Absolute continuity of Wasserstein barycenters on manifolds with a lower Ricci curvature bound

Jianyu MA

Institut de Mathématiques de Toulouse

October 23, 2025

Abstract

Given a complete Riemannian manifold M with a lower Ricci curvature bound, we consider barycenters in the Wasserstein space $\mathcal{W}_2(M)$ of probability measures on M. We refer to them as Wasserstein barycenters, which by definition are probability measures on M. The goal of this article is to present a novel approach to proving their absolute continuity. We introduce a new class of displacement functionals exploiting the Hessian equality for Wasserstein barycenters. To provide suitable instances of such functionals, we revisit Souslin space theory, Dunford-Pettis theorem and the de la Vallée Poussin criterion for uniform integrability. Our method shows that if a probability measure $\mathbb{P}$ on $\mathcal{W}_2(M)$ gives mass to absolutely continuous measures on M, then its unique barycenter is also absolutely continuous. This generalizes the previous results on compact manifolds by Kim and Pass [34].

1 Introduction

Barycenter is the notion of mean for probability measures on metric spaces. Given a probability measure μ on the Euclidean space $\mathbb{R}^m$, if its first and second moments are finite, then its mean $\int_{\mathbb{R}^m} x \, \mathrm{d}\mu(x)$ can be equivalently defined as the unique point where the infimum $\inf_{y \in \mathbb{R}^m} \int_{\mathbb{R}^m} \|y - x\|^2 \, \mathrm{d}\mu(x)$ is reached. This formulation in terms of minimization and metric is still valid for general metric spaces, and it leads to our definition of barycenter (see Definition 2.1). It is worth noting that the existence of barycenters is not guaranteed *a priori* for general metric spaces. We restrict our discussions to proper metric spaces to ensure that barycenters exist.

Wasserstein spaces are metric spaces extensively studied in the field of optimal transport theory. Their geometric properties have gained constant attention. By Wasserstein barycenter we mean a barycenter of some probability measure on a given Wasserstein space. In the simplest case, given two measures μ, ν in the Wasserstein space $\mathcal{W}_2(M)$ over a Riemannian manifold M, all minimal geodesics from μ to ν are made of barycenters of $(1 - \lambda)\delta_\mu + \lambda\,\delta_\nu$ with λ varying in $[0, 1]$. Their absolute continuity (in possibly generalized settings) was previously studied as the regularity of displacement in [8, 22, 24, 42, 58]. A more general case was first studied by Agueh and Carlier [1], where barycenters of $\mathbb{P} := \sum_{i=1}^n \lambda_i\,\delta_{\mu_i}$ on the Wasserstein space $(\mathcal{W}_2(\mathbb{R}^m), W_2)$ were considered. In this setting, Wasserstein barycenters are solutions to the following minimization problem:

$$\min_\nu \sum_{i=1}^n \lambda_i\, W_2(\nu, \mu_i)^2, \quad \text{for } \nu \in \mathcal{W}_2(\mathbb{R}^m).$$

They proved the existence of barycenters constructively using a dual formulation and showed that if at least one of μ_i's is absolutely continuous with bounded density function, then the unique barycenter is also absolutely continuous. Kim and Pass [34] conducted the same study for Wasserstein barycenters on compact Riemannian manifolds M with similar conclusions. Their generalization is applicable to general probability measures $\mathbb{P}$ on $\mathcal{W}_2(M)$ that give mass to the set of absolutely continuous measures with a uniform upper density bound. The absolute continuity of Wasserstein barycenters plays an indispensable role in their study of Jensen's type inequalities for Wasserstein barycenters. There is also a generalization [31] of Agueh and Carlier's results to compact Alexandrov spaces with curvature bounded below.

When $\mathbb{P}$ has the form $\sum_{i=1}^n \lambda_i\,\delta_{\mu_i}$ with μ_1 absolutely continuous, Kim and Pass' proof of the absolute continuity of the (unique) barycenter $\overline{\mu}$ of $\mathbb{P}$ remains valid for non-compact manifolds M. For a general measure $\mathbb{P}$ giving mass to absolutely continuous measures, the strategy is to approximate $\mathbb{P}$ with finitely supported measures $\mathbb{P}_j$ whose barycenters $\overline{\mu}_j$ are already shown to be absolutely continuous. Thanks to the law of large numbers for Wasserstein barycenters (Theorem 2.2), $\overline{\mu}_j$ converges to $\overline{\mu}$ weakly. However, this is not sufficient to ensure that $\overline{\mu}$ is also absolutely continuous. To overcome this difficulty, Kim and Pass [34] imposed a uniform upper density bound on $\overline{\mu}_j$'s, which forced them to include the assumption on $\mathbb{P}$.

2

In our work, instead of following their quantitative approach, we seek for proper integral functionals F on $\mathcal{W}_2(M)$ that admit finite values only for absolutely continuous measures. The continuity of these functionals has been studied in various sources, including [15], [58, Theorem 29.20], [50, Chapter 7], and [4, Chapter 15]. We summarize their assumptions and conclusions in Lemma 5.2. Additionally, we aim to control the value of F at $\overline{\mu}_j$ by those at the support of $\mathbb{P}_j$, which enables us to use the convergence $\mathbb{P}_j \to \mathbb{P}$ effectively. Classic references, such as Villani's monograph [58], focus on the λ-convexity of F, a widely studied property that would satisfy our requirements if we tolerate some independent constants in its inequality expression of convexity (Proposition 4.2). Functionals defined in this way generalize the entropy functional $f \cdot \mathrm{Vol} \mapsto \int_M f \log f \, \mathrm{d} \mathrm{Vol}$, which is an important example in the study of synthetic treatment of Ricci curvature lower bounds developed in [41,52,53]. Proposition 4.2 reveals how Ricci curvature affects the properties of Wasserstein barycenters and suggests possible extensions of our current work to general metric measures spaces.

The methodology previously described leads us to Proposition 5.5 on the absolute continuity of Wasserstein barycenters, where an extra assumption on $\mathbb{P}$ is needed. With the help of a generalized de la Vallée Poussin criterion (Theorem 5.13), this assumption can be further simplified: we ask that $\mathbb{P}$ gives mass to a compact subset in some weak topology of absolutely continuous measures. Although this topology is barely mentioned in the literature of optimal transport, it generates the same Borel sets as the topology induced by the Wasserstein metric according to the theory of Souslin space. This helps us to state our main result with a natural assumption on $\mathbb{P}$:

Theorem. *Let $(M, \mathfrak{g})$ be a complete Riemannian manifold with a lower Ricci curvature bound. If a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(M))$ gives mass to the set of absolutely continuous probability measures on M, then its unique barycenter is absolutely continuous.*

Structure of the paper

In Section 2, we introduce notation and definitions for Wasserstein barycenters, and then extend Kim and Pass' proof of their absolute continuity to non-compact manifolds. In Section 3, we present the Hessian equality for Wasserstein barycenters (Theorem 3.13), which is used in Section 4 to justify our displacement functionals (Proposition 4.2). Section 5 primarily concerns Polish spaces, and we use the Souslin space theory to provide appropriate instances of the previously defined displacement functionals. Our main result, Theorem 5.1, is a consequence of the intermediate result Proposition 5.5 after proving several auxiliary results.

2 Wasserstein barycenters

2.1 Notation and definitions

Definition 2.1 (Barycenter)**.** Let (E, d) be a metric space and let μ be a probability measure on E such that $\int_E d(x_0, y)^2 \, \mathrm{d} \mu(y) < \infty$ for some point $x_0 \in E$. We call $z \in E$ a barycenter of μ if

$$\int_E d(z, y)^2 \, \mathrm{d} \mu(y) = \min_{x \in E} \int_E d(x, y)^2 \, \mathrm{d} \mu(y).$$

A metric space is proper if its bounded closed subsets are also compact. Barycenters always exist in proper spaces since a minimizing sequence is bounded and thus pre-compact. We refer to Ohta [45] for more details and some other properties of barycenters in a proper space.

A metric space is Polish if it is complete and separable. Any proper space (E, d) is Polish, so are the Wasserstein spaces built over it. In this article, we consider the (2-)Wasserstein space $(\mathcal{W}_2(E), W_2)$ of probability measures on E with

$$\mathcal{W}_2(E) := \left\{ \mu \text{ is a probability measure on } E \,\middle|\, \exists\, x_0 \in E, \int_E d(x_0, y)^2 \,\mathrm{d}\,\mu(y) < \infty \right\},$$

$$W_2(\mu, \nu)^2 := \inf_{\pi \in \Pi(\mu, \nu)} \int_{E \times E} d(x, y)^2 \,\mathrm{d}\,\pi(x, y), \tag{1}$$

where $\Pi(\mu, \nu)$ is the set of probability measures on $E \times E$ with marginals μ and ν. The infimum in (1) is always reached by some measure $\pi \in \Pi(\mu, \nu)$, and we call it an optimal transport plan between μ and ν.

Since Wasserstein spaces are complete and separable, we can construct the Wasserstein space $(\mathcal{W}_2(\mathcal{W}_2(E)), \mathbb{W}_2)$ over the Wasserstein space $(\mathcal{W}_2(E), W_2)$. Symbols W_2 and $\mathbb{W}_2$ will always denote Wasserstein metrics in the rest of the paper. A Wasserstein space $\mathcal{W}_2(E)$ is not proper unless the base space E is compact [6, Remark 7.19]. Classic references on this topic are [58], [50], and [57].

As mentioned before, Wasserstein barycenters are barycenters of measures on Wasserstein spaces. We refer to the following result by Le Gouic and Loubes as the law of large numbers for Wasserstein barycenters since we can set $\mathbb{P}_j$ to be empirical measures for the law $\mathbb{P}$.

Theorem 2.2 (Law of large numbers for Wasserstein barycenters, [37])**.** *Let (E, d) be a proper space. Fix a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(E))$ on $\mathcal{W}_2(E)$. Given a sequence of measures $\mathbb{P}_j \in \mathcal{W}_2(\mathcal{W}_2(E))$ with their corresponding barycenters $\overline{\mu}_j \in \mathcal{W}_2(E)$, if $\mathbb{W}_2(\mathbb{P}_j, \mathbb{P}) \to 0$ as $j \to \infty$, then $W_2(\overline{\mu}_j, \overline{\mu}) \to 0$ for some barycenter $\overline{\mu}$ of $\mathbb{P}$ up to extracting a subsequence of $\overline{\mu}_j$.*

For two topological spaces E_1 and E_2, we denote by p_1 and p_2 the canonical projection maps defined on $E_1 \times E_2$, where p_1 maps $(x, y) \in E_1 \times E_2$ to $x \in E_1$ and p_2 maps (x, y) to $y \in E_2$. Recall that these projection maps are continuous and open (mapping open sets to open sets). The map p_1 (respectively p_2) is closed if E_2 (respectively E_1) is compact [13, Proposition 8.2]. By convention, Id denotes the identity map $x \mapsto x$. The following lemma is useful when compactness arguments are needed, whose proof is based on the previous property of projection maps.

Lemma 2.3. *Let (E, d) be a proper space. Given an integer $n \geq 1$, let $\lambda_i > 0, 1 \leq i \leq n$, be n positive real numbers such that $\sum_{i=1}^n \lambda_i = 1$. The set*

$$\Gamma := \left\{ (x_1, \ldots, x_n, z) \in E^{n+1} \,\middle|\, \sum_{i=1}^n \lambda_i \, d(z, x_i)^2 = \min_{y \in E} \sum_{i=1}^n \lambda_i \, d(y, x_i)^2 \right\}$$

is closed. Denote by $\mathrm{bary}(\boldsymbol{A})$ the set of all barycenters of the measures $\sum_{i=1}^n \lambda_i \, \delta_{x_i}$ when $(x_1, \ldots, x_n)$ runs through a subset $\boldsymbol{A} \subset E^n$. If $\boldsymbol{A}$ is compact, then $\mathrm{bary}(\boldsymbol{A})$ is compact.

Proof. For $\boldsymbol{x} := (x_1, x_2, \ldots, x_n) \in E^n$, we define $\eta(\boldsymbol{x}) := \sum_{i=1}^n \lambda_i \, \delta_{x_i} \in \mathcal{W}_2(E)$. The map $\eta : (E, d) \to (\mathcal{W}_2(E), W_2)$ is continuous by definition of Wasserstein metric: for $\boldsymbol{x}, \boldsymbol{y} \in E^n$,

$$W_2(\eta(\boldsymbol{x}), \eta(\boldsymbol{y}))^2 \leq \sum_{i=1}^n \lambda_i \, d(x_i, y_i)^2.$$

It follows from the triangle inequality that the map $\boldsymbol{x} \in E^n \mapsto \min_{y \in E} W_2(\eta(\boldsymbol{x}), \delta_y)$ is also continuous, which implies that the set Γ is closed as $W_2(\eta(\boldsymbol{x}), \delta_y)^2 = \sum_{i=1}^n \lambda_i \, d(y, x_i)^2$.

4

Note that $\mathrm{bary}(\boldsymbol{A}) = p_2\left(\Gamma \cap (\boldsymbol{A} \times E)\right)$, where $p_2 : \boldsymbol{A} \times E \to E$ is the canonical projection map. If $\boldsymbol{A}$ is compact, p_2 is a closed map and thus $\mathrm{bary}(\boldsymbol{A})$ is closed as Γ is closed. The set $\mathrm{bary}(\boldsymbol{A})$ is bounded since barycenters are located within the union of n bounded balls with centers x_i. $\qquad\square$

For a metric space E, denote by $\mathcal{B}(E)$ the σ-algebra of its Borel sets. If E is separable, then the support of any Borel measure μ on E exists [11, Proposition 7.2.9]. This support, denoted by $\mathrm{supp}(\mu)$, is the closed set whose complement is the union of all open sets $U \subset E$ satisfying $\mu(U) = 0$. We shall apply the following widely used measurable selection theorem to construct Wasserstein barycenters in the next subsection. Its proof could be found in [11, Theorem 6.9.3], [25], and [36].

Theorem 2.4 (Kuratowski and Ryll-Nardzewski measurable selection theorem). *Let E be a Polish metric space, and let Ψ be a map on a measurable space $(\Omega, \mathcal{B})$ with values in the set of nonempty closed subsets of E. Suppose that for every open set $U \subset E$, we have*

$$\{\omega \in \Omega \mid \Psi(\omega) \cap U \neq \emptyset\} \in \mathcal{B}.$$

Then Ψ has a selection that is measurable with respect to the pair of σ-algebras $\mathcal{B}$ and $\mathcal{B}(E)$.

The notion of conditional measures [11, Definition 10.4.2] will be used to prove Proposition 2.22.

Definition 2.5 (Conditional probability measures). Let E be a Polish metric space and let $n \geq 2$ be a positive integer. Denote by $\boldsymbol{x}' = (x_2, \ldots, x_n) \in E^{n-1}$ the last $n-1$ components of a point $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in E^n$. Given a Borel probability measures γ on E^n, define the measure $\pi := p_{2\#}\gamma$ on E^{n-1}, where p_2 is the projection $\boldsymbol{x} \in E \times E^{n-1} \mapsto \boldsymbol{x}' \in E^{n-1}$. We call $\gamma(\cdot, \cdot) : \mathcal{B}(E^n) \times E^{n-1} \to \mathbb{R}$ a conditional measure for γ, written as $\mathrm{d}\,\gamma(\boldsymbol{x}) = \gamma(\mathrm{d}\,\boldsymbol{x}, \boldsymbol{x}')\,\mathrm{d}\,\pi(\boldsymbol{x}')$, if

1. for all $\boldsymbol{x}' \in E^{n-1}$, $\gamma(\cdot, \boldsymbol{x}')$ is a Borel probability measure on E^n,

2. for π-almost every $\boldsymbol{x}' \in E^{n-1}$, $\gamma(\cdot, \boldsymbol{x}')$ is concentrated on $E \times \{\boldsymbol{x}'\}$,

3. for any Borel set $\boldsymbol{R} \subset E^n$, the function $\boldsymbol{x}' \mapsto \gamma(\boldsymbol{R}, \boldsymbol{x}')$ is measurable, and

4. for any Borel set $\boldsymbol{S} \subset E^{n-1}$, $\gamma[\boldsymbol{R} \cap (E \times \boldsymbol{S})] = \int_{\boldsymbol{S}} \gamma(\boldsymbol{R}, \boldsymbol{x}')\,\mathrm{d}\,\pi(\boldsymbol{x}')$.

Under our assumption that E is Polish, conditional measures always exist [11, Corollary 10.4.10]. For π-almost every $\boldsymbol{x}'$, the measure $\gamma(\cdot, \boldsymbol{x}')$ is unique [11, Lemma 10.4.3] and coincides with the disintegration [26, 452E] of γ that is consistent with the projection p_2.

Finally, throughout this document, we assume that (Riemannian) manifolds are connected and smooth without boundary. These assumptions enable us to apply the results by McCann [43, Proposition 6] and Cordero-Erausquin et al. [19]. In most propositions, we also assume that the manifolds are complete. We always denote by $(M, \mathfrak{g})$ such a manifold, by $d_\mathfrak{g}$ its intrinsic geodesic metric, by $\exp : TM \to M$ the exponential map on its tangent bundle, and by Vol the volume measure on it. Denote by $\mathcal{L}^m$ the Lebesgue measure on $\mathbb{R}^m$. $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$ is the set of natural numbers $\mathbb{N}$ with 0 excluded.

2.2 Construction, existence and uniqueness of Wasserstein barycenters

This subsection covers several fundamental properties of Wasserstein barycenters. We construct them via optimal transport theory and measurable barycenter selection maps. It is crucial to comprehend further features of these maps, as highlighted in Section 2.3.2. Once the construction is explained, we discuss the problem of existence and uniqueness of Wasserstein barycenters. These properties are closely related to the law of large numbers for Wasserstein barycenters.

We begin with the existence of measurable barycenter selection maps.

Lemma 2.6 (Measurable barycenter selection maps)**.** *Let* (E, d) *be a proper space. Given an integer* $n \geq 1$*, let* $\lambda_i > 0, 1 \leq i \leq n$*, be* n *positive real numbers such that* $\sum_{i=1}^n \lambda_i = 1$*. There exists a measurable barycenter selection map* $B : E^n \to E$ *such that* $B(x_1, \ldots, x_n)$ *is a barycenter of* $\sum_{i=1}^n \lambda_i \, \delta_{x_i} \in \mathcal{W}_2(E)$*.*

Proof. As in Lemma 2.3, for a subset $\boldsymbol{A} \subset E^n$, denote by $\mathrm{bary}(\boldsymbol{A}) \subset E$ the set of barycenters of $\sum_{i=1}^n \lambda_i \, \delta_{x_i}$ when $\boldsymbol{x} = (x_1, \ldots, x_n)$ runs through $\boldsymbol{A}$. For the existence of a measurable function $B : E^n \to E$ such that $B(\boldsymbol{x}) \in \mathrm{bary}(\{\boldsymbol{x}\})$, we shall apply the Kuratowski and Ryll-Nardzewski measurable selection theorem (Theorem 2.4). By Lemma 2.3, the set $\Gamma := \{(\boldsymbol{x}, z) \in E^{n+1} \mid z \in \mathrm{bary}(\{\boldsymbol{x}\})\}$ is closed. Let $C \subset E$ be a compact set, then

$$\{\boldsymbol{x} \mid \mathrm{bary}(\{\boldsymbol{x}\}) \cap C \neq \emptyset\} = p_1(\{(\boldsymbol{x}, z) \in E^n \times C \mid (\boldsymbol{x}, z) \in \Gamma\}),$$

where $p_1 : E^n \times C \to E^n$ is the canonical projection map. Since C is compact and Γ is closed, $p_1(\Gamma \cap (E^n \times C))$ is a closed set. As (E, d) is a proper space, any open subset $U \subset E$ is a countable union of compact sets. Indeed, fix a point $z \in E$ and define $C_j := \{x \in E \mid d(x, z) \leq j \text{ and } d(x, E \setminus U) \geq \frac{1}{j}\}$ for $j = 1, 2, \ldots$, then each C_j is compact and $U = \cup_{j \geq 1} C_j$. Therefore, as a union of countably many closed sets, the set

$$\{\boldsymbol{x} \mid \mathrm{bary}(\{\boldsymbol{x}\}) \cap U \neq \emptyset\} = \bigcup_{j \geq 1} \{\boldsymbol{x} \mid \mathrm{bary}(\{\boldsymbol{x}\}) \cap C_j \neq \emptyset\}$$

is measurable. Since $\mathrm{bary}(\{\boldsymbol{x}\})$ is compact for $\boldsymbol{x} \in E^n$ by Lemma 2.3, the assumptions of Theorem 2.4 are satisfied by the map $\boldsymbol{x} \mapsto \mathrm{bary}(\{\boldsymbol{x}\})$. This proves the lemma. $\square$

To construct Wasserstein barycenters of finitely many measures, we first recall the following particular type of multi-marginal optimal transport plans.

