \nu)^2 d\mathbb{Q}(\nu) &= \inf_{\mu \in \mathcal{W}_2(\vec{e})} \int_{\mathcal{W}_2(G)} d_W(\Phi(\mu), \Phi(\nu))^2 d\mathbb{P}(\nu) \\ &= \int_{\mathcal{W}_2(\mathbb{R})} d_W(\Phi(\mu_{\mathbb{P}}), \nu)^2 d\mathbb{Q}(\nu). \end{aligned} \quad (25)$$

Denote by $\mu_{\mathbb{Q}}$ the unique Wasserstein barycenter of $\mathbb{Q} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$. According to Theorem 2.26, the quantile function of $\mu_{\mathbb{Q}}$ satisfies:

$$f_{\mu_{\mathbb{Q}}}^{-1} = \int_{\mathcal{W}_2(\mathbb{R})} f_{\nu}^{-1} d\mathbb{Q}(\nu).$$

To further simplify (25), we build upon the isometric feature of $\nu \mapsto f_{\nu}^{-1}$ to get

$$\begin{aligned} \int_{\mathcal{W}_2(\mathbb{R})} d_W(\mu, \nu)^2 d\mathbb{Q}(\nu) &= \int_0^1 \int_{\mathcal{W}_2(\mathbb{R})} [f_{\mu}^{-1}(t) - f_{\nu}^{-1}(t)]^2 d\mathbb{Q}(\nu) dt \\ &= \int_0^1 \left[f_{\mu}^{-1}(t) - \int_{\mathcal{W}_2(\mathbb{R})} f_{\nu}^{-1}(t) d\mathbb{Q}(\nu) \right]^2 dt + I(\mathbb{Q}) \\ &= \int_0^1 [f_{\mu}^{-1} - f_{\mu_{\mathbb{Q}}}^{-1}]^2 d\lambda + I(\mathbb{Q}), \end{aligned}$$

where the abbreviated term $I(\mathbb{Q})$ is independent of μ :

$$I(\mathbb{Q}) := \int_0^1 \int_{\mathcal{W}_2(\mathbb{R})} [f_{\nu}^{-1}(t)]^2 d\mathbb{Q}(\nu) - \left(\int_{\mathcal{W}_2(\mathbb{R})} f_{\nu}^{-1}(t) d\mathbb{Q}(\nu) \right)^2 dt.$$

Hence, (25) is equivalent to

$$\inf_{\mu \in \mathcal{W}_2([0, l(\vec{e})])} \int_0^1 [f_{\mu}^{-1}(t) - f_{\mu_{\mathbb{Q}}}^{-1}(t)]^2 dt = \int_0^1 [f_{\Phi(\mu_{\mathbb{P}})}^{-1}(t) - f_{\mu_{\mathbb{Q}}}^{-1}(t)]^2 dt. \quad (26)$$

By Lemma 2.22, a probability measure $\mu \in \mathcal{W}_2(\mathbb{R})$ is in the space $\mathcal{W}_2([0, l(\vec{e})])$ if and only the image of f_{μ}^{-1} is contained in the interval $[0, l(\vec{e})]$. Hence, a solution $\mu \in \mathcal{W}_2([0, l(\vec{e})])$ of the minimization problem in the left-hand side of (26) must satisfy the requirements

$$f_{\mu}^{-1}(t) = 0 \text{ if } f_{\mu_{\mathbb{Q}}}^{-1}(t) < 0, \quad f_{\mu}^{-1}(t) = f_{\mu_{\mathbb{Q}}}^{-1}(t) \text{ if } 0 \leq f_{\mu_{\mathbb{Q}}}^{-1}(t) \leq l(\vec{e}), \quad f_{\mu}^{-1}(t) = l(\vec{e}) \text{ if } f_{\mu_{\mathbb{Q}}}^{-1}(t) > l(\vec{e}).$$

In particular, the measure $\Phi(\mu_{\mathbb{P}}) \in \mathcal{W}_2([0, l(\vec{e})])$ satisfies the above requirements, which concludes the proof. \square

By using Lemma 4.6, we can determine the barycenter of $\mathbb{Q} = \Phi_{\#}\mathbb{P}$ provided $\mu_{\mathbb{P}}$ is concentrated in the interior of e .

