RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

Université de Toulouse

Doctorat de l'Université de Toulouse

Régularité des barycentres de Wasserstein

Thèse présentée et soutenue, le 16 octobre 2025 par

Jianyu MA

École doctorale
EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse

Spécialité
Mathématiques et Applications

Unité de recherche
IMT : Institut de Mathématiques de Toulouse

Thèse dirigée par
Jérôme BERTRAND

Composition du jury
M. Nicolas JUILLET, Rapporteur, Université de Haute-Alsace
Mme Marjolaine PUEL, Examinatrice, Cergy Paris université
M. Filippo SANTAMBROGIO, Examinateur, Université Claude Bernard Lyon 1
M. Thierry KLEIN, Examinateur, École Nationale de l'Aviation Civile
M. Guillaume CARLIER, Examinateur, Université Paris Dauphine
M. Jérôme BERTRAND, Directeur de thèse, Université de Toulouse

My formula for greatness in a human being is *amor fati*: that one wants nothing to be different, not forward, not backward, not in all eternity. Not merely bear what is necessary, still less conceal it—all idealism is mendacity in the face of what is necessary—but *love* it.

— Friedrich Nietzsche

亦余心之所善兮，虽九死其犹未悔。[1]

—《离骚》(Li Sao)

热爱生命，丰盈存在；手托日月，直面命运。[2]

— The Author

[1] For that which my heart affirms as true, I would have no regrets, even if I were to die nine times for it.

[2] Will this life passionately, and cultivate a boundless being; hold aloft your sun and moon, and affirm your destiny steadfast.

Acknowledgements

This thesis is the culmination of a journey that began five years ago, during my Master 2 internship with my supervisor, Jérôme Bertrand, who introduced me to the wonderful research domain of Wasserstein barycenters. The memories from these five years remain vivid and have been deeply formative. Throughout this journey, I have been fortunate to receive immense support, and it is my great pleasure to now express my gratitude.

I am indebted to `LebEX CIMI` (Centre International de Mathématiques et d'Informatique) for funding my Master 2 studies and my three-year doctoral contract. For the final year of my thesis, `ENAC` (École Nationale de l'Aviation Civile) offered me a one-year `ATER` contract. The financial support from these two institutes provided the solid foundation for this research. I sincerely appreciate their support and trust, and I hope the work presented in this thesis proves worthy of them.

It is my supervisor who established the two main research directions of the thesis: absolute continuity of Wasserstein barycenters on non-compact manifolds and singular Wasserstein barycenters on metric trees. His visionary and profound mathematical intuition was essential to the success of this work. Throughout my doctoral studies, I have deeply appreciated his patience in offering critiques and corrections; this guidance was invaluable in my development as a professional mathematical scholar. I will always value his instruction, and I believe his influence on me will continue to manifest in my future career. In a personal note, I will also fondly remember his warmth and humor, and the pleasantness they brought to our many interactions.

I would also like to thank several other mathematicians who helped me along the way. Yuxin Ge of IMT (Institut de Mathématiques de Toulouse) initiated the international student program that first brought me to the Université de Toulouse. His tireless dedication to helping the Chinese students in our institute fostered a true sense of community, and we all felt his genuine kindness and care. I admire his character and hope that one day I can be as helpful to others as he has been. Nicolas Juillet, a leading expert in my research domain, served as a reporter and jury member for my thesis. Not only did he offer me precious opportunities to attend his academic conferences, but he was also generously responsive to my many questions. I am grateful to him and to Brendan Pass (University of Alberta), the two reporters for this thesis, for their careful reading of the manuscript and for the insightful discussions that helped to improve it.

Beyond this professional support, a thesis is also a deeply personal journey. The six-year Odyssey far from my hometown was made possible by friends who shared parts of this journey with me. Please allow me to dedicate the following words to them.

I must first mention the open-source community of Android modding, as it is impossible to name its many anonymous individuals. Under the pseudonym JingMatrix, I actively participate in and maintain projects for this community, many of which benefit users worldwide. The trust and appreciation I received from you gave me the courage to persevere through periods of intense self-doubt. Many of you treated me as a respected peer, and it was in that spirit of friendship that I could share my joy and passion for the open-source world. I truly appreciate the chance to connect with you, which has broadened my life significantly. It is my pleasure and honor to be a contributor.