Definition 2.7 (Multi-marginal optimal transport plans)**.** Let (E, d) be a proper space. Given an integer $n \geq 2$, let $\lambda_i > 0, 1 \leq i \leq n$, be n positive real numbers such that $\sum_{i=1}^n \lambda_i = 1$ and let $\mu_i \in \mathcal{W}_2(E), 1 \leq i \leq n$, be n probability measures on E. Denote by Π the set of probability measures on E^n with marginals $\mu_1, \ldots, \mu_n$ in this order. We call $\gamma \in \Pi$ a multi-marginal optimal transport plan (of its marginals) if

$$\int_{E^n} \min_{y \in E} \sum_{i=1}^n \lambda_i \, d(y, x_i)^2 \, \mathrm{d}\,\gamma(x_1, \ldots, x_n) = \min_{\theta \in \Pi} \int_{E^n} \min_{y \in E} \sum_{i=1}^n \lambda_i \, d(y, x_i)^2 \, \mathrm{d}\,\theta(x_1, \ldots, x_n). \tag{2}$$

In what follows, the marginal measures μ_i and constants λ_i will be clear from the context. In the proof of Lemma 2.3, it is shown that $\min_{y \in E} \sum_{i=1}^n \lambda_i \, d(x_i, y)^2$ is continuous with respect to $(x_1, \ldots, x_n) \in E^n$. Hence, we can conclude the existence of a multi-marginal optimal transport plan γ in the same way as the classic existence of optimal couplings between two measures [58, Theorem 4.1]. Now we are ready to construct Wasserstein barycenters.

6

Proposition 2.8 (Construction of Wasserstein barycenters of $\sum_{i=1}^{n} \lambda_i \delta_{\mu_i}$). *Let (E, d) be a proper space. Given an integer $n \geq 2$, let $\lambda_i > 0, 1 \leq i \leq n$, be n positive real numbers such that $\sum_{i=1}^{n} \lambda_i = 1$. Let $\mu_1, \ldots, \mu_n \in \mathcal{W}_2(E)$ be n probability measures and let γ be a multi-marginal optimal transport plan of them, i.e., satisfying (2). If $B : E^n \to E$ is a measurable map such that $B(x_1, \ldots, x_n)$ is a barycenter of $\sum_{i=1}^{n} \lambda_i \delta_{x_i}$, then*

1. *$\overline{\mu} := B_\# \gamma$ is a barycenter of $\mathbb{P} := \sum_{i=1}^{n} \lambda_i \delta_{\mu_i}$;*

2. *$(B, p_i)_\# \gamma$ is an optimal transport plan between $\overline{\mu}$ and μ_i, where p_i denotes the canonical projection $(x_1, \ldots, x_n) \in E^n \mapsto x_i \in E$;*

3. *if $X, X_1, \ldots, X_n$ are $n+1$ random variables from a probability space $(\Omega, \mathcal{B}, P)$ to (E, d) with law $\overline{\mu}, \mu_1, \ldots, \mu_n$ such that $\mathbb{E} \, d(X, X_i)^2 = W_2(\overline{\mu}, \mu_i)^2$, i.e., (X, X_i) is an optimal transport coupling between $\overline{\mu}$ and μ_i, then for P-almost every $\omega \in \Omega$, $X(\omega)$ is a barycenter of $\sum_{i=1}^{n} \lambda_i \delta_{X_i(\omega)}$.*

Proof. Given an arbitrary probability measure $\nu \in \mathcal{W}_2(E)$, thanks to the gluing lemma [57, Lemma 7.1], there are $n + 1$ random variables $X, X_1, \ldots X_n$ valued in E with laws $\nu, \mu_1, \ldots \mu_n$ such that $\mathbb{E} \, d(X, X_i)^2 = W_2(\nu, \mu_i)^2$. Since $\mu_i = p_{i\#}\gamma$, we have

$$
\sum_{i=1}^{n} \lambda_i \, W_2(\overline{\mu}, \mu_i)^2 \leq \sum_{i=1}^{n} \int_{E^n} \lambda_i \, d(B(\boldsymbol{x}), x_i)^2 \, \mathrm{d}\, \gamma(\boldsymbol{x}) = \int_{E^n} \min_{y \in E} \sum_{i=1}^{n} \lambda_i \, d(y, x_i)^2 \, \mathrm{d}\, \gamma(\boldsymbol{x})
$$

$$
\leq \mathbb{E} \min_{y \in E} \sum_{i=1}^{n} \lambda_i \, d(y, X_i)^2 \leq \mathbb{E} \sum_{i=1}^{n} \lambda_i \, d(X, X_i)^2
$$

$$
= \sum_{i=1}^{n} \lambda_i \, W_2(\nu, \mu_i)^2,
$$

where we sequentially applied the definitions of $\overline{\mu} = B_\# \gamma$, $W_2(\overline{\mu}, \mu_i)$, γ, and $X, X_1, \ldots X_n$. Since ν is arbitrary, it follows that $\overline{\mu}$ is a Wasserstein barycenter. By setting $\nu = \overline{\mu}$ in the above inequality, we actually obtain an equality. Firstly, this equality implies that $\sum_{i=1}^{n} \lambda_i \, W_2(\overline{\mu}, \mu_i)^2 \leq \sum_{i=1}^{n} \int_{E^n} \lambda_i \, d(B(\boldsymbol{x}), x_i)^2 \, \mathrm{d}\, \gamma(\boldsymbol{x})$ is indeed always an equality, which proves the second statement. Secondly, it also implies that the law of $(X_1, \ldots, X_n)$ is a multi-marginal optimal transport plan and $\min_{y \in E} \sum_{i=1}^{n} \lambda_i \, d(y, X_i(\omega))^2 = \sum_{i=1}^{n} \lambda_i \, d(X(\omega), X_i(\omega))^2$ for P-almost every $\omega \in \Omega$, which proves the third statement. $\square$

The general existence of Wasserstein barycenters was first established in [37]. Recall that finitely supported measures are dense in Wasserstein spaces [58, Theorem 6.18], so the above construction of Wasserstein barycenters together with the law of large numbers for Wasserstein barycenters (Theorem 2.2) implies the following theorem.

Theorem 2.9 (Existence of Wasserstein barycenters). *If (E, d) is a proper space, then any $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(E))$ has a barycenter.*

Note that in the law of large numbers for Wasserstein barycenters (Theorem 2.2), we may need to pass to a subsequence of Wasserstein barycenters $\overline{\mu}_j$ and the limit barycenter $\overline{\mu}$ is not known in advance. Hence, Theorem 2.2 will be enhanced if we can assert the uniqueness of barycenters under some additional assumptions, as follows.

Proposition 2.10 (Uniqueness of Wasserstein barycenters)**.** *Let (E, d) be a proper space. If a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(E))$ gives mass to a Borel subset $\mathcal{A} \subset \mathcal{W}_2(E)$ such that for $\mu \in \mathcal{A}$ and $\nu \in \mathcal{W}_2(E)$, any optimal transport plan between μ and ν is induced by a measurable map T pushing μ forward to ν, i.e., $\nu = T_{\#}\mu$ and $W_2(\mu, \nu)^2 = \int_E d(x, T(x))^2 \, \mathrm{d}\mu$, then $\mathbb{P}$ has a unique barycenter in $\mathcal{W}_2(E)$.*

Proof. The uniqueness follows from the strict convexity of the squared distance function to a point in $\mathcal{W}_2(E)$, as shown by [50, Theorem 7.19] and [34, Theorem 3.1]. We recall the proof for the sake of completeness.

Observe that any convex combination of probability measures in the space $\mathcal{W}_2(E)$ is still a probability measure in it. Fix $\mu \in \mathcal{A}$ and consider the squared Wasserstein distance function $W_2(\mu, \cdot)^2$ with respect to this convex structure. For $\lambda \in [0, 1]$ and two different probability measures $\nu_1, \nu_2 \in \mathcal{W}_2(E)$, by definition of Wasserstein metric we have

$$W_2(\mu, \lambda\nu_1 + (1 - \lambda)\nu_2)^2 \leq \lambda W_2(\mu, \nu_1)^2 + (1 - \lambda)W_2(\mu, \nu_2)^2. \tag{3}$$

By our assumptions, there are two measurable maps $T_1, T_2 : E \to E$ such that $\gamma_1 := (\mathrm{Id} \times T_1)_{\#}\mu$ and $\gamma_2 := (\mathrm{Id} \times T_2)_{\#}\mu$ are optimal transport plans between μ and the two measures ν_1 and ν_2 respectively. We claim that (3) cannot be an equality unless $\lambda = 0$ or $\lambda = 1$. Indeed, if (3) is an equality for some $0 < \lambda < 1$, then by setting $\gamma := \lambda\gamma_1 + (1 - \lambda)\gamma_2$ we have

$$\lambda W_2(\mu, \nu_1)^2 + (1 - \lambda)W_2(\mu, \nu_2)^2 = W_2(\mu, \lambda\nu_1 + (1 - \lambda)\nu_2)^2$$
$$\leq \int_{E \times E} d(x, y)^2 \, \mathrm{d}\gamma(x, y)$$
$$= \lambda W_2(\mu, \nu_1)^2 + (1 - \lambda)W_2(\mu, \nu_2)^2,$$

and thus γ is an optimal plan between μ and $\lambda\nu_1 + (1 - \lambda)\nu_2$. By assumptions, there exists a measurable map $T : E \to E$ such that $\gamma = (\mathrm{Id} \times T)_{\#}\mu$. Denote by $\mathrm{graph}(S) \subset E^2$ the graph of a map $S : E \to E$. Note that if S is a measurable map, then $\mathrm{graph}(S) = \{(x, y) \in E^2 \mid d(S(x), y) = 0\}$ is a Borel subset of E^2. Since $\gamma[\mathrm{graph}(T)] = \lambda\gamma_1[\mathrm{graph}(T)] + (1-\lambda)\gamma_2[\mathrm{graph}(T)] = 1$ and $0 < \lambda < 1$, we have $\gamma_1[\mathrm{graph}(T)] = \gamma_2[\mathrm{graph}(T)] = 1$. Hence, for $i \in \{1, 2\}$, $\mu(\{x \in E \mid T_i(x) = T(x)\}) = \gamma_i[\mathrm{graph}(T) \cap \mathrm{graph}(T_i)] = 1$. It follows that both T_1 and T_2 coincide with T almost everywhere with respect to μ and thus $\gamma_1 = \gamma_2$, which is a contradiction since $\nu_1 \neq \nu_2$.

This shows that $W_2(\mu, \cdot)^2$ is strictly convex on $\mathcal{W}_2(E)$ for $\mu \in \mathcal{A}$. Since $\mathbb{P}(\mathcal{A}) > 0$, the map

$$\nu \in \mathcal{W}_2(E) \mapsto \int_{\mathcal{W}_2(E)} W_2(\mu, \nu)^2 \, \mathrm{d}\mathbb{P}(\mu)$$

is also strictly convex on $\mathcal{W}_2(E)$ by the linearity and positivity of the above integral. It follows that the Wasserstein barycenter of $\mathbb{P}$ asserted by Theorem 2.9 is unique. $\qquad\square$

Remark 2.11. Under the assumptions of Proposition 2.10, the optimal transport plan between $\mu \in \mathcal{A}$ and $\nu \in \mathcal{W}_2(M)$ is unique. By setting $\nu_1 = \nu_2 = \nu$, (3) becomes an equality for any $\lambda \in [0, 1]$. It is shown above that any two optimal transport plans γ_1 and γ_2 between measures μ and ν coincide.

There are many setups in which we can apply Proposition 2.10. We typically choose $\mathcal{A}$ as the set of absolutely continuous measures with respect to some given reference measure. The following lemma ensures that $\mathcal{A}$ is a Borel set of $(\mathcal{W}_2(E), W_2)$.

Lemma 2.12. *Let E be a metric space with a σ-finite Borel measure μ on E. Assume that μ is outer regular, i.e., for any Borel set $N \in \mathcal{B}(E)$, $\mu(N) = \inf\{\mu(O) \mid O$ open neighborhood of $N\}$. Denote by $\mathcal{A}$ the set of probability measures in $\mathcal{W}_2(E)$ that are absolutely continuous with respect to μ. For $\epsilon, \delta > 0$, define the set*

$$\mathcal{E}_{\epsilon,\delta} := \{\nu \in \mathcal{W}_2(E) \mid \forall N \in \mathcal{B}(E), \mu(N) < \delta \implies \nu(N) \leq \epsilon\}.$$

It is a closed set with respect to the weak convergence topology of $\mathcal{W}_2(E)$, and we have

$$\mathcal{A} = \bigcap_{k \in \mathbb{N}} \bigcup_{l \in \mathbb{N}} \mathcal{E}_{2^{-k},2^{-l}}.$$

In particular, if E is a proper space and μ is a locally finite Borel measure, i.e., μ gives finite mass to some open neighborhood of every point in E, then for the Wasserstein space topology, $\mathcal{E}_{\epsilon,\delta}$ is a closed set and $\mathcal{A}$ is a Borel set.

Proof. Our proof is based on [34, Proposition 2.1, Remark 2.2] though we use different assumptions.

Suppose that $\nu_j \in \mathcal{E}_{\epsilon,\delta}$ converges weakly to $\nu \in \mathcal{W}_2(E)$. For any $N \in \mathcal{B}(E)$ such that $\mu(N) < \delta$, there exists an open set O such that $N \subset O$ and $\mu(O) < \delta$ since μ is outer regular. By the characterization of weak convergence of probability measures on metric spaces [11, Corollary 8.2.10], we have

$$\nu(N) \leq \nu(O) \leq \liminf_{j \to \infty} \nu_j(O) \leq \epsilon$$

and thus $\mathcal{E}_{\epsilon,\delta}$ is closed with respect to weak convergence topology of $\mathcal{W}_2(E)$.

The inclusion $\mathcal{A} \supset \bigcap_{k \in \mathbb{N}} \bigcup_{l \in \mathbb{N}} \mathcal{E}_{2^{-k},2^{-l}}$ follows from the definition of a measure ν being absolutely continuous with respect to μ: $\forall N \in \mathcal{B}(E), \mu(N) = 0 \implies \nu(N) = 0$. Fix a measure $\nu \in \mathcal{A}$. Since μ is σ-finite, we can apply the Radon-Nikodym theorem to write $\nu = f \cdot \mu$. The reverse inclusion $\mathcal{A} \subset \bigcap_{k \in \mathbb{N}} \bigcup_{l \in \mathbb{N}} \mathcal{E}_{2^{-k},2^{-l}}$ follows from the absolute continuity of Lebesgue integral [11, Theorem 2.5.7, Proposition 2.6.4].

Given a proper space E and a locally finite Borel measure μ, μ gives finite mass to compact sets, and every open subset of E is σ-compact. It follows that μ is outer regular [54, Theorem 6 of §2.7] and also σ-finite. Since Wasserstein convergence implies weak convergence, the set $\mathcal{E}_{\epsilon,\delta}$ is closed for the Wasserstein metric. It follows that $\mathcal{A}$ is a Borel set of $\mathcal{W}_2(E)$. $\qquad\square$

Remark 2.13. On a metric space, any finite Borel measure is outer regular, see [11, Definition 7.1.5, Theorem 7.1.7] or [10, Theorem 1.1]. However, this is not true for σ-finite Borel measures. For example, define the Borel measure μ on $\mathbb{R}$ such that for $N \in \mathcal{B}(\mathbb{R})$, μ counts the number of rational points in N. This measure is σ-finite but not outer regular since μ never gives finite mass to open sets. As for the assumption regarding the σ-compactness of open sets in the above cited theorem [54, Theorem 6 of §2.7], for metric spaces it can be replaced by assuming that μ gives finite mass to a sequence of open sets $O_i, i \geq 1$ such that $E = \bigcup_{i \geq 1} O_i$. We also mention that there exists a σ-finite and locally finite but not outer regular Borel measure on a locally compact Hausdorff space [12, problem 5 of Exercise §1, INT IV.119].

Thanks to Proposition 2.10 and Lemma 2.12, the Wasserstein barycenter of $\mathbb{P}$ is unique for the following spaces, provided that $\mathbb{P}$ gives mass to the set of absolutely continuous measures with respect to the corresponding canonical reference measure:

1. complete Riemannian manifolds, see Villani [58, Theorem 10.41] or Gigli [28, Theorem 7.4];

2. compact finite dimensional Alexandrov spaces, see Bertrand [9, Theorem 1.1];

3. for $K \in \mathbb{R}$ and $N \geq 1$, non-branching $\mathrm{CD}(K, N)$ spaces, see Gigli [29, Theorem 3.3];

4. for $K \in \mathbb{R}$ and $N \geq 1$, $\mathrm{RCD}^*(K, N)$ spaces, see Gigli, Rajala and Sturm [30, Theorem 1.1];

5. for $K \in \mathbb{R}$ and $N \geq 1$, essentially non-branching $\mathrm{MCP}(K, N)$ spaces, see Cavalletti and Mondino [16, Theorem 1.1];

6. (2-)essentially non-branching spaces with qualitatively non-degenerate reference measures, see Kell [32, Theorem 5.8].

The above spaces are listed in (nearly) ascending order of generality. For the metric measure spaces, we assume that the metric space is proper and the reference measure is locally finite. The references cited above demonstrate that the unique optimal transport plan (Remark 2.11) between an absolutely continuous probability measure and a given probability measure is induced by a measurable map, allowing us to apply Proposition 2.10.

Since the existence and uniqueness of Wasserstein barycenters (under mild assumptions) on Riemannian manifolds M are established, we are ready to prove the absolute continuity of Wasserstein barycenters with respect to Vol. In the next subsection, we consider Wasserstein spaces $\mathcal{W}_2(M)$ over Riemannian manifolds and show that the unique Wasserstein barycenter of finitely many measures is absolutely continuous if one of those measures is so.

2.3 The absolute continuity of Wasserstein barycenters of finitely many measures

Let (M, g) be a complete Riemannian manifold and let $\mathbb{P} = \sum_{i=1}^{n} \lambda_i \, \delta_{\mu_i}$ be a probability measure on $\mathcal{W}_2(M)$ with positive real numbers λ_i and compactly supported measures μ_i in $\mathcal{W}_2(M)$. Assuming that M is compact and μ_1 is absolutely continuous (with respect to Vol), Kim and Pass [34, Theorem 5.1] proved that unique barycenter $\overline{\mu}$ of $\mathbb{P}$ is absolutely continuous. For completeness and also for a rigorous foundation of our later arguments, we provide a proof for general non-compact manifolds.

The proof strategy is to investigate different cases for measures $\mu_i, 2 \leq i \leq n$, step by step. In the simplest case when $\mu_i = \delta_{x_i}, 2 \leq i \leq n$, are Dirac measures, the unique barycenter $\overline{\mu}$ is the push-forward of $\mu_1 \otimes \delta_{x_2} \otimes \ldots \otimes \delta_{x_n}$ by a measurable barycenter selection map B (Proposition 2.8). To deduce more properties of B, we shortly review c-concave functions.

2.3.1 c-concave functions

For $x, y \in M$, we define the function $c(x, y) := \frac{1}{2} d_{\mathsf{g}}(x, y)^2$ as the half of the squared distance between x and y in M, and define $d_y^2(\cdot) := d_{\mathsf{g}}(\cdot, y)^2$ to avoid ambiguity when fixing the point y.

Definition 2.14 (c-transforms and c-concave functions)**.** Let (M, g) be a Riemannian manifold. Let X and Y be two non-empty compact subsets of M. A function $\phi : X \to \mathbb{R}$ is c-concave if there exists a function $\psi : Y \to \mathbb{R}$ such that

$$\phi(x) = \inf_{y \in Y} c(x, y) - \psi(y), \quad \forall x \in X. \tag{4}$$

We write it as $\phi = \psi^c$ and call ϕ the c-transform of ψ. The set of all c-concave functions with respect to X and Y is denoted by $\mathcal{I}^c(X, Y)$.