Corollary 4.7. *Let (G, d) be a metric graph. Fix an oriented edge \vec{e} of (G, d) and a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(G))$. Suppose that \mathbb{P} has a barycenter $\mu_{\mathbb{P}} \in \mathcal{W}_2(G)$ that is supported in the edge \vec{e} of (G, d) and assigns no mass to the ends of \vec{e} .*

Then $\Phi(\mu_{\mathbb{P}})$ is the unique barycenter of $\mathbb{Q} := \Phi_{\#}\mathbb{P}$.

Proof. Since $\mu_{\mathbb{P}}$ is assumed to give no mass to the ends of \vec{e} , $\Phi(\mu_{\mathbb{P}})$ gives no mass to the endpoints of $[0, l(\vec{e})]$. Hence, [33, Chapter 0, Lemma (4.8)] implies

$$\begin{aligned} 0 &= f_{\Phi(\mu_{\mathbb{P}})}(0) = \inf\{t \in (0, 1) \mid f_{\Phi(\mu_{\mathbb{P}})}^{-1}(t) > 0\}, \\ 1 &= \lim_{s \uparrow l(\vec{e})} f_{\Phi(\mu_{\mathbb{P}})}(s) = \lim_{s \uparrow l(\vec{e})} \inf\{t \in (0, 1) \mid f_{\Phi(\mu_{\mathbb{P}})}^{-1}(t) > s\}. \end{aligned}$$

It follows from these two equalities that for $0 < t < 1$, $f_{\Phi(\mu_{\mathbb{P}})}^{-1}(t) \neq 0$ and $f_{\Phi(\mu_{\mathbb{P}})}^{-1}(t) \neq l(\vec{e})$. According to the requirements satisfied by $\Phi(\mu_{\mathbb{P}})$ in Lemma 4.6, we must have $0 \leq f_{\mu_{\mathbb{Q}}}^{-1}(t) \leq l(\vec{e})$ and $f_{\Phi(\mu_{\mathbb{P}})}^{-1}(t) = f_{\mu_{\mathbb{Q}}}^{-1}(t)$ for $t \in (0, 1)$. Hence, $\Phi(\mu_{\mathbb{P}}) = \mu_{\mathbb{Q}}$ is the unique barycenter of \mathbb{Q} . \square

The proof of our main result in the case $\mu_{\mathbb{P}}(\hat{e}) = 1$ is a straightforward consequence of the corollary above as explained below. We then use the restriction property of Wasserstein barycenters to prove the general case:

4.2.2 Proof of Theorem 4.5

Fix an oriented edge \vec{e} of (G, d) and denote by \hat{e} the interior of \vec{e} . We are going to prove that the restriction of $\mu_{\mathbb{P}}$ to \hat{e} is absolutely continuous with respect to \mathcal{H} . To do so, it suffices to prove that this property holds for any sufficiently small open subintervals of \hat{e} . Therefore, let us now fix $x \in \hat{e}$ and $I =]a, b[\ni x$ an open subinterval of \hat{e} such that the graph distance between a and b satisfies $d(a, b) = |b - a|$; such a I always exists. If $I \neq \hat{e}$ (for instance if e is *not* a minimising edge of (G, d)) then turn G into a new graph G_1 by adding a and b to the set of vertices and dividing e into three edges in the straightforward way. This new metric graph G_1 satisfies the same assumptions as (G, d) (up to reducing the positive lower bound on the edge length). Therefore, without loss of generality, we can assume that \vec{e} is a minimising edge. Let us now discuss all the alternatives.

If $\mu_{\mathbb{P}}$ gives no mass to the set \hat{e} , then its restriction on \hat{e} is null, and thus absolutely continuous. Consider now the case where $\mu_{\mathbb{P}}(\hat{e}) > 0$ and denote by $\mu_{\hat{e}} \in \mathcal{W}_2(G)$ the normalized probability measure of the restriction of $\mu_{\mathbb{P}}$ to \hat{e} . We still denote by Φ the branched Wasserstein covering introduced in Definition 3.9, recall that the properties of Φ rely on the fact that \vec{e} is minimising.