Living ten thousand miles from my family in China, I was fortunate to find friends with whom I feel the same ease as with family. Christmas, a holiday that was once foreign to my cultural background, became meaningful to me after spending the last five with my French teacher, Leena Jasani. Though the memory of my first two-year life living at a 9m^2 studio is fading, every surviving frame contains Eric Vahatra, highlighted by our laughter. Alonso Herrera was an elder brother who taught me much, but more profoundly, he was an irreplaceable company to my soul—a truth I felt most clearly in its absence after he left Toulouse. As those old days eventually passed, I was glad to see Qian Liu in the final part of the journey. Her insistent and curious deciphering of my inner world was touching and greatly furthered my own self-reflection. It has been a warmth in my heart to receive her trust and, together, to build a genuinely supportive connection.

Some friends appear in one's life like shining stars, rendering this world even more adorable. To my old friends Ziwei Tang and Lin Du: distance has not blurred the ideals we see in each other; instead, it has transformed them into a consoling source of strength against the solitude that accompanies the path of "becoming ourselves". This journey has also been illuminated by many other wonderful friends, whom I appreciate for countless specific reasons. To the following, I love you for the common reason that you love your lives passionately (and you'd better, if not!): Xinghan Liu, Mitja Briscik, Jiuwen Sun, Runze Zhang, Alexey Lazarev, Chenyan Wu, Tong Zhao, Kevin Yan, Bojian Shang, Bruno Dewer, Kunwei Li, Runqian Yang, Chung-Ming Pan, and Fu-Hsuan He. What a joy it is to write your names, and to see fragments of our shared past illuminated once more.

There were depressing days during this thesis, days I did not know how to handle and still do not know how to recall. Thank you, Luoting Lu and Chuyun Yuan, for witnessing my suffering and for comforting me in my retreat.

To my grandmother and my little sister: you stay in the softest part of my heart, and no sentence will ever be sufficient to express my love for you.

Contents

List of symbols

$\mathbb{Z}, \mathbb{N}, \mathbb{N}^*$	integers, natural numbers $(0, 1, 2, \ldots)$, $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$
$\mathbb{R}, \mathbb{R}_+, \mathbb{R}_-, \mathbb{C}$	real numbers, $\mathbb{R}_+ = [0, +\infty)$, $\mathbb{R}_- = (-\infty, 0]$, complex numbers
$\mathbb{R}^\infty$	infinite-dimensional real Hilbert space, whose elements are sequences of real numbers with finite square sum
$(M, d_{\mathfrak{g}})$	connected Riemannian manifold M without boundary equipped with the distance function $d_{\mathfrak{g}}$ induced by a $\mathcal{C}^2$ smooth metric tensor $\mathfrak{g}$
D, D_x	differential operator for maps (at point x)
$\mathrm{Hess}_x f$	(approximate) Hessian of some function f at given point x
(E, d)	metric space E with distance function d
$\overline{B}(x, r)$	closed metric ball centered at x with radius r
$\mathcal{L}^m$	Lebesgue measure on $\mathbb{R}^m$
$\mathfrak{u} = \mathcal{L}^1\|_{[0,1]}$	uniform probability measure on $[0, 1]$
$f_\# \mu$	push-forward of some measure μ by a measurable function f
$(\mathcal{W}_2(E), d_W)$	Wasserstein space over some metric space E
$(\mathcal{W}_2(\mathcal{W}_2(E)), d_{\mathbb{W}})$	Wasserstein space over some Wasserstein space $\mathcal{W}_2(E)$
$\mathbb{P}, \mathbb{Q}$	probability measures on the Wasserstein spaces $\mathcal{W}_2(\mathcal{W}_2(E))$
$\mu_{\mathbb{P}}$	a definitely chosen barycenter of some probability measure $\mathbb{P}$
f_μ	distribution function of probability measure μ, which is defined to be non-decreasing and right-continuous
f_μ^{-1}	the right-continuous inverse function of the distribution function of probability measure μ, also known as the quantile function of μ
$L^p([0, 1]), L^p(\mu)$	space of functions f such that $\|f\|^p$ is integrable w.r.t $\mathfrak{u}$ on $[0, 1]$ or the given measure μ, $p \in [1, +\infty]$
$\vec{e} = \overrightarrow{\{v_0, v_1\}}$	oriented edge with two ends v_0, v_1 in this given order
$\widetilde{\mu}$	dual measure of some probability measure $\mu \in \mathcal{W}_2([0, 1])$, i.e., $f_{\widetilde{\mu}}$ coincides with f_μ^{-1} on the interval $(0, 1)$