The significance of c-concave functions in optimal transport theory is highlighted by the following theorem of McCann [43], which extends Brenier's seminal theorem [57, Theorem 2.12] to Riemannian manifolds. Recall that given a c-concave function ϕ on a compact set $\overline{\mathcal{X}}$ with $\mathcal{X} \subset M$ open, its gradient $\nabla\phi$ exists on $\mathcal{X}$ almost everywhere with respect to Vol since ϕ is Lipschitz [43, Lemma 4].

Theorem 2.15 (Optimal transport on manifolds, [19, Theorem 3.2])**.** *Let (M, g) be a complete Riemannian manifold. Fix two measures $\mu, \nu \in \mathcal{W}_2(M)$ with compact support such that μ is absolutely continuous. Given two bounded open subsets $\mathcal{X}, \mathcal{Y} \subset M$ containing the supports of μ and ν respectively, there exists $\phi \in \mathcal{I}^c(\overline{\mathcal{X}}, \overline{\mathcal{Y}})$ such that $(\mathrm{Id}, F)_{\#}\mu$ is the unique optimal transport plan between μ and ν, where the function $F := \exp(-\nabla\phi)$ is μ-almost everywhere well-defined.*

The following lemma shows that the definition of barycenters for measures $\sum_{i=1}^n \lambda_i \delta_{x_i}$ on M involves c-concave functions.

Lemma 2.16. *Let (M, g) be a complete Riemannian manifold. Given an integer $n \geq 2$, let $\lambda_i > 0, 1 \leq i \leq n$, be n positive real numbers such that $\sum_{i=1}^n \lambda_i = 1$. We define*

$$f : (x_1, x_2, \ldots, x_n) \in M^n \mapsto \min_{w \in M} \sum_{i=1}^n \lambda_i \, c(w, x_i) = \frac{1}{2} \min_{w \in M} \sum_{i=1}^n \lambda_i \, d_{\mathsf{g}}(w, x_i)^2. \tag{5}$$

Fix a non-empty compact subset $X \subset M$ and $n - 1$ points $x_i \in M, 2 \leq i \leq n$. Denote by Y the set of all barycenters of $\sum_{i=1}^n \lambda_i \delta_{x_i}$ when x_1 runs through X. Define $f_1 : x_1 \in X \mapsto f(x_1, \ldots, x_n)/\lambda_1$ and $g_1 : y \in Y \mapsto -1/\lambda_1 \sum_{i=2}^n \lambda_i \, c(y, x_i)$, then $f_1 = g_1^c \in \mathcal{I}^c(X, Y)$ and $g_1 = f_1^c \in \mathcal{I}^c(Y, X)$.

Proof. The set $Y \subset M$ is compact by Lemma 2.3. Using the given definition of Y, we can replace the minimum over M in (5) by the minimum over Y, which shows the equality $f_1 = g_1^c \in \mathcal{I}^c(X, Y)$.

Since $f_1(x) + g_1(y) \leq c(x, y)$ for any $(x, y) \in X \times Y$, we have

$$g_1(y) \leq f_1^c(y) := \inf_{x \in X} c(x, y) - f_1(x). \tag{6}$$

Fix an arbitrary point $y \in Y$. Our definition of Y implies the existence of $x_1 \in X$ such that y is a barycenter of $\sum_{i=1}^n \lambda_i \delta_{x_i}$. For such a pair $(x_1, y) \in X \times Y$, $f_1(x_1) + g_1(y) = c(x_1, y)$ by the definitions of f_1 and g_1. It follows from the inequalities $f_1(x_1) + f_1^c(y) \leq c(x_1, y) = f_1(x_1) + g_1(y)$ and (6) that $g_1(y) = f_1^c(y)$. Since y is arbitrarily chosen, we conclude that $g_1 = f_1^c \in \mathcal{I}^c(Y, X)$. $\square$

The c-concave function $g_1 \in \mathcal{I}^c(Y, X)$ defined in Lemma 2.16 has simple expression unlike its c-transform f_1. Furthermore, thanks to the following lemma by Kim and Pass [33, Lemma 3.1], we conclude that g_1 is smooth. This differential property of g_1 (to be used in Lemma 2.20) is crucial to prove the absolute continuity of Wasserstein barycenters.

Lemma 2.17 (Barycenters and cut loci, [33, Lemma 3.1 and proof of Theorem 6.1])**.** *Let (M, g) be a complete Riemannian manifold. Given an integer $n \geq 1$, let $\lambda_i > 0, 1 \leq i \leq n$, be n positive real numbers such that $\sum_{i=1}^n \lambda_i = 1$ and let $x_i \in M, 1 \leq i \leq n$, be n points of M. For $1 \leq i \leq n$, x_i is out of the cut locus of any barycenters of $\sum_{i=1}^n \lambda_i \delta_{x_i}$.*

It will be shown in Lemma 2.20 that the map $\exp(-\nabla g_1)$ is an optimal transport map F as stated in Theorem 2.15. Given the above Lemma 2.17, the following lemma further illustrates how to differentiate such maps.

Lemma 2.18. *Let $(M, \mathfrak{g})$ be a complete Riemannian manifold. Fix an open set $U \subset M$, a point $x \in U$, and a $\mathcal{C}^2$ smooth function ϕ defined on U. Define $F := \exp(-\nabla\phi)$ on U. Assume that the (fixed) point $y := F(x)$ is out of the cut locus of x. If the two functions, ϕ and $d_y^2/2$, have the same gradient at x, then*

$$\mathrm{D}_x F = [\mathrm{D}_{-\nabla\phi(x)}\exp_x] \circ (\mathrm{Hess}_x\, d_y^2/2 - \mathrm{Hess}_x\, \phi). \tag{7}$$

In the above formula,

1. *Hess_x denotes the Hessian operator at x and its values are maps from $T_x M$ to $T_x M$;*

2. *$\mathrm{D}_{-\nabla\phi(x)}\exp_x : T_{-\nabla\phi(x)}T_x M \to T_y M$ denotes the differential of the exponential map $\exp_x : T_x M \to M$ at $-\nabla\phi(x)$;*

3. *the composition is valid since $T_{-\nabla\phi(x)}T_x M$ can be canonically identified with $T_x M$.*

Proof. The formula (7) is already proven in [19, Proposition 4.1], whose proof can be simplified thanks to our assumptions. Define $y := F(x)$. By the assumption that y is not in the cut locus of x, $\mathrm{Hess}_x\, d_y^2/2$ is well-defined. Shrink the neighborhood U of x if necessary so that for $(w, z) \in U \times U$, w is not in the cut loci of y and z [49, (2) of Proposition 4.1 in Chapter III]. Define the following function g on $U \times U$,

$$g(w, z) := \exp_w\left(-\nabla d_y^2(w)/2 + \Pi_{z \to w}\left[\nabla d_y^2(z)/2 - \nabla\phi(z)\right]\right),$$

where $\Pi_{z \to w} : T_z M \to T_w M$ denotes the parallel transport of tangent vectors along the minimal geodesic from z to w. For $z \in U$, since $\Pi_{z \to z}$ is the identity map on $T_z M$, $g(z, z) = F(z)$. For $w \in U$, $g(w, x) = \exp_w(-\nabla d_y^2(w)/2) \equiv y$ is a constant, where we used the assumption $\nabla d_y^2(x)/2 = \nabla\phi(x)$ for the first equality and used that w is not in the cut locus of y for the second one. It follows that

$$\mathrm{D}_x F = \partial_w g(x, x) + \partial_z g(x, x) = \partial_z g(x, x) \tag{8}$$

$$= [\mathrm{D}_{-\nabla\phi(x)}\exp_x] \circ (\mathrm{Hess}_x\, d_y^2/2 - \mathrm{Hess}_x\, \phi), \tag{9}$$

where we used $F(z) = g(z, z)$, the chain rule and $g(w, x) \equiv y$ in the line (8), and applied the relation between covariant derivative and parallel transport [39, Corollary 4.35] to the definition of Hessian as the covariant derivative of gradients [47, Proposition 2.2.6] in the line (9). $\square$

2.3.2 Lipschitz continuous optimal transport maps of Wasserstein barycenters

To better illustrate our approach towards the absolute continuity of Wasserstein barycenters of finitely many measures, we recall the following result corresponding to the case of two measures.

Proposition 2.19 (Regularity of displacement interpolations, [58, Theorem 8.5, Theorem 8.7])**.** *Let $(M, \mathfrak{g})$ be a complete Riemannian manifold. Let $t \in [0, 1] \mapsto \mu_t \in \mathcal{W}_2(M)$ be a minimal geodesic in the Wasserstein space $\mathcal{W}_2(M)$ such that both μ_0 and μ_1 have compact support. For any $0 < \lambda < 1$, μ_λ is the barycenter of $(1 - \lambda)\delta_{\mu_0} + \lambda\delta_{\mu_1}$. The optimal transport map from μ_λ to μ_0 is Lipschitz continuous, and it follows that μ_λ is absolutely continuous provided that μ_0 is absolutely continuous.*

The Lipschitz continuity in Proposition 2.19 can be shown as a consequence of Mather's shortening lemma [58, Chapter 8]. Since Lipschitz maps send Lebesgue negligible sets to Lebesgue negligible sets, the last statement on absolute continuity follows. Another approach to the Lipschitz continuity is given by Bernard and Buffoni [8] using the Hamilton-Jacobi equation, which is

12

generalized to non-compact settings by Fathi and Figalli [22]. For the case when both μ_0 and μ_1 are absolutely continuous measures on Euclidean spaces, McCann [42, Proposition 1.3] presented a concise proof of the Lipschitz continuity. See relevant references in Villani [58, Bibliographical notes of Chapter 8]. The goal of this subsection is to generalize Proposition 2.19.

We deduce the following Lipschitz continuity from the c-concave functions defined in Lemma 2.16, which are related to barycenter selection maps and thus barycenters of $\lambda_1\,\delta_{\mu_1} + \sum_{i=2}^{n}\lambda_i\,\delta_{\delta_{x_i}}$. Recall that a measurable barycenter selection map $B : M^n \to M$ (Theorem 2.4) sends $(x_1,\dots,x_n) \in M^n$ to a barycenter of $\sum_{i=1}^{n}\lambda_i\,\delta_{x_i}$. In the following propositions, the constant $\lambda_i, 1 \le i \le n$ for B are given in the context.

Lemma 2.20 (Lipschitz continuous maps $F = \exp(-\nabla g_1)$). *Let $(M, \mathfrak{g})$ be a complete Riemannian manifold. Given an integer $n \ge 2$, let $\lambda_i > 0, 1 \le i \le n$, be n positive real numbers such that $\sum_{i=1}^{n}\lambda_i = 1$. Fix a non-empty compact subset $X \subset M$ and a point $\boldsymbol{x}' = (x_2,\dots,x_n) \in M^{n-1}$. Denote by Y the compact set of all barycenters of $\sum_{i=1}^{n}\lambda_i\,\delta_{x_i}$ when x_1 runs through X. Define the function $g_1 : y \in M \mapsto -1/\lambda_1 \sum_{i=2}^{n}\lambda_i\, c(y,x_i)$. It is smooth in a neighborhood of Y and thus $F := \exp(-\nabla g_1) : Y \to M$ is a well-defined Lipschitz continuous function. We have $F(Y) = X$ and the following characterization of F:*

$$z \in Y \text{ and } x_1 = F(z) \iff x_1 \in X \text{ and } z \text{ is a barycenter of } \sum_{i=1}^{n}\lambda_i\,\delta_{x_i}. \tag{10}$$

Given a measure $\mu_1 \in \mathcal{W}_2(M)$ with support X and a measurable barycenter selection map $B : M^n \to M$, $\overline{\mu} := B_{\#}(\mu_1 \otimes \delta_{x_2} \otimes \cdots \otimes \delta_{x_n})$ is a barycenter of $\lambda_1\,\delta_{\mu_1} + \sum_{i=2}^{n}\lambda_i\,\delta_{\delta_{x_i}}$ and $(\mathrm{Id}, F)_{\#}\overline{\mu}$ is an optimal transport plan between $\overline{\mu}$ and μ_1.

Proof. Lemma 2.17 implies the differential property of g_1 and thus the Lipschitz continuity of F. Since g_1 restricted to Y is a c-concave function (Lemma 2.16) and ∇g_1 exists on Y, by defining $g_1^c : x \in X \mapsto \min_{y \in Y}\{c(x,y) - g_1(y)\}$, a well-known property of c-concave functions proven by McCann [43, Lemma 7] shows that

$z \in Y$ and $x_1 = \exp(-\nabla g_1)(z) = F(z) \iff (x_1, z) \in X \times Y$ and $g_1^c(x_1) + g_1(z) = c(x_1, z)$.

Note that though McCann's lemma is proven for compact manifolds, the arguments of its proof only depend on the existence of gradient ∇g_1 and the compactness of X and Y. For $x_1 \in X$, we have $g_1^c(x_1) = 1/\lambda_1 \inf_{w \in M} \sum_{i=1}^{n}\lambda_i\, c(w, x_i)^2$ (Lemma 2.16) and thus

$$z \in Y \text{ and } g_1^c(x_1) + g_1(z) = c(x_1, z) \iff \sum_{i=1}^{n}\lambda_i\, d_{\mathfrak{g}}(z, x_i)^2 = \inf_{w \in M}\sum_{i=1}^{n}\lambda_i\, d_{\mathfrak{g}}(w, x_i)^2,$$

which implies the characterization (10). $F(Y) = X$ follows from (10) and the definition of Y.

Since $\gamma := \mu_1 \otimes \delta_{x_2} \otimes \cdots \otimes \delta_{x_n}$ is the only measure on M^n with marginals $\mu_1, \delta_{x_2}, \dots, \delta_{x_n}$ in this order, it is the (unique) multi-marginal optimal transport plan of its marginals. Proposition 2.8 shows that $\overline{\mu} = B_{\#}\gamma$ is a Wasserstein barycenter. Denote by $p_1 : M \times M^{n-1} \to M$ the canonical projection map. Since $p_1(x_1, \boldsymbol{x}') = x_1 = F(B(x_1, \boldsymbol{x}'))$ for $x_1 \in X$ by the characterization (10), Proposition 2.8 shows that $(B, p_1)_{\#}\gamma = (B, F \circ B)_{\#}\gamma = (\mathrm{Id}, F)_{\#}\overline{\mu}$ is an optimal transport plan between $\overline{\mu}$ and μ_1. $\qquad\square$

Lemma 2.20 implies that any barycenter selection map on $X \times \{\boldsymbol{x}'\}$ is injective. The following lemma by Kim and Pass [33, Lemma 3.5] generalizes this injectivity, and it will help us to generalize Lemma 2.20.

Lemma 2.21. *Let (M, g) be a complete Riemannian manifold. Given an integer $n \geq 2$, let $\lambda_i > 0, 1 \leq i \leq n$, be n positive real numbers such that $\sum_{i=1}^n \lambda_i = 1$ and let $\mu_i \in \mathcal{W}_2(M), 1 \leq i \leq n$, be n measures with compact support. If γ is a multi-marginal optimal transport plan with marginals $\mu_1, \ldots, \mu_n$, then*

$$\boldsymbol{x}, \boldsymbol{y} \in \text{supp}(\gamma), \quad \boldsymbol{x} \neq \boldsymbol{y} \implies \text{bary}(\{\boldsymbol{x}\}) \cap \text{bary}(\{\boldsymbol{y}\}) = \emptyset,$$

where $\text{bary}(\{x_1, \ldots, x_n\})$ *is the set of barycenters of* $\sum_{i=1}^n \lambda_i \, \delta_{x_i}$.

To avoid being lengthy, we skip the proof of above lemma [33, Lemma 3.5], which is based on c-cyclical monotonicity and Lemma 2.17. Though the proof in the given reference is for the case when $\lambda_1 = \cdots = \lambda_n = 1/n$, there is no essential difficulty to apply it to the stated case [33, proof of Theorem 6.1]. The following proposition constructs an optimal transport map from $\overline{\mu} := B_\# \gamma$ to μ_1 when $\mu_i, 2 \leq i \leq n$ are discrete measures and thus generalizes Lemma 2.20. The optimal transport map may fail to be a Lipschitz map, but it is a disjoint union of Lipschitz maps. Recall that given (at most) countably many disjoint subsets $Y_j \subset M, j \in J \subset \mathbb{N}$ with functions $F_j : Y_j \to M$, the *disjoint union F of $F_j, j \in J$* is the function defined on $\cup_{j \in J} F_j$ such that $F|_{Y_j} = F_j$. We shall use conditional measures (Definition 2.5) to deduce further conclusions from F_j's Lipschitz continuity.

Proposition 2.22. *Let (M, g) be an m-dimensional complete Riemannian manifold. Given an integer $n \geq 2$, let $\lambda_i > 0, 1 \leq i \leq n$, be n positive real numbers such that $\sum_{i=1}^n \lambda_i = 1$. Let $\mu_1 \in \mathcal{W}_2(M)$ be a measure with compact support and let $\mu_i \in \mathcal{W}_2(M), 2 \leq i \leq n$, be $n - 1$ discrete measures, i.e., measures concentrated on at most countably many points. Given a multi-marginal optimal transport plan γ of $\mu_1, \ldots, \mu_n$ in this order and a measurable barycenter selection map $B : M^n \to M$, the measure $\overline{\mu} := B_\# \gamma$ is a barycenter of $\sum_{i=1}^n \lambda_i \, \delta_{\mu_i}$. This barycenter $\overline{\mu}$ assigns full mass to a disjoint union of at most countably many compact sets, and on each of them Lemma 2.20 defines a Lipschitz continuous map with compact subset $X \subset M$ and point $\boldsymbol{x}' \in M^{n-1}$ such that $X \times \{\boldsymbol{x}'\}$ is contained in the support of γ. Denote by F the disjoint union of the Lipschitz maps. $(\text{Id}, F)_\# \overline{\mu}$ is an optimal transport plan between $\overline{\mu}$ and μ_1.*

For positive real numbers $\delta, \epsilon > 0$, we define the set

$$\mathcal{E}_{\epsilon, \delta} := \{\mu \in \mathcal{W}_2(M) \mid \forall N \in \mathcal{B}(M), \text{Vol}(N) < \delta \implies \mu(N) \leq \epsilon\}.$$

If there is a common Lipschitz constant C of the Lipschitz maps, then $\mu_1 \in \mathcal{E}_{\epsilon, \delta} \implies \overline{\mu} \in \mathcal{E}_{\epsilon, \delta/C^m}$.

Proof. Proposition 2.8 shows that $\overline{\mu}$ is a Wasserstein barycenter. Let us reveal more details of γ. Denote by p_1 and p_2 the canonical projections sending $\boldsymbol{x} = (x_1, \boldsymbol{x}') \in M \times M^{n-1}$ to $x_1 \in M$ and $\boldsymbol{x}' \in M^{n-1}$ respectively. The measure $\pi := p_{2\#} \gamma$ on M^{n-1} is discrete since its marginals $\mu_2, \ldots, \mu_n$ are so. Denote by $\{\boldsymbol{x}'_j\}_{j \in J} \subset M^{n-1}$ ($J \subset \mathbb{N}$) the set of all atoms of π. For each $j \in J$, we introduce the following definitions. Define $\pi_j := \pi(\{\boldsymbol{x}'_j\}) > 0$ and define $X_j := p_1(\text{supp}\, \gamma \cap (M \times \{\boldsymbol{x}'_j\}))$. Applying Lemma 2.20 to X_j and $\boldsymbol{x}'_j \in M^{n-1}$, we obtain a compact set Y_j and a Lipschitz continuous map $F_j : Y_j \to M$ such that $F_j(Y_j) = X_j$. Since π assigns full mass to the union $\bigcup_{j \in J} \{\boldsymbol{x}_j\}$, γ assigns full mass to the union $\bigcup_{j \in J} X_j \times \{\boldsymbol{x}'_j\}$. As in Lemma 2.21, we denote by $\text{bary}(\{(x_1, \ldots, x_n)\})$ the set of barycenters of $\sum_{l=1}^n \lambda_l \, \delta_{x_l}$.