According to Item 3. of Theorem 4.1, for $\nu \in \mathcal{W}_2(G)$, $\Phi(\nu)$ is a.c. with respect to the Lebesgue measure iff $\nu \ll \mathcal{H}$. We prove the absolute continuity of $\mu_{\hat{e}}$ by discussing two different cases.

If $\mu_{\hat{e}} = \mu_{\mathbb{P}}$, then, according to Corollary 4.7, $\Phi(\mu_{\hat{e}})$ is the unique barycenter of $\mathbb{Q} := \Phi_{\#}\mathbb{P}$ which is absolutely continuous since \mathbb{Q} gives mass to absolutely continuous measures on \mathbb{R} by Theorem 2.27. It follows that $\mu_{\mathbb{P}}$ is absolutely continuous when $\mu_{\hat{e}} = \mu_{\mathbb{P}}$. We now prove that $\mu_{\hat{e}}$ is absolutely continuous when $\mu_{\hat{e}} \neq \mu_{\mathbb{P}}$. For the decomposition $\mu_{\mathbb{P}} = \lambda \mu_{\hat{e}} + (1 - \lambda) \mu^2$ with $\lambda := \mu_{\mathbb{P}}(\hat{e}) \in (0, 1)$ and $\mu^2 \in \mathcal{W}_2(G)$, Corollary 3.5 provides us with two measures $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{W}_2(\mathcal{W}_2(G))$ such that $\mu_{\hat{e}}$

is a barycenter of \mathbb{P}_1 and μ^2 is a barycenter of \mathbb{P}_2 . Moreover, Corollary 3.5 implies that both \mathbb{P}_1 and \mathbb{P}_2 give mass to absolutely continuous measures since \mathbb{P} does so. We then have $\mu_{\vec{e}} = \mu_{\mathbb{P}_1}$ is absolutely continuous as proven in the previous case.

Finally, since our argument holds for arbitrarily chosen \vec{e} , our theorem is proven.

4.2.3 Proof of Proposition 1.3

We first consider a tripod T made of segments of length 1 glued at the point v_0 as in Figure 2. Consider the oriented edge $\vec{e} = \{v_1, v_0\}$. Define $\nu_1 = \mathcal{H}|_{-\vec{e}}$ and $\nu_2 = \frac{1}{2}\delta_{v_2} + \frac{1}{2}\delta_{v_3}$ and let $\Phi : \mathcal{W}_2(T) \rightarrow$

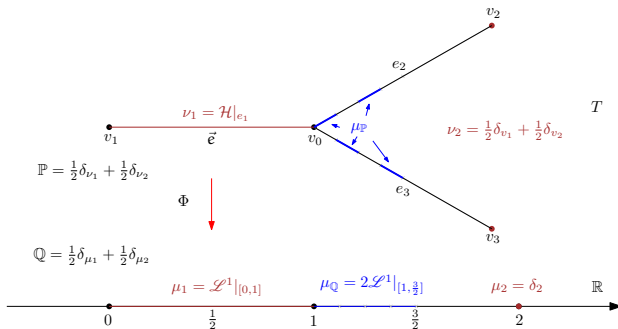


Figure 2: multiple a.c. barycenters

$\mathcal{W}_2(\mathbb{R})$ be the map associated to \vec{e} and $\bar{\mu} = \nu_1$ as in Definition 3.9. We set $\mathbb{P} = \frac{1}{2}\delta_{\nu_1} + \frac{1}{2}\delta_{\nu_2}$.