Introduction

This thesis investigates the regularity of Wasserstein barycenters, approached from a geometric perspective. Building upon the geometric study of optimal transport, this research explores how the geometric properties of the underlying space influence whether Wasserstein barycenters are absolutely continuous or singular with respect to a reference measure.

Key contributions of this work include: first, establishing the absolute continuity of Wasserstein barycenters on Riemannian manifolds with a lower Ricci curvature bound under weaker assumptions than previously known; and second, characterizing the nature of singular Wasserstein barycenters in the specific setting of metric trees, linking singularity to the tree's branching structure.

We begin by introducing the concept of Wasserstein barycenters and the motivation for our research project.

Background and motivations

The concept of a barycenter, or Fréchet mean, extends the familiar notion of the mean (expected value) from Euclidean spaces to general metric spaces. For a metric space (E, d), a barycenter $z_\mu \in E$ of a probability measure μ on (E, d) is defined as a minimizer of the mean squared distance:

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

This definition provides a natural generalization; if (E, d) is a Euclidean space and μ is the law of a random variable, z_μ is its standard mean.

Wasserstein barycenters are barycenters defined for Wasserstein spaces, the metric spaces extensively studied in optimal transport theory. Throughout the subsequent discussion, we consider (E, d) a proper metric space, meaning that bounded closed subsets of E are compact. Barycenters are known to exist on such spaces; furthermore, proper metric spaces are Polish (i.e., complete and separable). The Wasserstein space (of order 2) over (E, d), denoted by $(\mathcal{W}_2(E), d_W)$, comprises probability measures μ on (E, d) with finite second moments, i.e., $\int_E d(x_0, x)^2 \,\mathrm{d}\,\mu(x)$ is finite for some point (and thus any) $x_0 \in E$. The Wasserstein metric d_W quantifies the distance of two measures $\mu, \nu \in \mathcal{W}_2(E)$ via the optimal transport plans between them,

$$d_W(\mu, \nu) := \sqrt{\inf_\gamma \int_{E \times E} d(x, y)^2 \,\mathrm{d}\,\gamma(x, y)}, \tag{1}$$

where the infimum is taken over all transport plans γ, i.e., probability measures γ on $E \times E$ with marginals μ and ν. This infimum is attained because (E, d) is Polish. Crucially, $(\mathcal{W}_2(E), d_W)$ is itself a Polish space, which allows for the definition of a Wasserstein space over it, denoted by $(\mathcal{W}_2(\mathcal{W}_2(E)), d_{\mathbb{W}})$. A Wasserstein barycenter $\mu_{\mathbb{P}} \in \mathcal{W}_2(E)$ is a barycenter of a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(E))$, meaning $\mathbb{P}$ is a measure on measures in $\mathcal{W}_2(E)$. Such a barycenter $\mu_{\mathbb{P}}$ is thus a probability measure on E. This thesis focuses on establishing regularity properties of $\mu_{\mathbb{P}}$, such as its absolute continuity or singularity with respect to a reference measure on E, by leveraging geometric properties of the underlying space (E, d).

A primary motivation for studying Wasserstein barycenters is their remarkable ability to preserve certain geometric features when averaging data represented as probability distributions. This geometric fidelity is well-illustrated by the following example: the barycenter $\mu_{\mathbb{P}}$ of a finite collection of centered Gaussian measures $\{\mu_i\}_{i=1}^n$ on $\mathbb{R}^m$, with $\mathbb{P} = \sum_{i=1}^n \lambda_i \, \delta_{\mu_i}$ for non-negative weights λ_i summing to one, is itself a unique centered Gaussian measure [1, Theorem 6.1]. This property, capturing structural elements during averaging, has fueled increasing interest in applying Wasserstein barycenters across diverse fields such as image processing, machine learning, and statistics; see [78] for a survey of such applications. Although this thesis focuses on theoretical aspects, understanding the geometric properties of Wasserstein barycenters, including their regularity, is crucial to underpinning their effective and reliable use in practical settings. Moreover, investigating barycenter regularity aligns with the broader program of exploring the rich geometric structures inherent in optimal transport theory. This field has produced powerful concepts, notably the synthetic theory of Ricci curvature bounds for metric measure spaces [105, Part III]. This thesis utilizes tools developed from the geometric study of optimal transport, such as displacement convexity and related variational techniques. Conversely, we anticipate that our findings and the methodologies employed will offer further insights into the geometric analysis of Wasserstein spaces and the behavior of their barycenters.