We claim that $Y_i \cap Y_k = \emptyset$ for two different indices $i, k \in J$. Indeed, if $z \in Y_i \cap Y_k$ for $i, k \in J$, then by the characterization of F_i, F_k in Lemma 2.20, $z \in \text{bary}(\{(F_i(z), \boldsymbol{x}'_i)\}) \cap \text{bary}(\{(F_k(z), \boldsymbol{x}'_k)\})$. Since $\text{supp}\, \gamma = \bigcup_{j \in J} X_j \times \{\boldsymbol{x}'_j\}$ and $F_j(Y_j) = X_j$, Lemma 2.21 forces that $\boldsymbol{x}'_i = \boldsymbol{x}'_k$ and thus $i = k$. Define F as the disjoint union of $F_j, j \in J$, i.e., $F|_{Y_j} = F_j$. Since $p_1(x, \boldsymbol{x}'_j) = x = F(B(x, \boldsymbol{x}'_j))$ for

14

$x \in X_j$, Proposition 2.8 implies that $(B, p_1)_{\#}\gamma = (B, F \circ B)_{\#}\gamma = (\mathrm{Id}, F)_{\#}\overline{\mu}$ is an optimal transport plan between $\overline{\mu}$ and μ_1. Since $\bigcup_{j \in J} Y_j$ is the domain of F and $F_{\#}\overline{\mu} = \mu_1$, $\overline{\mu}$ assigns full mass to a union of at most countably many compact sets that satisfies our description.

We claim that $\mu_1(X_i \cap X_k) = 0$ for two different indices $i, k \in J$. Consider the conditional measure such that $\mathrm{d}\,\gamma(\boldsymbol{x}) = \gamma(\mathrm{d}\,\boldsymbol{x}, \boldsymbol{x}')\,\mathrm{d}\,\pi(\boldsymbol{x}')$. For $j \in J$, define $\nu_j := \frac{1}{\pi_j}\mu_1|_{X_j}$ and $\overline{\nu}_j := B_{\#}\gamma(\cdot, \boldsymbol{x}'_j)$. Note that for $j \in J$ and $\boldsymbol{R} \in \mathcal{B}(M^n)$, $\gamma[\boldsymbol{R} \cap (M \times \{\boldsymbol{x}'_j\})] = \gamma(\boldsymbol{R}, \boldsymbol{x}'_j)\,\pi_j$ by Definition 2.5, so $\gamma(\cdot, \boldsymbol{x}'_j)$ is concentrated on $X_j \times \{\boldsymbol{x}'_j\}$. Since γ assigns full mass to the union $\bigcup_{j \in J} X_j \times \{\boldsymbol{x}'_j\}$, we obtain the following equality by choosing $\boldsymbol{R}$ of the form $A \times M^{n-1}$ with $A \in \mathcal{B}(M)$,

$$\gamma(A \times M^{n-1}, \boldsymbol{x}'_j) = \frac{1}{\pi_j}\gamma[A \times \{\boldsymbol{x}'_j\}] = \frac{1}{\pi_j}\gamma[(A \cap X_j) \times M^{n-1}] = \frac{1}{\pi_j}\mu_1(A \cap X_j),$$

which implies that the first marginal of $\gamma(\cdot, \boldsymbol{x}'_j)$ is ν_j as A is arbitrarily chosen. Furthermore, for a measurable map $f : M^n \to M$,

$$\forall N \in \mathcal{B}(M), \quad [f_{\#}\gamma](N) = \gamma(f^{-1}(N)) = \sum_{j \in J}\gamma(f^{-1}(N), \boldsymbol{x}'_j)\,\pi_j = \sum_{j \in J}[f_{\#}\gamma(\cdot, \boldsymbol{x}'_j)](N)\,\pi_j.$$

Taking $f = p_1$ and $f = B$, we obtain $\mu_1 = \sum_{j \in J}\pi_j\,\nu_j$ and $\overline{\mu} = \sum_{j \in J}\pi_j\,\overline{\nu}_j$. Hence, given $i \in J$, $\mu_1(X_i) = \sum_{j \in J}\mu_1|_{X_j}(X_i)$ and thus $\mu_1(X_i \cap X_k) = 0$ for $k \in J$ different from i.

Assume that C is a common Lipschitz constant of all $F_j, j \in J$. For any Borel set $N \in \mathcal{B}(M)$, there exist Borel sets $W_j, j \in J$ such that $F_j(N \cap Y_j) \subset W_j \subset X_j$ and $\mathrm{Vol}(W_j) \le C^m\,\mathrm{Vol}(N \cap Y_j)$ [55, Proposition 12.6, Proposition 12.12, Remark after Proposition 12.12] (c.f. [58, Proof of Theorem 8.7]). For $j \in J$, since $\gamma(\cdot, \boldsymbol{x}'_j)$ is the product measure of its marginals, Lemma 2.20 shows that $F_{j\#}\overline{\nu}_j = \nu_j$. It follows that $\overline{\nu}_j(N \cap Y_j) \le \overline{\nu}_j(F_j^{-1}(W_j)) \le \nu_j(W_j)$ for $j \in J$ and thus

$$\overline{\mu}(N) = \sum_{j \in J}\pi_j\,\overline{\nu}_j(N \cap Y_j) \le \sum_{j \in J}\pi_j\frac{1}{\pi_j}\mu_1|_{X_j}(W_j) = \sum_{j \in J}\mu_1(W_j) = \mu_1(\bigcup_{j \in J}W_j), \tag{11}$$

where we used $W_j \subset X_j$ and $\mu_1(X_i \cap X_k) = 0$ if $i \ne k \in J$. Since $Y_j, j \in J$ are disjoint, $\mathrm{Vol}(\bigcup_{j \in J}W_j) \le C^m\sum_{j \in J}\mathrm{Vol}(N \cap Y_j) \le C^m\,\mathrm{Vol}(N)$. Assuming that $\mu_1 \in \mathcal{E}_{\epsilon,\delta}$, then for any $N \in \mathcal{B}(M)$ with $\mathrm{Vol}(N) < \delta/C^m$, we have $\mathrm{Vol}(\bigcup_{j \in J}W_j) < \delta$ and thus $\mu_{\mathbb{P}}(N) \le \mu_1(\bigcup_{j \in J}W_j) \le \epsilon$ by (11). Therefore, the asserted implication $\mu_1 \in \mathcal{E}_{\epsilon,\delta} \implies \mu_{\mathbb{P}} \in \mathcal{E}_{\epsilon,\delta/C^m}$ is proven. $\qquad\square$

Remark 2.23. Figuratively speaking, the sets $X_j, j \in J$ create a tiling of the support of μ_1 and the points $\boldsymbol{x}'_j, j \in J$ pull them apart (via barycenter selection maps) into disjoint sets $Y_j, j \in J$, which contain different pieces of the support of $\overline{\mu}$ separately.

2.3.3 Proof of absolute continuity

Consider the probability measure $\mathbb{P} = \sum_{i=1}^{n}\lambda_i\,\delta_{\mu_i}$ with positive real numbers λ_i and compactly supported measures $\mu_i \in \mathcal{W}_2(M)$. We can approximate each μ_i for $2 \le i \le n$ with discrete measures to apply Proposition 2.22. If μ_1 is absolutely continuous, then $\mathbb{P}$ has a unique barycenter $\overline{\mu}$, which is approximated by the barycenters of the approximating sequence (Theorem 2.2). Recall that the sets $\mathcal{E}_{\epsilon,\delta}$ (Lemma 2.12) fully characterize absolutely continuous measures and are closed with respect to weak convergence. Hence, to prove the absolute continuity of $\overline{\mu}$, it remains to find a common Lipschitz constant C for F defined as in Lemma 2.20 valid for any element of the whole

15

approximating sequence. Then we get the result thanks to Proposition 2.22. Note that the domain Y of F is at least varying along the approximating sequence, so the existence of C is not simply a direct consequence of compactness. More precisely, we shall prove:

Theorem 2.24 (Absolute continuity of the barycenter of $\sum_{i=1}^{n} \lambda_i \, \delta_{\mu_i}$)**.** *Let* (M, g) *be a complete Riemannian manifold. Given an integer* $n \geq 2$*, let* $\lambda_i > 0, 1 \leq i \leq n$*, be* n *positive real numbers such that* $\sum_{i=1}^{n} \lambda_i = 1$ *and let* $\mu_i \in \mathcal{W}_2(M), 1 \leq i \leq n$*, be* n *probability measures with compact support. If* μ_1 *is absolutely continuous, then the unique barycenter* $\overline{\mu}$ *of* $\sum_{i=1}^{n} \lambda_i \, \delta_{\mu_i}$ *is absolutely continuous with compact support.*

Proof. The uniqueness and compact support of $\overline{\mu}$ follow from Section 2.2 and Lemma 2.3. We approximate each μ_i for $2 \leq i \leq n$ in $(\mathcal{W}_2(M), W_2)$ by a sequence of discrete measures $\{\mu_i^j\}_{j \geq 1}$ whose supports are contained in the compact support of μ_i. Then $\mathbb{P}_j := \lambda_1 \delta_{\mu_1} + \sum_{i=2}^{n} \lambda_i \, \delta_{\mu_i^j}$ converges to $\mathbb{P}$ in $\mathcal{W}_2(\mathcal{W}_2(M))$. By the law of large numbers for Wasserstein barycenters (Theorem 2.2), the unique barycenter $\overline{\mu}_j$ of $\mathbb{P}_j$ converges in $(\mathcal{W}_2(M), W_2)$ to the unique barycenter $\overline{\mu}$ of $\mathbb{P}$.

Denote by γ_j a multi-marginal optimal transport plan of marginal measures $\mu_1, \mu_2^j, \ldots, \mu_n^j$ in this order. Fix an index j, a non-empty compact subset $X \subset M$ and a point $\boldsymbol{x}' := (x_2, \ldots, x_n) \in M^{n-1}$ such that $X \times \{\boldsymbol{x}'\} \subset \operatorname{supp} \gamma_j$. Applying Lemma 2.20 to X and $\boldsymbol{x}'$, we obtain a Lipschitz continuous function $F = \exp(-\nabla g_1)$ on a compact set Y. We claim that there exists a Lipschitz constant C of F on Y independent of j, X and $\boldsymbol{x}'$. Recall that $g_1(y) := -1/\lambda_1 \sum_{i=2}^{n} \lambda_i \, c(y, x_i)$ is smooth in a neighborhood of Y. Given $z \in Y$, since z is a barycenter of $\sum_{i=1}^{n} \lambda_i \, \delta_{x_i}$ (Lemma 2.20) with $x_1 := F(z)$, we have $\sum_{i=1}^{n} \nabla d_{x_i}^2(z) = 0$ thanks to Lemma 2.17 and thus $\nabla d_{x_1}^2 / 2(z) = \nabla g_1(z)$. Moreover, Lemma 2.17 enables us to apply Lemma 2.18, which implies

$$\mathrm{D}_z F = \mathrm{D}_z \exp(-\nabla g_1) = [\mathrm{D}_{-\nabla g_1(z)} \exp_z] \circ (\operatorname{Hess}_z d_{x_1}^2 / 2 - \operatorname{Hess}_z g_1)$$

$$= [\mathrm{D}_{-\nabla g_1(z)} \exp_z] \circ \frac{1}{2\lambda_1} \sum_{i=1}^{n} \lambda_i \operatorname{Hess}_z d_{x_i}^2. \tag{12}$$

In (12), $\sum_{i=1}^{n} \lambda_i \operatorname{Hess}_z d_{x_i}^2$ is positive semi-definite since z reaches the global minimum of $w \in M \mapsto \sum_{i=1}^{n} \lambda_i \, d_{\mathsf{g}}(w, x_i)^2$. We now bound (12) using compactness as follows. By Lemma 2.3 and our construction of $\mathbb{P}_j$, the union of the supports of $\overline{\mu}, \mu_i, \overline{\mu}_j$ and μ_i^j for $1 \leq i \leq n$ and $j \geq 1$ is compact. Hence, independent of z, j and $\boldsymbol{x}'$, $\mathrm{D}_{-\nabla g_1(z)} \exp_z$ is uniformly bounded (in norm) and $\sum_{i=1}^{n} \lambda_i \operatorname{Hess}_z d_{x_i}^2$ is uniformly bounded from above by the Rauch comparison theorem for Hessians of distance functions, which is applicable here and provides a constant upper bound thanks to the compactness, see [19, Lemma 3.12 and Corollary 3.13] or [47, Theorem 6.4.3]. This shows the existence of the claimed Lipschitz constant C. We remark that the absolute continuity of μ_1 is not needed for the existence of C.

Applying Proposition 2.22 to measures $\mu_1, \mu_2^j, \ldots, \mu_n^j$, we have for $\epsilon, \delta > 0$, $\mu_1 \in \mathcal{E}_{\epsilon, \delta} \implies \overline{\mu}_j \in \mathcal{E}_{\epsilon, \delta/C^m}$ since $\overline{\mu}_j$ is the unique barycenter of $\mathbb{P}_j$. As $\overline{\mu}_j$ converges to $\overline{\mu}$ weakly, Lemma 2.12 shows that all measures $\overline{\mu}_j$ for $j \geq 1$ and $\overline{\mu}$ are absolutely continuous since μ_1 is so. $\qquad\square$

3 Hessian equality for Wasserstein barycenters

In this section, we prove the Hessian equality for Wasserstein barycenters of finitely many measures (Theorem 3.13). A similar property is named as the *second order balance* (inequality) by Kim and Pass [34, Theorem 4.4], but being an equality instead of an inequality is crucial for our proof

of Proposition 4.2. Let us take a special case to illustrate this equality. Consider the reduced case in Lemma 2.20. Namely, take n positive numbers $\lambda_i > 0$ such that $\sum_{i=1}^n \lambda_i = 1$ and denote by $\overline{\mu}$ the barycenter of $\sum_{i=1}^n \lambda_i \delta_{\mu_i}$, where μ_1 is absolutely continuous with compact support and $\mu_i = \delta_{x_i}, 2 \le i \le n$, are Dirac measures. Let us set $\phi_1(z) := g_1(z) := -1/\lambda_1 \sum_{i=2}^n \lambda_i c(z, x_i)$ and $\phi_i(z) := c(z, x_i), 2 \le i \le n$. Thanks to Lemma 2.3 and Lemma 2.17, if z is in the support of $\overline{\mu}$, then z is not in the cut locus of any x_i, which implies $\exp(-\nabla\phi_i)_{\#}\overline{\mu} = \mu_i$ for $2 \le i \le n$. Besides, by definition of the ϕ_i's, $\sum_{i=1}^n \lambda_i \phi_i \equiv 0$; therefore $\sum_{i=1}^n \lambda_i \nabla\phi_i(z) = 0$. Consequently we get $\sum_{i=1}^n \lambda_i \operatorname{Hess}_z \phi_i = 0$, which is the Hessian equality we are referring to.

3.1 Approximate differentiability

We justify the definition of *density point* for Riemannian manifolds by comparing it to its usual Euclidean counterpart. Denote by $\overline{B}(x, r)$ the closed metric ball centered at x with radius r.

Lemma 3.1 (Density points). *Let (M, g) be a Riemannian manifold and let A be a Borel subset of M. We call $x \in M$ a density point of A (with respect to* Vol*) if*

$$\lim_{r \downarrow 0} \frac{\operatorname{Vol}[\overline{B}(x, r) \setminus A]}{\operatorname{Vol}[\overline{B}(x, r)]} = 0.$$

This definition is equivalent to the standard one with respect to the Lebesgue measure after pulling x and A back to the Euclidean space through an arbitrary chart around x. In particular, almost every point of A is a density point of A with respect to Vol.

Proof. Denote by m the dimension of M. In a (smooth) local chart (φ, U) with U a small enough neighborhood of $x \in M$, the metric of M is bounded (from both sides) by the metric of $\mathbb{R}^m$ with constant scales $0 < c_1 < c_2$. It follows that $c_1^m \mathcal{L}^m(\varphi(N)) \le \operatorname{Vol}(N) \le c_2^m \mathcal{L}^m(\varphi(N))$ for any measurable subset $N \subset U$ [55, Proposition 12.6 and 12.7]. Hence, x is a density point of A if and only if

$$\lim_{r \downarrow 0} \frac{\mathcal{L}^m[\varphi(\overline{B}(x, r)) \setminus \varphi(A \cap U)]}{\mathcal{L}^m[\varphi(\overline{B}(x, r))]} = 0. \tag{13}$$

Applying again the relation between the metric of M and the metric of $\mathbb{R}^m$, for any $r > 0$, we have $\overline{B}(\varphi(x), c_1 r) \subset \varphi(\overline{B}(x, r)) \subset \overline{B}(\varphi(x), c_2 r)$. Therefore, (13) is equivalent to that $\varphi(x)$ is a density point of $\varphi(A)$ with respect to $\mathcal{L}^m$. $\qquad \square$

We now recall the definition of *approximate derivatives* first on Euclidean space (see [11, 5.8(v)] and [23, 3.1.2] for more detailed discussions), then on manifolds.

Definition 3.2 (Approximate derivatives on Euclidean spaces). Let $m, n \ge 1$ be two positive integers. Given a function $F : \Omega \to \mathbb{R}^n$ defined on a subset Ω of $\mathbb{R}^m$, $l \in \mathbb{R}^n$ is an *approximate limit* of F at a point $x \in \mathbb{R}^m$, for which we write $l = \operatorname{ap}\lim_{y \to x} F(y)$, if there exists a Borel set $\Omega_x \subset \Omega$ such that x is a density point of Ω_x and $\lim_{y \in \Omega_x, y \to x} F(y) = l$. The approximate derivatives of F are defined via the approximate limits of its difference quotients as follows.

A linear map $L : \mathbb{R}^m \to \mathbb{R}^n$ is called the *approximate derivative* of a function $F : \Omega \to \mathbb{R}^n$ at a point $x \in \Omega \subset \mathbb{R}^m$ if

$$\operatorname{ap}\lim_{y \to x} \frac{|F(y) - F(x) - L(y - x)|}{|y - x|} = 0. \tag{14}$$

The approximate derivative L will be denoted by $\operatorname{ap} \mathrm{D}_x F$.

The previous definition can be extended to the Riemannian setting as follows:

Lemma 3.3 (Approximate derivatives on manifolds)**.** *Let (M, g) be an m-dimensional Riemannian manifold M and let $f : A \to \mathbb{R}^n$ be a function defined on a subset A of M. Given an arbitrary local chart (φ, U) around a point $x \in A$, f is said to be approximately differentiable at x if the approximate derivative $\mathrm{ap}\, \mathrm{D}_{\varphi(x)}[f \circ \varphi^{-1}|_{\varphi(A \cap U)}]$ exists. The approximate derivative of f at x is then defined as*

$$\mathrm{ap}\, \mathrm{D}_x f := \mathrm{ap}\, \mathrm{D}_{\varphi(x)}[f \circ \varphi^{-1}|_{\varphi(A \cap U)}] \circ \mathrm{D}_x \varphi : T_x M \to \mathbb{R}^n,$$

where $\mathrm{D}_x \varphi : T_x M \to T_{\varphi(x)} \mathbb{R}^m$ denotes the differential map of φ at x and the tangent space $T_{\varphi(x)} \mathbb{R}^m$ is canonically identified with $\mathbb{R}^m$ in the above composition of functions. In particular, a constant function has null approximate derivative at density points located in its domain.

Proof. In Euclidean space, approximate derivatives are unique when they exist [21, Theorem 6.3]. Since density points are well-defined for Riemannian manifolds by Lemma 3.1 and coordinate changes for M are smooth diffeomorphisms, it follows from (14) that the existence of approximate derivative at a given point is independent of the choice of the chart and the change of variables rule applies. To show our last statement, note that $L = 0$ satisfies (14) whenever $F := f \circ \varphi^{-1}$ is a constant function. $\square$

3.2 Approximate Hessian of locally semi-concave functions

The properties of locally semi-concave functions provide a valuable toolbox for analyzing optimal transport maps on manifolds. In this section, we examine the weak second-order regularity of these functions.

In a Riemannian manifold (M, g), a subset C of M is said to be a *geodesically convex* (or *simple and convex*) set if, given any two points in C, there is a unique minimizing geodesic contained within C that joins those two points. A function $f : C \to \mathbb{R}$ defined on a geodesically convex set $C \subset M$ is said to be *geodesically convex* (respectively *geodesically concave*) if the composition $f \circ \gamma$ of f and any geodesic curve γ contained within C is convex (respectively concave). It is noteworthy that for any point $x \in M$, there exists an open ball centered at x that is geodesically convex [35,59], and such a ball is referred to as a geodesically convex ball.