Now, consider $\mu_1 := \Phi(\nu_1) = \mathcal{L}^1|_{[0,1]}$, $\mu_2 := \Phi(\nu_2) = \delta_2$, and let us introduce

$$\mu_{\mathbb{P}} := \frac{1}{2}\eta_2 + \frac{1}{2}\eta_3$$

with η_2 (respectively η_3) being a probability measure supported in the edge e_2 (respectively e_3) such that

$$\Phi(\mu_{\mathbb{P}}) = \frac{1}{2}\Phi(\eta_2) + \frac{1}{2}\Phi(\eta_3) = 2\mathcal{L}^1|_{[1, \frac{3}{2}]}$$

Our goal is to *prove* that any such $\mu_{\mathbb{P}}$ is a barycenter of \mathbb{P} .

We finally define $\mathbb{Q} := \Phi_{\#}\mathbb{P} = \frac{1}{2}\delta_{\mu_1} + \frac{1}{2}\delta_{\mu_2}$. Recall that when a generic \mathbb{P} gives mass only to two distinct measures with equal weight, any barycenter of \mathbb{P} is a midpoint (relative to d_W) of the involved measures and vice versa. Therefore, $\mu_{\mathbb{Q}} := 2\mathcal{L}^1|_{[1, \frac{3}{2}]}$ is the barycenter of \mathbb{Q} by its very definition.

Thanks to Theorem 4.1, Item 3., we have

$$\begin{aligned} d_W(\nu_1, \mu_{\mathbb{P}}) &= d_W(\Phi(\nu_1), \Phi(\mu_{\mathbb{P}})) = d_W(\mu_1, \mu_{\mathbb{Q}}) \\ d_W(\nu_1, \nu_2) &= d_W(\Phi(\nu_1), \Phi(\nu_2)) = d_W(\mu_1, \mu_2). \end{aligned} \tag{27}$$

The proof is complete if we check that

$$d_W(\nu_1, \mu_{\mathbb{P}}) + d_W(\mu_{\mathbb{P}}, \nu_2) = d_W(\nu_1, \nu_2).$$

According to (27), and since $\mu_{\mathbb{Q}}$ is a midpoint of μ_1 and μ_2 , we are done if we prove $d_W(\mu_2, \mu_{\mathbb{Q}}) = d_W(\nu_2, \mu_{\mathbb{P}})$.

To this aim, we use the fact that $e_2 \cup e_3$ is isometric to a segment $[0, 2]$ (where v_2 is identified with 0 and v_3 with 2), and the image of η_2 (resp. η_3) in $[0, 2]$ is a probability measure supported in $[0, 1]$ (resp. $[1, 2]$). Therefore the following transport plan γ between $\mu_{\mathbb{P}}$ and ν_2 is optimal:

$$\gamma := \frac{1}{2}\eta_2 \otimes \delta_{v_2} + \frac{1}{2}\eta_3 \otimes \delta_{v_3}.$$

We can now estimate the distance between $\mu_{\mathbb{P}}$ and ν_2 :

$$\begin{aligned} d_W(\mu_{\mathbb{P}}, \nu_2)^2 &= \int_{\Gamma} d(x, y)^2 d\gamma(x, y) \\ &= \frac{1}{2} \int_{e_2} d(x, v_2)^2 d\eta_2(x) + \frac{1}{2} \int_{e_3} d(x, v_3)^2 d\eta_3(x) \\ &= \frac{1}{2} d_W(\Phi(\eta_2), \mu_2)^2 + \frac{1}{2} d_W(\Phi(\eta_3), \mu_2)^2 \\ &= d_W(\Phi(\mu_{\mathbb{P}}), \mu_2)^2. \end{aligned}$$

From our construction of $\mu_{\mathbb{P}}$, there are infinitely many possible choices and all of them are absolutely continuous. We have thus proven that the absolutely continuous part of the Wasserstein barycenter as in Theorem 4.5 is not unique.