Definition 3.4 (Semi-concavity)**.** Let (M, g) be a Riemannian manifold. Fix an open subset $O \subset M$. A function $\phi : O \to \mathbb{R}$ is semi-concave at $x \in O$ if there exists an open and geodesically convex set $C(x)$ centered at x and a $\mathcal{C}^2$ function $V : C(x) \to \mathbb{R}$ such that $\phi + V$ is geodesically concave throughout $C(x)$. The function ϕ is locally semi-concave on O if it is semi-concave at each point of O.

Bangert [7, (2.3) Satz] proved that the notion of local semi-concavity is independent of the Riemannian metric. This property also follows from the following characterization of locally semi-concave functions (with a linear module), whose proof for the Euclidean case is detailed in [56, Proposition 4.3, Proposition 4.8] and [17, Theorem 5.1]. In [22, Appendix A], it is adopted as the definition of local semi-concavity. Denote by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|_2$ respectively the Euclidean inner product and its associated norm. To stress that certain points are coordinate representations of manifold points, we denote them by tilde symbols $\widetilde{x}$ and $\widetilde{z}$.

Proposition 3.5 (Characterization of local semi-concavity, [58, Proposition 10.12])**.** *Let (M, g) be an m-dimensional Riemannian manifold. Fix an open subset O of M. A function $f : O \to \mathbb{R}$ is*

locally semi-concave if and only if for each point in O, there exist a chart (φ, U) defined around the point and a positive constant $C > 0$ such that $\forall \widetilde{x} \in \varphi(U)$, $\exists l_{\widetilde{x}} \in \mathbb{R}^m$, $\forall \widetilde{z} \in \varphi(U)$,

$$\left(f \circ \varphi^{-1}\right)(\widetilde{z}) \leq \left(f \circ \varphi^{-1}\right)(\widetilde{x}) + \langle l_{\widetilde{x}}, \, \widetilde{z} - \widetilde{x} \rangle \; + C \, \|\widetilde{z} - \widetilde{x}\|_2^2.$$

Hence, a function is locally semi-concave if and only if it is so when expressed in local charts [22, discussion after Lemma A.9]. We shall apply this chart-independence, along with Alexandrov's theorem, to establish the weak second-order regularity of locally semi-concave functions.

In the following theorem, we revisit Alexandrov's theorem stated via approximate derivatives. The proofs of this theorem can be found in [58, Theorem 14.1] and [44, Theorem D.2.1]. To maintain clarity of notation, for a function $f : U \to \mathbb{R}$ defined on an open subset $U \subset \mathbb{R}^m$, we define its Euclidean gradient $\nabla^E f(x) \in \mathbb{R}^m$ at $x \in U$ as the (column) vector $(\partial_1 f(x), \partial_2 f(x), \ldots, \partial_m f(x))$ when all of these partial derivatives of f exist at x. By contrast, the symbol ∇f is reserved to denote the gradient of functions $f : U \to \mathbb{R}$ defined on some open subset U of a Riemannian manifold, which is a (possibly not continuous) vector field defined at points where f is differentiable.

Theorem 3.6 (Alexandrov's theorem)**.** *Let $f : U \subset \mathbb{R}^m \to \mathbb{R}$ be a semi-concave function. Then the Euclidean gradient $\nabla^E f$ of f is defined $\mathcal{L}^m$-almost everywhere on U:*

$$\nabla^E f : A \longrightarrow \mathbb{R}^m \qquad \text{with } A \in \mathcal{B}(\mathbb{R}^m) \text{ and } \mathcal{L}^m(U \setminus A) = 0.$$

For $\mathcal{L}^m$-almost everywhere on A, the function $\nabla^E f$ is approximately differentiable and its approximate derivative $(\partial_{ij}^2 f)_{1 \leq i, j \leq m}$ forms a symmetric matrix. Moreover, at every point x where such approximate derivative of $\nabla^E f$ exists, f admits a second-order Taylor expansion:

$$f(z) = f(x) + \langle \nabla^E f(x), z - x \rangle + \frac{1}{2} \langle \mathrm{ap}\, \mathrm{D}_x \nabla^E f(z - x), z - x \rangle + o(\|z - x\|_2^2). \tag{15}$$

Remark 3.7. In the literature, the weak second-order regularity in Alexandrov's theorem is expressed in different formulations, including the one that differentiates super-gradients of semi-concave functions [44, Theorem D.2.1, Theorem D.2.2]. Their equivalence to (15) is proven in [58, Theorem 14.25]. Compared to these equivalent formulations, our Theorem 3.6 further requires x to be a density point of A for the existence of $\mathrm{ap}\, \mathrm{D}_x \nabla^E f$. However, under our assumption that U is an open set, the condition $\mathcal{L}^m(U \setminus A) = 0$ implies that every point of A is a density point.

To extend our results to the Riemannian setting, we provide a concise review of the Riemannian Hessian. For a $\mathcal{C}^2$ function defined on a Riemannian manifold (M, g), the Hessian at a point $x \in M$ can be interpreted either as a self-adjoint linear map from the tangent space $T_x M$ to itself or as a symmetric bilinear form on $T_x M \times T_x M$. These two interpretations are related by duality through the Riemannian metric g at x [47, Proposition 2.2.6]. While we shall primarily adopt the linear map perspective in the subsequent sections, we shall utilize the bilinear form viewpoint in the following two paragraphs. This choice is motivated by the fact that the chart-based expression of the Hessian is simpler when viewed as a bilinear form.

In what follows, the Hessian of a $\mathcal{C}^2$ function on a Riemannian manifold is a particular instance of a continuous $(0, 2)$-tensor S. Namely, for any two given charts φ, ψ defined on a common open subset $U \subset M$, there exist two bilinear forms S_φ and S_ψ whose coefficients are continuous functions such that $\forall \widetilde{x} \in \varphi(U) \subset \mathbb{R}^m$, $\forall u, v \in \mathbb{R}^m$,

$$\left[S_\varphi(\widetilde{x})\right](u, v) = \left[S_\psi(T(\widetilde{x}))\right](\mathrm{D}_{\widetilde{x}} T(u), \mathrm{D}_{\widetilde{x}} T(v)),$$

where $T = \psi \circ \varphi^{-1}$ is assumed to be a smooth (transition) map defined on $\varphi(U)$. In the case of the Hessian of a $\mathcal{C}^2$ function f, its expression in a chart φ is given by

$$\mathrm{Hess}_{\widetilde{x}}(f \circ \varphi^{-1})(\partial_i, \partial_j) = \partial_{ij}^2(f \circ \varphi^{-1})(\widetilde{x}) - \sum_{k=1}^m \Gamma_{ij}^k(\widetilde{x}) \, \partial_k(f \circ \varphi^{-1})(\widetilde{x}),$$

where ∂_i are the coordinate vectors associated with the given coordinate system [38, p.60 of Chapter 3], and Γ_{ij}^k are the Christoffel symbols of the chart, see [47, Chapter 2] for more details.

In the particular case of a chart φ inducing a normal coordinate system at $x_0 \in M$ [49, §2 of Chapter II], i.e., $\varphi^{-1}(u) = \exp_{x_0}(u)$ after identifying $T_{x_0}M$ with $\mathbb{R}^m$ by choosing an orthonormal basis of $T_{x_0}M$, the matrix made with the metric components g_{ij} is the identity at $\widetilde{x}_0 = \varphi(x_0)$, and all its first-order partial derivatives (and thus the Christoffel symbols) vanish at $\widetilde{x}_0$ [27, 2.89 bis]. Hence, the above formula at the point $\widetilde{x}_0$ is simplified into

$$\mathrm{Hess}_{\widetilde{x}_0}(f \circ \varphi^{-1})(\partial_i, \partial_j) = \partial_{ij}^2(f \circ \varphi^{-1})(\widetilde{x}_0). \tag{16}$$

Since the metric matrix $(g_{ij})_{1 \le i,j \le m}$ at $\widetilde{x}_0$ is the identity, if we consider $\mathrm{Hess}_{\widetilde{x}_0}(f \circ \varphi^{-1})$ as a linear map from $\mathbb{R}^m \cong T_{x_0}M$ to itself, then it coincides with the derivative of $\nabla^E(f \circ \varphi^{-1})$ at $\widetilde{x}_0$.

As a consequence, we are led to the following definition of Hessian for semi-concave functions on a Riemannian manifold.

Definition 3.8 (Hessian of semi-concave functions)**.** Let $(M, \mathfrak{g})$ be an m-dimensional complete Riemannian manifold, $f : O \to \mathbb{R}$ be a semi-concave function defined on an open subset $O \subset M$, and $A \subset O$ be the subset of points where f is differentiable.

The function f is said to have an *approximate Hessian* or simply a *Hessian* at a point $x \in A$ if there exists a chart (φ, U) inducing a *normal* coordinate system around x such that $\nabla^E(f \circ \varphi^{-1})$ is approximately differentiable at $\varphi(x)$, and its approximate derivative is symmetric. Then the Hessian of f at x is the function $\mathrm{Hess}_x f$ from T_xM to T_xM defined by

$$\mathrm{Hess}_x f(u) := (\mathrm{D}_x\varphi)^{-1} \circ \mathrm{ap}\, \mathrm{D}_{\varphi(x)} \nabla^E(f \circ \varphi^{-1}) \circ \mathrm{D}_x\varphi(u), \quad \forall\, u \in T_xM. \tag{17}$$

Remark 3.9. To justify Definition 3.8, first note that if (ψ, V) is another chart defined in a neighborhood of x, then $\nabla^E(f \circ \varphi^{-1})$ is approximately differentiable at $\varphi(x)$ if and only if $\nabla^E(f \circ \psi^{-1})$ is approximately differentiable at $\psi(x)$; indeed both vector fields are related by the formula

$$^t(\mathrm{D}_{\psi(z)}T) \cdot [\nabla^E(f \circ \varphi^{-1})(\varphi(z))] = \nabla^E(f \circ \psi^{-1})(\psi(z)), \tag{18}$$

where z is close to x, $T := \varphi \circ \psi^{-1}$ is a $\mathcal{C}^\infty$ diffeomorphism defined around $\psi(x)$ and $^t(\mathrm{D}_{\psi(z)}T)$ is the transpose of T's differential at $\psi(z)$. See the proof of Lemma 3.3 for a similar argument. Moreover, in our definition (17) of $\mathrm{Hess}_x f(u)$, we can justify the independence of charts (inducing normal coordinate systems) in two different ways. Since the Hessian of a $\mathcal{C}^2$ function defined on manifolds is a tensor, the required independence is guaranteed by its simplified local expressions (16) in normal coordinate systems. Alternatively, we suppose that (ψ, V) also induces a normal coordinate system around x, which implies that the transition map $T = \varphi \circ \psi^{-1}$ is linear. By applying the chain rule to (17) for the chart (ψ, V), the independence follows from the linearity of $\mathrm{D}_{\psi(z)}T = T$ and the equality (18).

To summarize the content of this part, we have obtained the following analog of Alexandrov's theorem for locally semi-concave functions on Riemannian manifolds.

Proposition 3.10. *Let $(M, \mathfrak{g})$ be a complete Riemannian manifold. Fix an open subset $O \subset M$ and a locally semi-concave function $f : O \to \mathbb{R}$. For* Vol-*almost every $x \in O$, there exists a function $\operatorname{Hess}_x f : T_x M \to T_x M$, called the Hessian of f at x, such that*

- $\operatorname{Hess}_x f$ *is a self-adjoint operator on $T_x M$;*

- *the function f satisfies the following second-order expansion at x,*

$$f(\exp_x u) = f(x) + \mathrm{D}_x f(u) + \frac{1}{2} \mathfrak{g}_x(\operatorname{Hess}_x f(u), u) + o(\|u\|^2), \tag{19}$$

for $u \in T_x M$.

3.3 Differentiating optimal transport maps

In this part, we collect some properties of optimal transport maps between absolutely continuous measures on a Riemannian manifold, which are taken from [19, Sections 4 & 5]. These properties will be used in Section 4. To justify them, we remark that our definition of Hessian enjoys the second-order expansion (19), which allows us to apply properties proven for the Hessian defined in [19, Definition 3.9]. See Remark 3.7 and [19, Discussion after Definition 3.9] for more details.

Proposition 3.11 (Differentiating optimal transport maps, [19, Proposition 4.1]). *Let $(M, \mathfrak{g})$ be a complete Riemannian manifold. Given a c-concave function ϕ defined on $\overline{\mathcal{X}} \subset M$ with $\mathcal{X}$ a bounded open set, we set $F := \exp(-\nabla \phi)$, which is* Vol-*almost everywhere well-defined on $\mathcal{X}$. Fix a point $x \in \mathcal{X}$ such that $\operatorname{Hess}_x \phi$ exists (17). Then the point $y := F(x)$ is not in the cut locus of x, $\nabla \phi(x) = \nabla d_y^2/2(x)$, and $\operatorname{Hess}_x d_y^2/2 - \operatorname{Hess}_x \phi$ is positive semi-definite. Define the differential $\mathrm{D}_x F : T_x M \to T_y M$ of F at x as*

$$\mathrm{D}_x F := [\mathrm{D}_{-\nabla \phi(x)} \exp_x] \circ (\operatorname{Hess}_x d_y^2/2 - \operatorname{Hess}_x \phi), \tag{20}$$

and define $\operatorname{Jac} F(x) := \det \mathrm{D}_x F$ as the Jacobian determinant of $\mathrm{D}_x F$.

The Jacobian determinant of the differential $\mathrm{D}_x F$, as defined in Proposition 3.11, is calculated with respect to normal coordinate systems of the tangent spaces $T_x M$ and $T_y M$ [19, Lemma 2.1]. By [19, Claim 4.5], these algebraic Jacobians are equivalent to their geometric counterparts, which results in the following change of variables formula. For further details, see [58, p.364 of Chapter 14].

Proposition 3.12 (Interpolation and change of variable formula). *Let $(M, \mathfrak{g})$ be a complete Riemannian manifold. Fix two absolutely continuous measures $\mu, \nu \in \mathcal{W}_2(M)$ with supports contained in two bounded open sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let $F := \exp(-\nabla \phi)$ be the optimal transport map that pushes μ forward to ν, where $\phi \in \mathcal{I}^c(\overline{\mathcal{X}}, \overline{\mathcal{Y}})$ is a c-concave function given by Theorem 2.15.*

Denote by $\phi^c \in \mathcal{I}^c(\overline{\mathcal{Y}}, \overline{\mathcal{X}})$ the c-conjugate of ϕ. The set

$$\Omega := \left\{ x \in \mathcal{X} \mid F(x) \in \mathcal{Y}, \operatorname{Hess}_x \phi \text{ and } \operatorname{Hess}_{F(x)} \phi^c \text{ exist } \right\}$$

satisfies the following properties:

1. *$\mu(\Omega) = 1$;*

2. *defining $F^t := \exp(-t\nabla \phi)$ for $0 \leq t \leq 1$, we have $\operatorname{Jac} F^t > 0$ on Ω;*

3. *denote by f and g the density functions of μ and ν respectively; there exists a measurable subset $N \subset \Omega$ depending on these two density functions such that $\mu(N) = 1$ and for $x \in N$,*

$$f(x) = g(F(x)) \operatorname{Jac} F(x) > 0;$$

4. *for any Borel function A on $[0, +\infty)$ with $A(0) = 0$, with the set N as in Property 3,*

$$\int_M A(g) \, \mathrm{d} \operatorname{Vol} = \int_N A\left(\frac{f}{\operatorname{Jac} F}\right) \operatorname{Jac} F \, \mathrm{d} \operatorname{Vol}. \tag{21}$$

(Either both integrals are undefined or both take the same value in $\mathbb{R} \cup \{+\infty, -\infty\}$.)

Proof. All the statements follow from [19, Claim 4.4, Theorem 4.2, Corollary 4.7] except Property 2 for $t \in (0, 1)$. Recall that for any c-concave function ϕ, we always have $\det[\mathrm{D}_{-t\nabla\phi(x)} \exp_x] > 0$ since $\exp_x(-t\nabla\phi(x))$ is not in the cut locus of x. As $t\phi$ is c-concave for $0 < t < 1$, it suffices to show that

$$\det[\operatorname{Hess}_x d^2_{F(x)}/2 - \operatorname{Hess}_x \phi] > 0 \implies \forall t \in (0, 1), \det[\operatorname{Hess}_x d^2_{F^t(x)}/2 - t \operatorname{Hess}_x \phi] > 0. \tag{22}$$

Indeed, $\operatorname{Hess}_x d^2_{F^t(x)}/2 - t \operatorname{Hess}_x d^2_{F(x)}/2$ is positive semi-definite for $0 < t < 1$ [19, Lemma 2.3], which implies (22) according to Minkowski's determinant inequality [57, (5.23)]. $\square$

3.4 Proof of Hessian equality

The Hessian equality (23) to prove is a second-order relation. We first demonstrate a first-order counterpart of this equality using the conclusion of Proposition 2.8 that relates barycenters in manifolds to Wasserstein barycenters.

Theorem 3.13 (Hessian equality for Wasserstein barycenters)**.** *Let (M, g) be a complete Riemannian manifold. Given an integer $n \geq 2$, let $\lambda_i > 0, 1 \leq i \leq n$, be n positive real numbers such that $\sum_{i=1}^n \lambda_i = 1$ and let $\mu_i \in \mathcal{W}_2(M), 1 \leq i \leq n$, be n probability measures with compact support. We assume that μ_1 is absolutely continuous. The unique barycenter $\overline{\mu}$ of $\mathbb{P} := \sum_{i=1}^n \lambda_i \delta_{\mu_i}$ is absolutely continuous with compact support. For $1 \leq i \leq n$, let $F_i = \exp(-\nabla\phi_i)$ be the optimal transport map pushing $\overline{\mu}$ forward to μ_i, where ϕ_i is a c-concave function given by Theorem 2.15.*

For $\overline{\mu}$-almost every $x \in M$, x is a barycenter of $\sum_{i=1}^n \lambda_i \delta_{F_i(x)}$, and we have the Hessian equality

$$\sum_{i=1}^n \lambda_i \operatorname{Hess}_x \phi_i = 0. \tag{23}$$

Proof. By Theorem 2.24, $\overline{\mu}$ is absolutely continuous with compact support. We now apply Proposition 2.8 to $\mathbb{P}$. Since $\overline{\mu}$ is the unique barycenter of $\mathbb{P}$, it coincides with the barycenter constructed in Proposition 2.8. Consider the identity map $\operatorname{Id} : (M, \mathcal{B}(M), \overline{\mu}) \to M$ as a random variable taking values in M. It has law $\overline{\mu}$, and the random variable $F_i = F_i \circ \operatorname{Id}$ has law μ_i for $1 \leq i \leq n$. Proposition 2.8 implies that for $\overline{\mu}$-almost every $x \in M$, x is a barycenter of $\sum_{i=1}^n \lambda_i \delta_{F_i(x)}$.

Let Ω be a Borel subset of M with $\overline{\mu}(\Omega) = 1$ such that for $x \in \Omega$, $\nabla\phi_i(x)$ exists for $1 \leq i \leq n$ and x is a barycenter of $\sum_{i=1}^n \lambda_i \delta_{F_i(x)}$. Fix a point $x \in \Omega$. By definition, x reaches the minimum of the function

$$h : w \in M \mapsto W_2(\delta_w, \sum_{i=1}^n \lambda_i \delta_{F_i(x)})^2 = \sum_{i=1}^n \lambda_i d_{\mathsf{g}}(w, F_i(x))^2.$$

By Lemma 2.17, the fixed point x is out of the cut locus of any point $F_i(x)$ for $1 \leq i \leq n$. We can thus differentiate h at $w = x$ and get $\nabla h|_{w=x} = 0$. Since $\nabla \phi_i(x) = \frac{1}{2} \nabla d_{F_i(x)}^2|_{w=x}$ holds as both gradients exist [19, Lemma 3.3], it follows that $\sum_{i=1}^{n} \lambda_i \nabla \phi_i(x) = \frac{1}{2} \nabla h|_{w=x} = 0$.

Define $f := \sum_{i=1}^{n} \lambda_i \phi_i$ on a neighborhood of Ω that is a common domain for $\phi_i, 1 \leq i \leq n$. The function f is locally semi-concave as each ϕ_i is so, and for $x \in \Omega$, $\nabla f(x) = \sum_{i=1}^{n} \lambda_i \nabla \phi_i(x) = 0 \in T_x M$ by the previous arguments. Let $\Omega_1 \subset \Omega$ be the set where the Hessians of f and $\phi_i, 1 \leq i \leq n$, all exist. Let Ω_2 be the set of density points of Ω. We have $\mathrm{Vol}(\Omega \setminus \Omega_1) = 0$ by Proposition 3.10, and $\mathrm{Vol}(\Omega \setminus \Omega_2) = 0$ by [21, Theorem 1.35].

For $x \in \Omega_1$, using the linearity of the Hessian operator, we get $\mathrm{Hess}_x f = \sum_{i=1}^{n} \lambda_i \, \mathrm{Hess}_x \phi_i$ by (17). Besides, noting that ∇f is constant on Ω, we infer from the last statement of Lemma 3.3 that for $x \in \Omega_2 \cap \Omega$, $\mathrm{Hess}_x f = 0$. It follows that for $x \in \Omega_1 \cap \Omega_2$, $\sum_{i=1}^{n} \lambda_i \, \mathrm{Hess}_x \phi_i = 0$. This proves the theorem since $\bar{\mu}(\Omega_1 \cap \Omega_2) = 1$ thanks to the absolute continuity of $\bar{\mu}$. $\qquad\square$

4 Lower Ricci curvature bounds and displacement functionals

In this section, we introduce a class of displacement functionals exploiting the Hessian equality in Theorem 3.13. This is one of the primary difference between our approach and the one proposed by Kim and Pass [34] regarding the absolute continuity of the barycenter.

In Section 3, the notion of Hessian plays a central role in differentiating optimal transport maps. There is also the following widely used connection between $\mathrm{Hess}_x \phi$ and Jacobi equations involving $\exp(-\nabla \phi)$, which is demonstrated in various works including Sturm [51], Lott and Villani [41, §7], Cordero-Erausquin et al. [20] and Villani [58, Chapter 14]. The function $J(t)$ defined below is actually $\mathrm{D}_x \exp(-\nabla t \, \phi)$ using (20). By convention, for a function f with variable t, we denote by $\dot{f}$ its derivative with respect to t.

Proposition 4.1. *Let $(M, \mathfrak{g})$ be an m-dimensional complete Riemannian manifold and let ϕ be a c-concave function defined on $\overline{\mathcal{X}} \subset M$ with $\mathcal{X}$ a bounded open set. Fix a point $x \in \mathcal{X}$ such that $\mathrm{Hess}_x \phi$ (Proposition 3.10) exists. Then $t \in [0,1] \mapsto \gamma(t) = \exp(-t\nabla\phi)(x)$ is a minimal geodesic. Define*

$$J : t \in [0,1] \mapsto \mathrm{D}_{-t\nabla\phi(x)} \exp_x \cdot (\mathrm{Hess}_x d_{\gamma(t)}^2/2 - t \, \mathrm{Hess}_x \phi).$$

Denote by $\Delta\phi(x)$ the trace of $\mathrm{Hess}_x \phi$ and by $\det J(t), 0 \leq t \leq 1$ the determinant of $J(t)$ calculated in coordinates using orthonormal bases of $T_x M$ and $T_{\gamma(t)} M$. If $-K \in \mathbb{R}$ is a lower Ricci curvature bound of M along γ and $\det J > 0$, then $\ell := -\log \det J$ defined on $[0,1]$ satisfies

$$\ddot{\ell} \geq \dot{\ell}^2/m - K\|\nabla\phi(x)\|^2$$

with $\ell(0) = 0$ and $\dot{\ell}(0) = \Delta\phi(x)$. In particular,

$$l \geq \Delta\phi(x) - K\|\nabla\phi(x)\|^2/2,$$

where we define $l := \ell(1) = -\log \det J(1)$.

The following displacement functionals $f \, \mathrm{d\,Vol} \in \mathcal{W}_2(M) \mapsto \int G(f) \, \mathrm{d\,Vol}$ are inspired by the entropy functional, where $G(x) := x \log x$. To uniformly bound (from above) their values of the sequence of barycenter measures in the law of large numbers for Wasserstein barycenters, we add the assumption of bounded derivatives. Examples of G can be constructed according to Theorem 5.13.

Proposition 4.2 (Displacement functionals)**.** *Let $(M, \mathfrak{g})$ be an m-dimensional complete Rieman-nian manifold with a lower Ricci curvature bound $-K$ ($K \geq 0$). Given an integer $n \geq 2$, let $\lambda_i > 0, 1 \leq i \leq n$, be n positive real numbers such that $\sum_{i=1}^{n} \lambda_i = 1$ and let $\mu_i \in \mathcal{W}_2(M), 1 \leq i \leq n$, be n probability measures with compact support. Assume that there is an integer $1 \leq k \leq n$ such that for any index $1 \leq i \leq k$, μ_i is absolutely continuous with density function g_i. Denote by $\overline{\mu}$ the unique Wasserstein barycenter of $\mathbb{P} := \sum_{i=1}^{n} \lambda_i \delta_{\mu_i} \in (\mathcal{W}_2(\mathcal{W}_2(M)), \mathbb{W}_2)$, which is absolutely continuous, and we denote by f its density function.*

Let G be a function on $[0, \infty)$ such that $G(0) = 0$, and the function $H : x \in \mathbb{R} \mapsto G(e^x) e^{-x}$ is continuously differentiable with non-negative derivative bounded above by some constant $L_H > 0$. The following inequality holds,

$$\int_M G(f) \, \mathrm{d} \operatorname{Vol} \leq \sum_{i=1}^{k} \frac{\lambda_i}{\Lambda} \int_M G(g_i) \, \mathrm{d} \operatorname{Vol} + \frac{L_H K}{2\Lambda} \mathbb{W}_2(\mathbb{P}, \delta_{\overline{\mu}})^2 + \frac{L_H}{2\Lambda} (m^2 + 2m), \tag{24}$$

where we define the constant $\Lambda := \sum_{i=1}^{k} \lambda_i$.

Remark 4.3. The following example helps to understand (24). Take $\mathbb{P} = \lambda \delta_{\mu_1} + (1 - \lambda) \delta_{\mu_2}$ with $0 < \lambda < 1$ and absolutely continuous measures $\mu_1, \mu_2 \in \mathcal{W}_2(M)$. Set $G(x) := x \log x$. Since $H(x) = x$, we choose $L_H = 1$. Define $\operatorname{Ent}(f \cdot \operatorname{Vol}) := \int_M G(f) \, \mathrm{d} \operatorname{Vol}$. The inequality (24) becomes

$$\operatorname{Ent}(\overline{\mu}) \leq \lambda \operatorname{Ent}(\mu_1) + (1 - \lambda) \operatorname{Ent}(\mu_2) + \frac{K}{2} \lambda (1 - \lambda) W_2(\mu_1, \mu_2)^2 + \frac{m^2}{2} + m,$$

which has exactly one additional term $L_H(m^2 + 2m)/(2\Lambda)$ compared to the λ-convexity expression of Ent used to define lower Ricci curvature bound $-K$ for metric measure spaces in [52, §4,2] and [41, Definition 0.7].

Moreover, $L_H(m^2 + 2m)/(2\Lambda)$ is also the only additional term when we compare inequality (24) with the Wasserstein Jensen's inequality proven by Kim and Pass [34, Theorem 7.11], which corresponds to the case $k = n$. However, our inequality (24) for the case $k < n$ is crucial to the proof of our main result in the next section.

Proof of Proposition 4.2. For $1 \leq i \leq n$, let $F_i := \exp(-\nabla \phi_i)$ be the optimal transport map from $\overline{\mu}$ to μ_i with ϕ_i a c-concave function given by Theorem 2.15. According to Theorem 3.13 and Proposition 3.12, there exists a Borel set $\Omega \subset M$ with $\overline{\mu}(\Omega) = 1$ such that $\sum_{i=1}^{n} \lambda_i \operatorname{Hess}_x \phi_i = 0$ for $x \in \Omega$, $\operatorname{Jac} \exp(-t\nabla \phi_i) > 0$ on Ω for $t \in [0, 1]$ and $1 \leq i \leq k$, and

$$\int_M G(g_i) \, \mathrm{d} \operatorname{Vol} = \int_{N_i} G\left(\frac{f}{\operatorname{Jac} F_i}\right) \operatorname{Jac} F_i \, \mathrm{d} \operatorname{Vol}, \quad 1 \leq i \leq k, \tag{25}$$

where $N_i \subset \Omega$ for $1 \leq i \leq k$ are Borel sets such that $\overline{\mu}(N_i) = 1$ and $f = g_i(F_i) \operatorname{Jac} F_i > 0$ on N_i. Hence, $\log f$ is well-defined on $\cup_{i=1}^{k} N_i$. Define $l_i(x) := -\log \operatorname{Jac} F_i(x)$ on Ω. It follows from (25) that

$$\int_M G(g_i) \, \mathrm{d} \operatorname{Vol} = \int_{N_i} H(\log f + l_i) \, \mathrm{d} \overline{\mu}, \quad 1 \leq i \leq k. \tag{26}$$

Applying Proposition 4.1 to ϕ_i for $1 \leq i \leq k$, we have on Ω,

$$l_i \geq \Delta \phi_i - K \|\nabla \phi_i\|^2 / 2, \quad 1 \leq i \leq k. \tag{27}$$

For $x \in \Omega$ and $1 \le i \le n$, since $\mathrm{Hess}_x \, d_{F_i(x)}^2/2 - \mathrm{Hess}_x \, \phi_i$ is positive semi-definite (Proposition 3.11), we can also bound $\Delta\phi_i(x)$ from above using the upper bound of the Laplacian of distance functions observed by Kim and Pass [34, Lemmma 2.7]:

$$\Delta\phi_i(x) \le \Delta d_{F_i(x)}^2/2 \le m\frac{\sqrt{K}d_{\mathsf{g}}(x, F_i(x))}{\tanh(\sqrt{K}d_{\mathsf{g}}(x, F_i(x)))}$$
$$\le m(1 + \sqrt{K}d_{\mathsf{g}}(x, F_i(x))) \le m + m^2/2 + K\,\|\nabla\phi_i(x)\|^2/2, \qquad (28)$$

where we used the general inequality $\alpha/\tanh\alpha \le 1 + \alpha$ for $\alpha \ge 0$ [1], applied the inequality of arithmetic and geometric means to $\sqrt{K\,d_{\mathsf{g}}(x, F_i(x))^2} \cdot \sqrt{m^2}$, and employed the equality $d_{\mathsf{g}}(x, F_i(x)) = \|\nabla\phi_i(x)\|$ for $x \in \Omega$. With our assumptions on H, (27) and (28) imply that for $1 \le i \le k$, on the set $\cup_{i=1}^k N_i$ (where $\log f$ is well-defined),

$$H(\log f + l_i) - H(\log f) = H'(\xi)\,l_i \ge H'(\xi)[\Delta\phi_i - K\|\nabla\phi_i\|^2/2]$$
$$\ge H'(\xi)[\Delta\phi_i - K\|\nabla\phi_i\|^2/2 - m - m^2/2]$$
$$\ge L_H(\Delta\phi_i - K\|\nabla\phi_i\|^2/2) - L_H(m + m^2/2), \qquad (29)$$

where we applied the mean value theorem to H that gave the real number ξ between $\log f + l_i$ and $\log f$. Sum up k inequalities as (29) with coefficients λ_i/Λ on the set $\cup_{i=1}^k N_i$,

$$H(\log f) \le \sum_{i=1}^k \frac{\lambda_i}{\Lambda} H(\log f + l_i) - \frac{L_H}{\Lambda} \sum_{i=1}^k \lambda_i(\Delta\phi_i - K\|\nabla\phi_i\|^2/2) + L_H(m + m^2/2)$$
$$= \sum_{i=1}^k \frac{\lambda_i}{\Lambda} H(\log f + l_i) + \frac{L_H}{\Lambda} \sum_{i>k}^n \lambda_i\,\Delta\phi_i + \frac{L_H K}{2\Lambda} \sum_{i=1}^k \lambda_i\,\|\nabla\phi_i\|^2 + L_H(m + m^2/2)$$
$$\le \sum_{i=1}^k \frac{\lambda_i}{\Lambda} H(\log f + l_i) + \frac{L_H K}{2\Lambda} \sum_{i=1}^n \lambda_i\,\|\nabla\phi_i\|^2 + \frac{L_H}{2\Lambda}(m^2 + 2m), \qquad (30)$$

where we used $\sum_{i=1}^n \lambda_i\,\Delta\phi_i = 0$ derived from the Hessian equality for the first equality and used (28) for the last inequality. Finally, (24) follows from (26) after integrating (30) over $N_1 \cap \ldots \cap N_k$ against $\overline{\mu}$ since $\overline{\mu}(N_i) = 1$ for $1 \le i \le k$ and $W_2(\overline{\mu}, \mu_i)^2 = \int_M \|\nabla\phi_i\|^2 \, \mathrm{d}\,\overline{\mu}$ for $1 \le i \le n$. $\qquad\square$

5   Proof of our main result

In this section, we prove our main result, i.e., the following theorem.

Theorem 5.1. *Let (M, g) be a complete Riemannian manifold with a lower Ricci curvature bound. If a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(M))$ gives mass to the set of absolutely continuous probability measures on M, then its unique Wasserstein barycenter is absolutely continuous.*

New auxiliary results in this section no longer require Riemannian structure, so we usually consider a Polish metric space equipped with a σ-finite Borel measure.

---

[1] Since $\lim_{\alpha\downarrow 0} \frac{\alpha}{\tanh\alpha} = 1$, it suffices to show that the function $f(\alpha) := \sinh\alpha + \alpha\sinh\alpha - \alpha\cosh\alpha$ is non-negative for $\alpha \ge 0$. As $f(0) = 0$ and $f'(\alpha) = \sinh\alpha + \alpha(\cosh\alpha - \sinh\alpha) = \sinh\alpha + \alpha\,e^{-\alpha}$, we have $f'(\alpha) \ge 0$ and thus $f(\alpha) \ge f(0) = 0$.

5.1 Wasserstein barycenters' absolute continuity by approximation

We first deduce an intermediate result by applying the law of large numbers for Wasserstein barycenters to the displacement functionals introduced in Proposition 4.2.

The following lemma, taken from Santambrogio [50, Proposition 7.7, Remak 7.8], originates from Buttazzo and Freddi [15, Theorem 2.2], which was slightly generalized later in [5, Theorem 2.34]. One can find another slightly generalized version by Ambrosio et al. [4, Theorem 15.8, Theorem 15.9] with a proof for the case of Euclidean spaces.

Lemma 5.2. *Let E be a Polish metric space with a σ-finite Borel measure μ. Let G be a function defined on $[0, \infty)$ such that*

1. *$G(x) \geq 0$;*

2. *G continuous and convex;*

3. *$\lim\limits_{x \to \infty} G(x)/x = \infty$.*

With respect to the reference measure μ, if there is a sequence of absolutely continuous probability measures $\nu_i = f_i \, \mathrm{d}\mu$, $i \geq 1$ converging weakly to a probability measure ν such that $\liminf\limits_{i \to \infty} \int_E G(f_i) \, \mathrm{d}\mu$ is finite, then ν is also absolutely continuous and

$$\int_E G(f) \, \mathrm{d}\mu \leq \liminf_{i \to \infty} \int_E G(f_i) \, \mathrm{d}\mu < \infty, \tag{31}$$

where f is the density of ν.

Since convergence in Wasserstein metric implies weak convergence, Lemma 5.2 ensures that the set below is closed in $\mathcal{W}_2(E)$.

Definition 5.3 (B(G, L) sets)**.** Let E be Polish space with a σ-finite Borel measure μ. Let G be a function on $[0, \infty)$ such that

1. G is non-negative and $G(x) = 0$ for $x \in [0, 1]$;

2. G is non-decreasing, continuous and convex;

3. $\lim\limits_{x \to \infty} G(x)/x = \infty$;

4. the function $H(x) := G(e^x)/e^x$ has continuous non-negative bounded derivative.

Given $L > 0$, the following set of measures,

$$\mathrm{B}(G, L) := \left\{ \nu \in \mathcal{W}_2(E) \mid \nu = f \cdot \mu, \ \int_M G(f) \, \mathrm{d}\mu \leq L \right\},$$

is a closed subset of $\mathcal{W}_2(E)$.

The function $\widehat{G} : x \mapsto x \log x$ on $[0, +\infty)$ is not always positive and non-decreasing, so it fails to meet the above assumptions. Since $\widehat{G}(e^{-1}) = -e^{-1}$ is the minimum value of $\widehat{G}$, we can consider the function that is equal to 0 on $[0, 1]$ and is equal to $\widehat{G}(x/e) + e^{-1}$ on $x \in [1, +\infty)$, which is a valid example. Indeed, we include the property that G is non-decreasing to ensure that each element in B(G, L) can be approximated by elements in B$(G, L + 1)$ with compact support, as shown in the following lemma.

26

Lemma 5.4. *Let (E, d) be a proper metric space equipped with a σ-finite Borel measure μ. Fix a $\mathrm{B}(G, L)$ set as defined in Definition 5.3. For any probability measure $\nu \in \mathrm{B}(G, L)$, there exists a sequence of probability measures in $\mathrm{B}(G, L+1)$ with compact support that converges to ν with respect to the Wasserstein metric.*

Proof. Let f be the density function of ν with respect to μ, i.e., $\nu = f \cdot \mu$. Since the integral $\int_E f \, \mathrm{d}\mu = 1$ is non-zero, there exists a positive number $l > 0$ such that the set $\{x \in E \mid f(x) \leq l\}$ is not μ-negligible. Since μ is σ-finite, there exists a bounded subset $Y \subset E$ such that $f(y) \leq l$ for $y \in Y$ and $0 < \mu(Y) < +\infty$. We define for $(k, x) \in \mathbb{N}^* \times E$,

$$g(k, x) := f(x) \mathbb{1}_{\overline{B}(x_0, k)}(x) + \alpha_k \, \mathbb{1}_{Y \cap \overline{B}(x_0, k)}(x), \tag{32}$$

where we set $\alpha_k := 0$ if $\mu(Y \cap \overline{B}(x_0, k)) = 0$ and $\alpha_k := [1 - \nu(\overline{B}(x_0, k))]/\mu(Y \cap \overline{B}(x_0, k))$ if $\mu(Y \cap \overline{B}(x_0, k)) > 0$. Since $\lim_{k \to \infty} \mu(Y \cap \overline{B}(x_0, k)) = \mu(Y) > 0$, for k sufficiently large such that $a_k > 0$, the sequence α_k is decreasing with $\lim_{k \to +\infty} \alpha_k = 0$. Let $k_0 \in \mathbb{N}^*$ be the smallest integer such that $\alpha_{k_0} > 0$. Our choices of α_k and k_0 ensure that for $n \in \mathbb{N}^*$, $\alpha_{k_0+n} > 0$ and $g(k_0 + n, \cdot)$ is a probability density function with respect to μ. Define $\nu_n := g(k_0 + n, \cdot) \cdot \mu$. Since (E, d) is a proper metric space, Y is pre-compact set, which implies that ν_n is a probability measure with compact support and thus $\nu_n \in \mathcal{W}_2(E)$. We now prove the convergence $\nu_n \to \nu$ with respect to W_2 using test functions. For a continuous function $\phi : E \to \mathbb{R}$ such that $|\phi(x)| \leq 1 + d(x_0, x)^2$, note that

$$|\phi(x) \, g(k_0 + n, x)| \leq \left(1 + d(x_0, x)^2\right) \cdot \left(f(x) + \alpha_{k_0} \, \mathbb{1}_Y(x)\right) \quad \text{and} \quad \lim_{n \to \infty} g(k_0 + n, x) = f(x).$$

As Y is pre-compact with $\mu(Y) < +\infty$ and $\nu \in \mathcal{W}_2(E)$, it follows from the dominated convergence theorem that

$$\lim_{n \to \infty} \int_E \phi \, \mathrm{d}\nu_n = \lim_{n \to \infty} \int_E \phi(x) \, g(k_0 + n, x) \, \mathrm{d}\mu(x) = \int_E \phi(x) \, f(x) \, \mathrm{d}\mu(x) = \int_E \phi \, \mathrm{d}\nu,$$

which implies $\lim_{n \to \infty} W_2(\nu_n, \nu) = 0$ according to the characterization of W_2 using the test function ϕ [58, (iv) of Definition 6.8 and Theorem 6.9].