Let us now prove the non-uniqueness of the atomic part. Let us consider two copies of the tripod described in the introduction and drawn in Figure 1. We obtain a new metric graph (G, d) by gluing each segment $[1/2, 1]$ of one copy with the corresponding segment of the other one as in Figure 3. As we did in the introduction, let $\mathbb{P} := \frac{1}{3} \sum_{i=1}^3 \delta_{\nu_i}$ be a probability measure on the Wasserstein space over the glued tripod, where each ν_i is an absolutely continuous measure supported on the outer half $[1/2, 1]$ of a distinct branch, for instance the uniform probability measure on this interval. As explained in the introduction in the case of a tripod, the Dirac mass at the common point 0 is the barycenter of the corresponding \mathbb{P} . For each copy of the tripod isometrically embeds into the glued tripod (G, d) , we infer that any convex combination of Dirac masses at 0^+ and 0^- is a Wasserstein barycenter of \mathbb{P} and the non-uniqueness is proven.

Note that by isometrically embedding the two previous examples into a single metric graph (G', d) where the images are far from each other, and then by considering a probability measure \mathbb{P}' which is a balanced convex combination of (the images of) the previous \mathbb{P} 's, one gets a metric graph where the Wasserstein barycenter has both an atomic and an absolutely continuous part, each of which is non-unique.

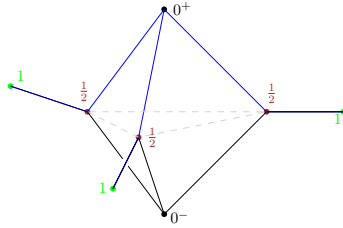


Figure 3: $\mathbb{P} = \sum_{i=1}^3 \frac{1}{3} \delta_{\nu_i}$ on the glued tripod

References

- [1] M. AGUEH AND G. CARLIER, *Barycenters in the Wasserstein space*, SIAM Journal on Mathematical Analysis, 43 (2011), pp. 904–924.
- [2] A. AHIDAR-COUTRIX, T. LE GOUIC, AND Q. PARIS, *Convergence rates for empirical barycenters in metric spaces: curvature, convexity and extendable geodesics*, Probab. Theory Relat. Fields, 177 (2020), pp. 323–368.
- [3] L. AMBROSIO, E. BRUÉ, AND D. SEMOLA, *Lectures on Optimal Transport*, vol. 130 of UNITEXT, 2021.
- [4] L. AMBROSIO, N. GIGLI, AND G. SAVARÉ, *Gradient flows: in metric spaces and in the space of probability measures*, Lectures in Mathematics. ETH Zürich, Birkhäuser, Basel, 2005.
- [5] ———, *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, Inventiones mathematicae, 195 (2014), pp. 289–391.
- [6] J.-D. BENAMOU AND Y. BRENIER, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.
- [7] J. BERTRAND, *Existence and uniqueness of optimal maps on Alexandrov spaces*, Advances in Mathematics, 219 (2008), pp. 838–851.
- [8] J. BERTRAND AND B. R. KLOECKNER, *A geometric study of Wasserstein spaces: Hadamard spaces*, Journal of Topology and Analysis, 4 (2012), pp. 515–542.
- [9] J. BIGOT AND T. KLEIN, *Characterization of barycenters in the Wasserstein space by averaging optimal transport maps*, ESAIM: Probability and Statistics, 22 (2018), pp. 35–57.
- [10] S. BOBKOV AND M. LEDOUX, *One-dimensional empirical measures, order statistics, and Kantorovich transport distances*, vol. 1259 of Mem. Am. Math. Soc., Providence, RI: American Mathematical Society (AMS), 2019.
- [11] V. I. BOGACHEV, *Measure theory*, Springer, Berlin, 2007.
- [12] D. BURAGO, I. D. BURAGO, Y. BURAGO, S. IVANOV, S. V. IVANOV, AND S. A. IVANOV, *A course in metric geometry*, vol. 33, American Mathematical Soc., 2001.
- [13] E. ÇINLAR, *Probability and Stochastics*, Graduate Texts in Mathematics, Springer New York, 2011.
- [14] M. ERBAR, D. L. FORKERT, J. MAAS, AND D. MUGNOLO, *Gradient flow formulation of diffusion equations in the Wasserstein space over a metric graph*, Networks and Heterogeneous Media, 17 (2022).
- [15] M. ERBAR, K. KUWADA, AND K.-T. STURM, *On the equivalence of the entropic curvature-dimension condition and Bochner’s inequality on metric measure spaces*, Invent. Math., 201 (2015), pp. 993–1071.
- [16] L. C. EVANS AND R. F. GARZEPY, *Measure theory and fine properties of functions*, Routledge, Oxfordshire, 2018.