Since $f(y) \leq l$ for $y \in Y$ and G is non-decreasing, we have

$$\forall (n, x) \in \mathbb{N}^* \times E, \quad G(g(k_0 + n, x)) \leq G(f(x)) + G(l + \alpha_{k_0}) \, \mathbb{1}_Y(x). \tag{33}$$

Since G is a continuous function and $\mu(Y) < 0$, we can apply the dominated convergence theorem to (33) and obtain

$$\lim_{n \to \infty} \int_E G(g(k_0 + n, x)) \, \mathrm{d}\mu(x) = \int_E G(f(x)) \, \mathrm{d}\mu(x).$$

Hence, for n sufficiently large, $\nu_n \in \mathrm{B}(G, L+1)$, which concludes the proof. $\square$

As the assumptions in Definition 5.3 include the ones we used to construct displacement functionals in Proposition 4.2, we obtain the following intermediate result.

Proposition 5.5. *Let (M, g) be a complete Riemannian manifold with a lower Ricci curvature bound. If $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(M))$ gives mass to some closed set $\mathrm{B}(G, L)$ defined in Definition 5.3, i.e., $\mathbb{P}(\mathrm{B}(G, L)) > 0$, then the unique barycenter of $\mathbb{P}$ is absolutely continuous.*

27

Proof. Write $\mathbb{P} = \mathbb{P}(\mathsf{B}(G,L))\,\mathbb{P}^1 + (1 - \mathbb{P}(\mathsf{B}(G,L)))\,\mathbb{P}^2$ with $\mathbb{P}^1, \mathbb{P}^2 \in \mathcal{W}_2(\mathcal{W}_2(M))$ such that $\mathbb{P}^1$ is supported in $\mathsf{B}(G,L)$. We approximate $\mathbb{P}$ in the Wasserstein metric $\mathbb{W}_2$ with finitely supported measures $\mathbb{P}_j \in \mathcal{W}_2(\mathcal{W}_2(M))$ by approximating $\mathbb{P}^1$ and $\mathbb{P}^2$ as follows.

Since $\mathsf{B}(G,L)$ equipped with the Wasserstein metric $\mathbb{W}_2$ is a non-empty closed subspace of $\mathcal{W}_2(M)$, we can construct the Wasserstein space $\mathcal{W}_2(\mathsf{B}(G,L))$ and treat $\mathbb{P}^1$ as an element in it. Recall that the set of finitely supported measures is dense in Wasserstein spaces [58, Theorem 6.18]. Applying this property to the Wasserstein spaces $\mathcal{W}_2(\mathsf{B}(G,L))$ and $\mathcal{W}_2(\mathcal{W}_2(M))$, we obtain two sequences of finitely supported probability measures $\{\mathbb{P}_j^1\}_{j\geq 1}$ and $\{\mathbb{P}_j^2\}_{j\geq 1}$ satisfying $\mathbb{W}_2(\mathbb{P}_j^1, \mathbb{P}^1) \to 0$, $\mathbb{W}_2(\mathbb{P}_j^2, \mathbb{P}^2) \to 0$ when $j \to \infty$. Furthermore, thanks to Lemma 5.4, we can further refine the two approximating sequences to ensure that all $\mathbb{P}_j^1, \mathbb{P}_j^2$ for $j \geq 1$ are supported in probability measures with compact support and $\mathbb{P}_j^1(\mathsf{B}(G,L+1)) = 1$. Define $\mathbb{P}_j := \mathbb{P}(\mathsf{B}(G,L))\,\mathbb{P}_j^1 + (1 - \mathbb{P}(\mathsf{B}(G,L)))\,\mathbb{P}_j^2$. It follows that $\mathbb{W}_2(\mathbb{P}_j, \mathbb{P}) \to 0$ as $j \to \infty$.

Consider the displacement functional $\mathcal{G} : f \cdot \mathrm{Vol} \mapsto \int_M G(f)\,\mathrm{d}\,\mathrm{Vol}$. Proposition 4.2 implies the following estimate of $\mathcal{G}(\mu_{\mathbb{P}_j})$ at the barycenter $\mu_{\mathbb{P}_j}$ of $\mathbb{P}_j$,

$$\mathcal{G}(\mu_{\mathbb{P}_j}) \leq \int_{\mathcal{W}_2(M)} \mathcal{G}(\nu)\,\mathrm{d}\,\mathbb{P}_j^1(\nu) + \frac{L_H K}{2\Lambda}\mathbb{W}_2(\mathbb{P}_j, \delta_{\mu_{\mathbb{P}_j}})^2 + \frac{L_H}{2\Lambda}(m^2 + 2m), \tag{34}$$

where $\Lambda := \mathbb{P}(\mathsf{B}(G,L))$, $-K$ is a lower Ricci curvature bound of M, m is the dimension of M, and L_H is an upper bounded of the H' with $H(x) := G(e^x)e^{-x}$. Denote by $\mu_{\mathbb{P}}$ the unique barycenter of $\mathbb{P}$, Theorem 2.2 implies that $W_2(\mu_{\mathbb{P}_j}, \mu_{\mathbb{P}}) \to 0$ and thus $\mathbb{W}_2(\mathbb{P}_j, \delta_{\mu_{\mathbb{P}_j}}) \to \mathbb{W}_2(\mathbb{P}, \delta_{\mu_{\mathbb{P}}})$ as $j \to \infty$. Since the support of $\mathbb{P}_j^1$ is a subset of $\mathsf{B}(G,L+1)$ and $\mathbb{W}_2(\mathbb{P}_j, \delta_{\mu_{\mathbb{P}_j}})$ is bounded for $j \geq 1$, by setting

$$L' := (L+1) + \frac{L_H K}{2\Lambda}\sup_{j\geq 1} \mathbb{W}_2(\mathbb{P}_j, \delta_{\mu_{\mathbb{P}_j}})^2 + \frac{L_H}{2\Lambda}(m^2 + 2m),$$

we have $\mu_{\mathbb{P}_j} \in \mathsf{B}(G,L')$ for all $j \geq 1$. It follows from Lemma 5.2 that $\mu_{\mathbb{P}}$ is absolutely continuous. $\qquad\square$

We replace the assumption $\mathbb{P}(\mathsf{B}(G,L)) > 0$ by a more natural one in the next subsection.

5.2 Compactness using Souslin space theory

The last step towards our main result is to show that the closed subset $\mathsf{B}(G,L)$ needed in Proposition 5.5 always exists if $\mathbb{P}$ gives mass to the set of absolutely continuous measures. Our inspiration is the criterion of uniform integrability by Charles-Jean de la Vallée Poussin. This criterion [11, Theorem 4.5.9] constructs a functional $f \mapsto \int G(f)\,\mathrm{d}\,\mu$ that is uniformly bounded for a family of uniformly integrable functions. We have enough freedom in its construction to impose the properties required by Definition 5.3 on the function G. Pre-compact sets of measures with respect to the topology τ defined below are closely related to uniformly integrable families.

Definition 5.6 (The set $\mathbb{A}$ and four topologies $\tau_w, \tau_W, \tau, \tau_L$)**.** Let E be a Polish space with a σ-finite reference measure μ. Pick a point $x_0 \in E$ and define the following set of measurable functions on E,

$$\mathbb{A} := \left\{ f \in L^1(\mu) \,\Big|\, f \geq 0, \int_E f\,\mathrm{d}\,\mu = 1, \int_E d(x_0, x)^2 f(x)\,\mathrm{d}\,\mu(x) < \infty \right\}, \tag{35}$$

which is independent of the chosen point x_0. The set $\mathbb{A}$ is naturally identified via $f \leftrightarrow f \cdot \mu$ with the set of probability measures in $\mathcal{W}_2(E)$ that are absolutely continuous with respect to μ. We

introduce the following four topologies. Denote by τ_w the topology on $\mathcal{W}_2(E)$ with respect to the weak convergence, denote by τ_W the topology of the Wasserstein space $\mathcal{W}_2(E)$, denote by τ the weak topology on $L^1(\mu)$ induced by its dual space $L^\infty(\mu)$ [11, Theorem 4.4.1] and denote by τ_L the topology of the Lebesgue space $L^1(\mu)$. By definition, $\tau_w \subset \tau_W$ and $\tau \subset \tau_L$. Denote by $(\mathbb{A}, \tau_w), (\mathbb{A}, \tau_W), (\mathbb{A}, \tau)$ and $(\mathbb{A}, \tau_L)$ the four topological subspaces induced by these topologies on the set $\mathbb{A}$.

Consider the case when E is a complete Riemannian manifold and μ is the volume measure on E. By Lemma 2.12, $\mathbb{A}$ is a Borel set for the topology τ_W. Given a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(E))$ such that $\mathbb{P}(\mathbb{A}) > 0$, our goal is to find a compact subset $\mathcal{F}$ in $(\mathbb{A}, \tau)$ with $\mathbb{P}(\mathcal{F}) > 0$. If we can accomplish this, then $\mathcal{F}$ forms a family of uniformly integrable functions by the Dunford-Pettis theorem (Proposition 5.12), bringing us closer to the main result. To find such an $\mathcal{F}$, a direct but problematic approach is to argue that $\mathbb{P}$ is a Radon measure. However, this argument overlooks that crucial point that $\mathbb{P}$ (restricted on $\mathbb{A}$) must be a Borel measure with respect to the Borel sets of $(\mathbb{A}, \tau)$. To address this issue, we revisit the Souslin space theory. Our main reference is Bogachev [11, Section 6.6, Section 6.7, Section 7.4].

Definition 5.7 (Souslin space)**.** A set in a Hausdorff space is called Souslin if it is the image of a Polish metric space under a continuous map. A Souslin space is a Hausdorff space that is a Souslin set. The empty set is Souslin as well.

By definition, Polish spaces are Souslin. Here are some properties of Souslin spaces:

1. Every Borel subset of a Souslin space is a Souslin space [11, Theorem 6.6.7];

2. Let E and F be Souslin spaces and let $f : E \mapsto F$ be a measurable map. If f is bijective, then E and F share the same Borel sets, see [26, Proposition 423F] or [11, Theorem 6.7.3];

3. If E is a Souslin space, then every finite Borel measure μ on E is Radon [11, Theorem 7.4.3].

These properties are used in the following lemma to justify the previous arguments with Radon measures.

Lemma 5.8. *Let (E, d) be a Polish space with an outer regular and σ-finite Borel measure μ on E. Let $\mathbb{A}$ be as (35). The four topological subspaces, $(\mathbb{A}, \tau_w), (\mathbb{A}, \tau_W), (\mathbb{A}, \tau)$, and $(\mathbb{A}, \tau_L)$ share the same Borel sets.*

In particular, if $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(E))$ gives mass to the set $\mathbb{A}$, then it gives mass to a compact subset of $(\mathbb{A}, \tau)$.

Proof. For spaces $(\mathbb{A}, \tau_w)$ and $(\mathbb{A}, \tau_W)$, the first statement is already proven in [46, Lemma 2.4.2], and we recall its arguments here. By Lemma 2.12, $\mathbb{A}$ is a Borel set for both τ_w and τ_W. Since $(\mathcal{W}_2(E), W_2)$ is a Polish space, $(\mathbb{A}, \tau_W)$ is then a Souslin space as a Borel subset of $(\mathcal{W}_2(E), W_2)$. Consider the identity map Id : $(\mathbb{A}, \tau_W) \to (\mathbb{A}, \tau_w)$, it is continuous and bijective. By definition, $(\mathbb{A}, \tau_w)$ is a Souslin space as the image of the Souslin space $(\mathbb{A}, \tau_W)$ under the continuous map Id. Moreover, $(\mathbb{A}, \tau_W)$ and $(\mathbb{A}, \tau_w)$ share the same Borel sets since the measurable map Id is bijective.

We claim that $(\mathbb{A}, \tau_L)$ is also a Souslin space. We first prove that the Lebesgue space $L^1(\mu)$ is complete and separable using the assumption that E is Polish. $L^1(\mu)$ is complete for any measurable space E [11, Theorem 4.1.3]. Its separability is asserted in Brézis [14, Theorem 4.13] and Bogachev [11, Section 1.12(iii), Corollary 4.2.2, Exercise 4.7.63] but only proven for the case of Euclidean spaces. Here is a brief proof of it. Every Polish space is homeomorphic to a closed subspace of

$\mathbb{R}^\infty$ [11, Theorem 6.1.12]. Moreover, one can show that $L^1(\mu)$ is separable when $E = \mathbb{R}^\infty$ using the same arguments for Euclidean spaces. It follows that $L^1(\mu)$ is a Polish space. We then prove that $\mathbb{A}$ is a Borel set for the topology τ_L. Fix a point $x_0 \in E$. Define the following sets for integers $k, j \geq 1$,

$$A_{k,j} := \left\{ f \in L^1(\mu) \,\middle|\, f \geq 0, \int_E f \,\mathrm{d}\mu = 1, \int_E \min\{d(x_0, x)^2, k\} f(x) \,\mathrm{d}\mu(x) \leq j \right\}.$$

Fix two integers $k, j \geq 1$. We show that the set $A_{k,j}$ is a closed subset of $L^1(\mu)$. Let $\{f_i\}_{i \geq 1} \subset A_{k,j}$ be a sequence converging to $f \in L^1(\mu)$ in $L^1(\mu)$. Since $\{f_i\}_{i \geq 1}$ has a subsequence converging almost everywhere to f, f is non-negative (μ-almost everywhere). It follows that $\int_E f \,\mathrm{d}\mu = \|f\|_{L^1(\mu)} = \lim_{i \to \infty} \|f_i\|_{L^1(\mu)} = 1$. Noting that as $i \to \infty$,

$$\| \min\{d(x_0, \cdot)^2, k\} f_i - \min\{d(x_0, \cdot)^2, k\} f \|_{L^1(\mu)} \leq k \|f_i - f\|_{L^1(\mu)} \to 0,$$

which implies that $f \in A_{k,j}$. Hence, $A_{k,j}$ is a closed subset of $L^1(\mu)$. By the monotone convergence theorem, we have $\mathbb{A} = \cup_{j \geq 1} \cap_{k \geq 1} A_{k,j}$, which proves that $\mathbb{A}$ is a Borel set. Finally, $(\mathbb{A}, \tau_L)$ is a Souslin space as $\mathbb{A}$ is a Borel set of the Polish space $L^1(\mu)$.

By definition of τ_w and τ, we have the topological inclusions $(\mathbb{A}, \tau_w) \subset (\mathbb{A}, \tau) \subset (\mathbb{A}, \tau_L)$. Using the identity map as before, we conclude that the three topological spaces, $(\mathbb{A}, \tau_w)$, $(\mathbb{A}, \tau)$ and $(\mathbb{A}, \tau_L)$, share the same Borel sets since $(\mathbb{A}, \tau_L)$ is a Souslin space.

$\mathbb{P}$, restricted on $\mathbb{A}$, is then a Radon measure with respect to the common Borel sets for the four topological subspaces since finite Borel measures on Souslin spaces are Radon. Hence, $\mathbb{P}(\mathbb{A}) > 0$ can be approximated by the $\mathbb{P}$ measure of compact subsets of $(\mathbb{A}, \tau)$. $\qquad \square$

We now recall some results from functional analysis to prove Proposition 5.12 later, a slightly generalized Dunford-Pettis theorem that connects uniform integrability and the weak topology τ. The Eberlein–Šmulian theorem characterizes compact sets with respect to the weak topology of a Banach space. For its proof, see [2, Theorem 1.6.3] or [40, Theorem II.3].

Theorem 5.9 (Eberlein–Šmulian theorem). *A subset K of a Banach space E is pre-compact with respect to the weak topology $\sigma(E^*, E)$ if and only if, from each sequence of elements of K, we can extract a weakly convergent subsequence.*

To deal with set-wise convergence of countably additive set functions (i.e., signed measures), we recall the following Vitali–Hahn–Saks theorem. For its proof, see [54, §3.14], [11, Theorem 4.6.3] or [3, Theorem A8.15].

Theorem 5.10 (Vitali–Hahn–Saks theorem). *Let $(\Omega, \mathcal{B})$ be a measurable space with a probability measure μ on it. Let $\nu_n : \mathcal{B} \to \mathbb{R}$, $n \in \mathbb{N}$ be a sequence of real-valued countably additive set functions such that*

1. the limit $\lim_{n \to \infty} \nu_n(A) \in \mathbb{R}$ exists and is finite for any $A \in \mathcal{B}$;

2. each ν_n is absolutely continuous with respect to μ, i.e., for $A \in \mathcal{B}$, $\mu(A) = 0$ implies $\nu_n(A) = 0$.

Then $\{\nu_n\}_{n \in \mathbb{N}}$ is uniformly absolutely continuous with respect to μ, i.e.,

$$\sup_{n \in \mathbb{N}} |\nu_n(A)| \to 0 \ \text{as } \mu(A) \to 0.$$

As a corollary, we illustrate how to apply Theorem 5.10 with a σ-finite measure μ.

Corollary 5.11. *Let $(\Omega, \mathcal{B})$ be a measurable space with a σ-finite measure μ on it. Let $\{f_n\}_{n \in \mathbb{N}} \subset L^1(\mu)$ be a sequence of μ-integrable function such that there exists a μ-integrable function $f \in L^1(\mu)$ satisfying*

$$\forall A \in \mathcal{B}, \quad \lim_{n \to \infty} \int_A f_n \, d\mu = \int_A f \, d\mu.$$

Then for any $\epsilon > 0$, there exists $\delta > 0$ such that for $A \in \mathcal{B}$,

$$\mu(A) < \delta \implies \sup_{n \in \mathbb{N}} \int_A f_n \, d\mu < \epsilon.$$

Proof. Since μ is σ-finite, there exists an at most countable family of pairwise disjoint measurable sets, $\{E_j, j \in J\}$ ($J \subset \mathbb{N}$), such that $0 < \mu(E_j) < +\infty$ and $\mu(\Omega \setminus \cup_{j \in J} E_j) = 0$. Define the measure $\eta := \sum_{j \in J} \lambda_j \frac{1}{\mu(E_j)} \mu|_{E_j}$ with $\lambda_j := 2^{-j} / \sum_{k \in J} 2^{-k}$. Since $\sum_{j \in J} \lambda_j = 1$, η is a probability measure satisfying

$$\forall A \in \mathcal{B}, \quad \eta(A) = \int_A \sum_{j \in J} \frac{\lambda_j}{\mu(E_j)} \mathbb{1}_{E_j} \, d\mu. \tag{36}$$

As $\mu(\Omega \setminus \cup_{j \in J} E_j) = 0$, (36) implies that $\eta(A) = 0$ if and only if $\mu(A) = 0$.

For $n \in \mathbb{N}$, define the countably additive function $\nu_n : \mathcal{B} \to \mathbb{R}$,

$$\nu_n(A) := \int_A f_n \, d\mu, \quad A \in \mathcal{B}.$$

As $f \in L^1(\mu)$, the limit $\lim_{n \to \infty} \mu_n(A) = \int_A f \, d\mu$ always exists and is finite. Since $\eta(A) = 0$ implies $\mu(A) = 0$ and thus $\nu_n(A) = 0$, Theorem 5.10 is applicable to $\{\nu_n\}_{n \in \mathbb{N}}$ with the probability measure η, which implies that $\sup_{n \in \mathbb{N}} |\nu_n(A)| \to 0$ as $\eta(A) \to 0$. Moreover, since η is finite measure that is absolutely continuous with respect to μ, the convergence $\mu(A) \to 0$ implies $\eta(A) \to 0$ [18, Lemma 4.2.1]. Hence, $\mu(A) \to 0$ implies $\sup_{n \in \mathbb{N}} |\nu_n(A)| \to 0$, which concludes the proof. $\square$

We are ready to prove the following Dunford-Pettis theorem, where the σ-finiteness of μ is not a standard assumption.

Proposition 5.12 (Dunford-Pettis theorem). *Let $(E, \mathcal{B})$ be a measurable space with a σ-finite Borel measure μ on it. Let $\mathcal{F} \subset L^1(\mu)$ be a set of μ-integrable functions. If $\mathcal{F}$ has compact closure in the weak topology induced by the dual space $L^\infty(\mu)$ of $L^1(\mu)$, then $\mathcal{F}$ is* uniformly integrable, *i.e.,*

$$\lim_{C \to \infty} \sup_{f \in \mathcal{F}} \int_{\{|f| > C\}} |f| \, d\mu = 0.$$

Proof. We need the assumption of μ being σ-finite to ensure that $L^\infty(\mu)$ is the dual space of $L^1(\mu)$, see [11, Theorem 4.4.1] and [48, Exercise 6.12]. The above definition of uniform integrability is taken from Bogachev [11, Definition 4.5.1]. When μ is finite, the equivalence between pre-compactness in the weak topology and uniform integrability is already proven by Bogachev [11, Theorem 4.7.18]. The following arguments for the general case are based on his proof.

We prove our statement for σ-finite measures by contradiction. Suppose that $\mathcal{F}$ has compact closure in the weak topology, but is not uniformly integrable. Then, there are $\epsilon > 0$ and a sequence $\{f_n\}_{n \geq 1} \subset \mathcal{F}$ such that

$$\inf_{n \geq 1} \int_{\{|f_n| > n\}} |f_n| \, d\mu \geq \epsilon. \tag{37}$$

31

Applying the Eberlein–Šmulian theorem (Theorem 5.9) to $\{f_n\}$ and the Banach space $L^1(\mu)$ [14, Theorem 4.8], we obtain a subsequence $\{f_{n_k}\}_{k \geq 1}$ convergent to some function $f \in L^1(\mu)$ in the weak topology. In particular, for every measurable set $A \in \mathcal{B}$ we have

$$\lim_{k \to \infty} \int_A f_{n_k} \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu. \tag{38}$$

It follows from the Vitali–Hahn–Saks theorem (Corollary 5.11) that sequence $\{f_{n_k}\}_{k \geq 1}$ has uniformly absolutely continuous integrals, i.e., for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$\mu(A) < \delta \implies \sup_{k \geq 1} \int_A |f_{n_k}| \, \mathrm{d}\mu < \epsilon. \tag{39}$$

Via the isometric embedding of $L^1(\mu)$ into the dual space of $L^\infty(\mu)$ [14, Corollary 1.4], the Banach–Steinhaus theorem [11, Theorem 4.4.3] is applicable to the Banach space $L^\infty(\mu)$ and the convergent sequence of functional $\{f_{n_k}\}_{k \geq 1}$, which implies that $C := \sup_{k \geq 1} \|f_{n_k}\|_{L^1(\mu)} < \infty$ is finite. Take the δ given by (39) for the ϵ in (37), and let n be an integer bigger than C/δ. Then by Chebyshev's inequality,

$$\sup_{k \geq 1} \mu(\{|f_{n_k}| > n\}) \leq \frac{1}{n} \sup_{k \geq 1} \|f_{n_k}\|_{L^1(\mu)} < \delta,$$

which leads to a contradiction between (37) and (39). $\qquad \square$

We also generalize the de la Vallée Poussin criterion to construct the function G in Definition 5.3. In the following proposition, the σ-finiteness of μ allows us to apply Fubini's theorem.

Theorem 5.13 (De la Vallée Poussin criterion)**.** *Let $(E, \mathcal{B})$ be a measurable space with a σ-finite Borel measure μ on it. A subset $\mathcal{F} \subset L^1(\mu)$ is uniformly integrable, i.e.,*

$$\lim_{C \to \infty} \sup_{f \in \mathcal{F}} \int_{\{|f| > C\}} |f| \, \mathrm{d}\mu = 0$$

if and only if there exists a function G defined on $[0, +\infty)$ such that

1. *$G(x) = 0$ for $0 \leq x \leq 1$;*

2. *G is a non-decreasing and convex function that is smooth on $(0, +\infty)$;*

3. *$\sup_{f \in \mathcal{F}} \int_E G(|f|) \, \mathrm{d}\mu \leq 1$;*

4. *if we define the function $H(x) := G(e^x)e^{-x}$ on $\mathbb{R}$, then $\lim_{x \to +\infty} H(x) = +\infty$, and its derivative H' is smooth with $0 \leq H'(x) \leq 1$.*

Proof. If we have the asserted function G for some subset $\mathcal{F} \subset L^1(\mu)$, then for every $\epsilon > 0$, we can find a real number $C > 0$ such that $G(t)/t \geq 2/\epsilon$ for any $t > C$. It implies that $|f(x)| \leq \epsilon \, G(|f(x)|)/2$ for all $f \in \mathcal{F}$ when $|f(x)| > C$. Hence,

$$\int_{\{|f| > C\}} |f| \, \mathrm{d}\mu \leq \frac{\varepsilon}{2} \int_{\{|f| > C\}} G \circ |f| \, \mathrm{d}\mu \leq \epsilon,$$

which shows that $\mathcal{F}$ is uniformly integrable.

Now assume that we are given a uniformly integrable subset $\mathcal{F} \subset L^1(\mu)$. To better motivate our construction of G, we postpone the definition of a smooth function H with $H(x) = 0, x \leq 0$ to (43) but use it here to define $G(x) := H(\log x)\, x$. Differentiate this equation twice, we obtain $G''(x) = [H'(\log x) + H''(\log x)]/x$. By our requirements on H, $G(x) = 0$ for $0 \leq x \leq 1$. Hence, we have $G(x) = \int_0^x \int_0^s G''(t)\, \mathrm{d}\, t \, \mathrm{d}\, s$ for $x > 0$ and thus

$$
\begin{aligned}
\int_E G(|f|)\, \mathrm{d}\, \mu &= \int_E \int_0^{|f|} \int_0^s G''(t)\, \mathrm{d}\, t \, \mathrm{d}\, s \, \mathrm{d}\, \mu = \int_E \int_{\mathbb{R}} \int_{\mathbb{R}} G''(t) \cdot \mathbb{1}_{0<t<s<|f|}\, \mathrm{d}\, t \, \mathrm{d}\, s \, \mathrm{d}\, \mu \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} G''(t) \cdot \mathbb{1}_{0<t<s} \cdot \mu(|f| > s)\, \mathrm{d}\, t \, \mathrm{d}\, s \\
&= \int_{\mathbb{R}} G''(t) \cdot \mathbb{1}_{t>0} \int_t^{\infty} \mu(|f| > s)\, \mathrm{d}\, s \, \mathrm{d}\, t \\
&= \int_0^{\infty} \frac{H'(\log t) + H''(\log t)}{t} \int_t^{\infty} \mu(|f| > s)\, \mathrm{d}\, s \, \mathrm{d}\, t \\
&= \int_{\mathbb{R}} [H'(y) + H''(y)] \int_{e^y}^{\infty} \mu(|f| > s)\, \mathrm{d}\, s \, \mathrm{d}\, y,
\end{aligned}
\tag{40}
$$

where we applied Fubini's theorem twice and a change of variable $y := \log t$. According to (40), we need to control $H' + H''$ and the integral of $\mu(|f| > s)$ at the same time. For the integral, note that by Fubini's theorem again, we have for $t > 0$ and $f \in L^1(\mu)$ that

$$
\begin{aligned}
\int_{\{|f|>t\}} |f|\, \mathrm{d}\, \mu &= \int_{\{|f|>t\}} \int_{\mathbb{R}} \mathbb{1}_{0<s<|f|}\, \mathrm{d}\, s \, \mathrm{d}\, \mu = \int_{\mathbb{R}} \int_E \mathbb{1}_{|f|>t} \cdot \mathbb{1}_{0<s<|f|}\, \mathrm{d}\, \mu \, \mathrm{d}\, s \\
&= \int_{\mathbb{R}} \int_E \mathbb{1}_{0<s<t<|f|} + \mathbb{1}_{0<t\leq s<|f|}\, \mathrm{d}\, \mu \, \mathrm{d}\, s \\
&= t\, \mu(|f| > t) + \int_t^{\infty} \mu(|f| > s)\, \mathrm{d}\, s.
\end{aligned}
\tag{41}
$$

Let $\alpha : \mathbb{N} \to \mathbb{N}$ be a strictly increasing function such that $\alpha(0) \geq 0$ and

$$
\sup_{f \in \mathcal{F}} \int_{e^{\alpha(n)}}^{\infty} \mu(|f| > s)\, \mathrm{d}\, s \leq \sup_{f \in \mathcal{F}} \int_{\{|f|>e^{\alpha(n)}\}} |f|\, \mathrm{d}\, \mu \leq 2^{-(n+1)},
$$

where we used (41) for the first inequality and the uniform integrability of $\mathcal{F}$ for the second one. It follows that

$$
\sup_{f \in \mathcal{F}} \sum_{n \geq 0} \int_{e^{\alpha(n)}}^{\infty} \mu(|f| > s)\, \mathrm{d}\, s \leq 1.
\tag{42}
$$

For the term $H' + H''$ in (40), we bound it from above with a function that is non-zero only on selected intervals based on our choice of $\alpha(n)$, allowing us to convert the integral of $\int_{e^y}^{\infty} \mu(|f| > s)\, \mathrm{d}\, s$ into the series summation (42). To achieve this, we first select a smooth function $\gamma : \mathbb{R} \to [0,1]$ such that $\gamma(x) = 1$ for $x \in [\alpha(n) + 1/3, \alpha(n) + 2/3]$ and $\gamma(x) = 0$ for $x \notin (\alpha(n), \alpha(n) + 1)$. Then we define

$$
H(x) := \begin{cases} \int_0^x e^{-s} \int_0^s \gamma(t)e^t\, \mathrm{d}\, t \, \mathrm{d}\, s, & x > 0 \\ 0, & x \leq 0 \end{cases}.
\tag{43}
$$

33

In this way, we have $H''(x) + H'(x) = \gamma(x)$. Using this construction, (40) and (42) imply that

$$\sup_{f \in \mathcal{F}} \int_E G(|f|) \, \mathrm{d}\mu = \sup_{f \in \mathcal{F}} \sum_{n \geq 0} \int_{\alpha(n)}^{\alpha(n)+1} \gamma(y) \int_{e^y}^{\infty} \mu(|f| > s) \, \mathrm{d}s \, \mathrm{d}y \leq \sup_{f \in \mathcal{F}} \sum_{n \geq 0} \int_{e^{\alpha(n)}}^{\infty} \mu(|f| > s) \, \mathrm{d}s \leq 1.$$

For the first derivative of H, we have

$$0 \leq H'(x) = e^{-x} \int_0^x \gamma(t) e^t \, \mathrm{d}t \leq e^{-x}(e^x - 1) \leq 1.$$

And by direct calculation we have that the difference

$$H(\alpha(n) + 1) - H(\alpha(n)) > \int_{\alpha(n)+\frac{2}{3}}^{\alpha(n)+1} e^{-s} \int_{\alpha(n)+\frac{1}{3}}^{\alpha(n)+\frac{2}{3}} e^t \, \mathrm{d}t \, \mathrm{d}s = (1 - e^{-\frac{1}{3}})^2$$

is bigger than a constant independent of n, which implies that $\lim_{x \to +\infty} H(x) = +\infty$ since H is non-decreasing. It follows from $0 \leq \gamma \leq 1$ that G is non-decreasing and convex as $G''(x) = \gamma(\log x)/x \geq 0$ for $x > 1$ and $G(x) = 0$ for $0 \leq x \leq 1$. □

5.3 Final step of the proof

To prove Theorem 5.1, it remains to combine the previous auxiliary propositions to replace the assumption in Proposition 5.5 that $\mathbb{P}(\mathsf{B}(G, L)) > 0$ for some set $\mathsf{B}(G, L)$ (Definition 5.3).

As in Definition 5.6, we denote by $\mathbb{A}$ the set of absolutely continuous measures in $\mathcal{W}_2(M)$. If $\mathbb{P}(\mathbb{A}) > 0$, then Lemma 5.8 provides a compact subset $\mathcal{F}$ of $(\mathbb{A}, \tau)$ such that $\mathbb{P}(\mathcal{F}) > 0$. Applying the Dunford-Pettis theorem (Proposition 5.12) to $\mathcal{F}$ with $\mu := \mathrm{Vol}$, we see that $\mathcal{F}$ is uniformly integrable. Then the de la Vallée Poussin criterion (Theorem 5.13) asserts the existence of a smooth function G such that $\mathcal{F} \subset \mathsf{B}(G, 1) \subset \mathbb{A}$. Therefore, our theorem follows from Proposition 5.5 and the property $\mathbb{P}(\mathsf{B}(G, 1)) \geq \mathbb{P}(\mathcal{F}) > 0$.

Conflicts of Interest Statement The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Statement Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

References

[1] M. AGUEH AND G. CARLIER, *Barycenters in the Wasserstein space*, SIAM Journal on Mathematical Analysis, 43 (2011), pp. 904–924.

[2] F. ALBIAC AND N. J. KALTON, *Topics in Banach space theory*, vol. 233 of Graduate Texts in Mathematics, Springer, New York, 2006.

[3] H. Alt and R. Nürnberg, *Linear Functional Analysis: An Application-Oriented Introduction*, Universitext, Springer, London, 2016.

[4] L. Ambrosio, E. Brué, and D. Semola, *Lectures on Optimal Transport*, vol. 130 of UNITEXT, 2021.

[5] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, 2000.

[6] 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.

[7] V. Bangert, *Analytische Eigenschaften konvexer Funktionen auf Riemannschen Mannigfaltigkeiten*, Journal für die reine und angewandte Mathematik, (1979), pp. 309–324.

[8] P. Bernard and B. Buffoni, *Optimal mass transportation and Mather theory*, Journal of the European Mathematical Society, 9 (2007), pp. 85–121.

[9] J. Bertrand, *Existence and uniqueness of optimal maps on Alexandrov spaces*, Advances in Mathematics, 219 (2008), pp. 838–851.

[10] P. Billingsley, *Convergence of probability measures*, Wiley series in probability and statistics Probability and statistics section, 2 ed., 1999.

[11] V. I. Bogachev, *Measure theory*, Springer, Berlin, 2007.

[12] N. Bourbaki, *Extension of a measure. L^p spaces*, in Elements of Mathematics: Integration I, Springer, Berlin, 2004, pp. 94–241.

[13] G. E. Bredon, *Topology and geometry*, vol. 139 of Graduate Texts in Mathematics, 1993.

[14] H. Brézis, *Functional analysis, Sobolev spaces and partial differential equations*, Universitext, Springer, New York, 2011.

[15] G. Buttazzo and L. Freddi, *Functionals defined on measures and applications to non equi-uniformly elliptic problems*, Annali di matematica pura ed applicata, 159 (1991), pp. 133–149.

[16] F. Cavalletti and A. Mondino, *Optimal maps in essentially non-branching spaces*, Communications in Contemporary Mathematics, 19 (2017), p. 1750007.

[17] F. H. Clarke, R. J. Stern, and P. R. Wolenski, *Proximal smoothness and the lower-$\mathcal{C}^2$ property*, Journal of Convex Analysis, 2 (1995), pp. 117–144.

[18] D. L. Cohn, *Measure theory*, Birkhäuser Advanced Texts Basler Lehrbücher, Birkhäuser, New York, 2 ed., 2013.

[19] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger, *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, Inventiones mathematicae, 146 (2001), pp. 219–257.

[20] ———, *Prékopa–leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport*, Annales de la Faculté des sciences de Toulouse : Mathématiques, Ser. 6, 15 (2006), pp. 613–635.

[21] L. C. Evans and R. F. Garzepy, *Measure theory and fine properties of functions*, Routledge, Oxfordshire, 2018.

[22] A. Fathi and A. Figalli, *Optimal transportation on non-compact manifolds*, Israel Journal of Mathematics, 175 (2010), pp. 1–59.

[23] H. Federer, *Geometric measure theory*, Springer, Berlin, 1996.

[24] A. Figalli and N. Juillet, *Absolute continuity of Wasserstein geodesics in the Heisenberg group*, Journal of Functional Analysis, 255 (2008), pp. 133–141.

[25] D. Fremlin, *Measurable selections and measure-additive coverings*, in Measure Theory Oberwolfach 1981, vol. 945 of Lecture Notes in Mathematics, 2006, pp. 425–431.

[26] D. H. FREMLIN, *Topological Measure Theory*, vol. 4 of Measure Theory, Torres Fremlin, Colchester CO3 3AT, England, 2005.

[27] S. GALLOT, D. HULIN, AND J. LAFONTAINE, *Riemannian Geometry*, Universitext, Springer Science & Business Media, Berlin, 2004.

[28] N. GIGLI, *On the inverse implication of Brenier-McCann theorems and the structure of* $(\mathcal{P}_2(M), W_2)$, Methods and Applications of Analysis, 18 (2011), pp. 127–158.

[29] ——, *Optimal maps in non branching spaces with Ricci curvature bounded from below*, Geom. Funct. Anal, 22 (2012), pp. 990–999.

[30] 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.

[31] Y. JIANG, *Absolute continuity of Wasserstein barycenters over Alesxandrov spaces*, Canadian Journal of Mathematics, 69 (2017), pp. 1087–1108.

[32] M. KELL, *Transport maps, non-branching sets of geodesics and measure rigidity*, Advances in Mathematics, 320 (2017), pp. 520–573.

[33] Y.-H. KIM AND B. PASS, *Multi-marginal optimal transport on Riemannian manifolds*, American Journal of Mathematics, 137 (2015), pp. 1045–1060.

[34] ——, *Wasserstein barycenters over Riemannian manifolds*, Advances in Mathematics, 307 (2017), pp. 640–683.

[35] S. KOBAYASHI AND K. NOMIZU, *Foundations of Differential Geometry, Volume 1*, vol. 1, John Wiley & Sons, Hoboken, 1996.

[36] G. KOUMOULLIS AND K. PRIKRY, *The Ramsey property and measurable selections*, Journal of the London Mathematical Society, s2-28 (1983), pp. 203–210.

[37] T. LE GOUIC AND J.-M. LOUBES, *Existence and consistency of Wasserstein barycenters*, Probability Theory and Related Fields, 168 (2017), pp. 901–917.

[38] J. M. LEE, *Introduction to Smooth Manifolds*, vol. 218 of Graduate Texts in Mathematics, Springer SZcience+Business Media, New York, 2 ed., 2013.

[39] ——, *Introduction to Riemannian manifolds*, Springer International Publishing AG, Chanm, Switzerland, 2 ed., 2018.

[40] D. LI AND H. QUEFFÉLEC, *Introduction to Banach Spaces: Analysis and Probability*, vol. 166 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge CB2 8BS, United Kingdom, 2017.

[41] J. LOTT AND C. VILLANI, *Ricci curvature for metric-measure spaces via optimal transport*, Annals of Mathematics, 169 (2009), pp. 903–991.

[42] R. J. MCCANN, *A convexity principle for interacting gases*, Advances in mathematics, 128 (1997), pp. 153–179.

[43] ——, *Polar factorization of maps on Riemannian manifolds*, Geometric & Functional Analysis GAFA, 11 (2001), pp. 589–608.

[44] C. NICULESCU AND L. PERSSON, *Convex Functions and Their Applications: A Contemporary Approach*, CMS Books in Mathematics, Springer International Publishing, Cham, Switzerland, 2 ed., 2018.

[45] S.-I. OHTA, *Barycenters in Alesxandrov spaces of curvature bounded below*, Advances in geometry, 12 (2012), pp. 571–587.

[46] V. M. PANARETOS AND Y. ZEMEL, *An invitation to statistics in Wasserstein space*, SpringerBriefs in Probability and Mathematical Statistics, Springer, Cham, Switzerland, 2020.

[47] P. PETERSEN, *Riemannian Geometry*, vol. 171 of Graduate Texts in Mathematics, 2016.

[48] W. RUDIN, *Real and Complex Analysis*, McGraw-Hill Book, Singapore, 3 ed., 1987.

[49] T. SAKAI, *Riemannian Geometry*, Translations of mathematical monographs, American Mathematical Soc., Providence, Rhode Island, 1996.

[50] 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.

[51] K.-T. STURM, *Convex functionals of probability measures and nonlinear diffusions on manifolds*, Journal de Mathématiques Pures et Appliquées, 84 (2005), pp. 149–168.

[52] ———, *On the geometry of metric measure spaces. I*, Acta Math, 196 (2006), pp. 65–131.

[53] ———, *On the geometry of metric measure spaces. II*, Acta Math, 196 (2006), pp. 133–177.

[54] C. W. SWARTZ, *Measure, integration and function spaces*, World Scientific, Singapore, 1994.

[55] M. E. TAYLOR, *Measure theory and integration*, vol. 76 of Graduate Studies in Mathematics, American Mathematical Society, Michigan, 2006.

[56] J.-P. VIAL, *Strong and weak convexity of sets and functions*, Mathematics of Operations Research, 8 (1983), pp. 231–259.

[57] C. VILLANI, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, Rhode Island, 2003.

[58] ———, *Optimal transport: old and new*, vol. 338 of Grundlehren der mathematischen Wissenschaften, 2009.

[59] J. H. C. WHITEHEAD, *Convex regions in the geometry of paths*, The Quarterly Journal of Mathematics, (1932), pp. 33–42.