- [17] N. GIGLI, T. RAJALA, AND K.-T. STURM, *Optimal maps and exponentiation on finite-dimensional spaces with Ricci curvature bounded from below*, The Journal of geometric analysis, 26 (2016), pp. 2914–2929.
- [18] B.-X. HAN, D. LIU, AND Z. ZHU, *On the geometry of Wasserstein barycenter I*, arXiv preprint arXiv:2412.01190, (2024).
- [19] T. HOTZ, S. SKWERER, S. HUCKEMANN, H. LE, J. MARRON, J. C. MATTINGLY, E. MILLER, J. NOLEN, M. OWEN, AND V. PATRANGENARU, *Sticky central limit theorems on open books*, (2013).
- [20] Y. JIANG, *Absolute continuity of Wasserstein barycenters over Alexandrov spaces*, Canadian Journal of Mathematics, 69 (2017), pp. 1087–1108.
- [21] Y.-H. KIM AND B. PASS, *Wasserstein barycenters over Riemannian manifolds*, Advances in Mathematics, 307 (2017), pp. 640–683.
- [22] T. LE GOUIC AND J.-M. LOUBES, *Existence and consistency of Wasserstein barycenters*, Probability Theory and Related Fields, 168 (2017), pp. 901–917.
- [23] S. LISINI, *Characterization of absolutely continuous curves in Wasserstein spaces*, Calc. Var. Partial Differ. Equ., 28 (2007), pp. 85–120.
- [24] J. MA, *Régularité des barycentres de Wasserstein*, phd, université de Toulouse, 2025.
- [25] ———, *Absolute continuity of Wasserstein barycenters on manifolds with a lower Ricci curvature bound*, Calc. Var. Partial Differ. Equ., 65 (2026), p. 36. Id/No 9.
- [26] L. MANELLA PEREIRA AND M. HADI AMINI, *A survey on optimal transport for machine learning: Theory and applications*, IEEE Access, 13 (2025), pp. 26506–26526.
- [27] J. M. MAZÓN, J. D. ROSSI, AND J. TOLEDO, *Optimal mass transport on metric graphs*, SIAM Journal on Optimization, 25 (2015), pp. 1609–1632.
- [28] R. J. MCCANN, *Polar factorization of maps on Riemannian manifolds*, Geometric & Functional Analysis GAFA, 11 (2001), pp. 589–608.
- [29] S.-I. OHTA, *Barycenters in Alexandrov spaces of curvature bounded below*, Advances in geometry, 12 (2012), pp. 571–587.
- [30] F. OTTO, *The geometry of dissipative evolution equations: The porous medium equation*, Commun. Partial Differ. Equations, 26 (2001), pp. 101–174.
- [31] V. M. PANARETOS AND Y. ZEMEL, *An invitation to statistics in Wasserstein space*, Springer-Briefs in Probability and Mathematical Statistics, Springer, Cham, Switzerland, 2020.
- [32] Q. PARIS, *Jensen's inequality in geodesic spaces with lower bounded curvature*, Preprintv2, (2021).
- [33] D. REVUZ AND M. YOR, *Continuous martingales and Brownian motion*, vol. 293, Springer Science & Business Media, 2013.

- [34] F. SANTAMBROGIO, *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, vol. 87 of Progress in Nonlinear Differential Equations and Their Applications, 2015.
- [35] K.-T. STURM, *Probability measures on metric spaces of nonpositive curvature*, SFB 611, 2003.
- [36] ———, *Convex functionals of probability measures and nonlinear diffusions on manifolds*, Journal de Mathématiques Pures et Appliquées, 84 (2005), pp. 149–168.
- [37] C. VILLANI, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, Rhode Island, 2003.
- [38] ———, *Optimal transport: old and new*, vol. 338 of Grundlehren der mathematischen Wissenschaften, 2009.