Absolutely continuous Wasserstein barycenters on manifolds

Early investigations in optimal transport concerning the regularity of Wasserstein barycenters often focused on displacement interpolations. Namely, given two probability measures μ, ν in the Wasserstein space $\mathcal{W}_2(M)$ over a complete Riemannian manifold (M, d_{g}), any minimal geodesic from μ to ν consists of points μ_λ, which are barycenters of the measure $(1-\lambda)\delta_\mu + \lambda\, \delta_\nu \in \mathcal{W}_2(\mathcal{W}_2(M))$ as λ varying in $[0, 1]$. These barycenters μ_λ are termed displacement interpolations (or McCann interpolants), and their absolute continuity, under various (and sometimes generalized) conditions, have been extensively studied [72, 12, 36, 38, 105].

Agueh and Carlier [1] initiated the study of Wasserstein barycenters for finite collections of probability measures on Euclidean spaces. These barycenters are solutions to the minimization problem:

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

They established the existence of such barycenters constructively via a dual formulation and demonstrated that if at least one of the input measures μ_i is absolutely continuous with a bounded density function, then the unique barycenter inherits this absolute continuity. Kim and Pass [58] extended this line of inquiry to Wasserstein barycenters on compact Riemannian manifolds M, reaching similar conclusions. Their framework accommodates general probability measures $\mathbb{P}$ on $\mathcal{W}_2(M)$, provided $\mathbb{P}$ assigns positive mass to the set of absolutely continuous measures whose densities are uniformly bounded from above. The absolute continuity of Wasserstein barycenters was indispensable for their subsequent study of Jensen-type inequalities. Agueh and Carlier's results were later generalized by Jiang [52] to compact Alexandrov spaces with curvature bounded from below.

To contextualize our contributions to the geometric investigation of absolute continuity, we first outline some pertinent known properties of Wasserstein barycenters. For a proper metric space (E, d), the existence of a Wasserstein barycenter $\mu_{\mathbb{P}}$ for a finitely supported measure $\mathbb{P} = \sum_{i=1}^{n} \lambda_i \, \delta_{\mu_i}$ (c.f. formulation (2)) is closely related to a multi-marginal optimal transport problem (see Definition 2.11 for more details). Specifically, an optimal multi-marginal plan γ is a minimizer for

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

where Θ denotes the set of all multi-marginal transport plans (probability measures on E^n) with prescribed marginals $\mu_1, \dots, \mu_n$ in this given order. A barycenter selection map is a measurable map $B : E^n \to E$ sending $(x_1, \dots, x_n)$ to a barycenter of $\sum_{i=1}^{n} \lambda_i \, \delta_{x_i}$, i.e.,

$$\sum_{i=1}^{n} \lambda_i \, d(B(x_1, \dots, x_n), x_i)^2 = \min_{y \in E} \sum_{i=1}^{n} \lambda_i \, d(y, x_i)^2.$$

If γ is a multi-marginal optimal transport plan and B is a barycenter selection map, then the push-forward measure $\mu_{\mathbb{P}} := B_{\#}\gamma$ is a barycenter of $\mathbb{P}$. For a general measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(E))$, the existence of its barycenter $\mu_{\mathbb{P}}$ is typically established by approximating $\mathbb{P}$ with a sequence of finitely supported measures $\{\mathbb{P}_j\}$, leveraging the guaranteed existence of a barycenter of $\mathbb{P}_j$, which in turn follows from the existence of multi-marginal optimal transport plans and barycenter selection maps. Thanks to the consistency of Wasserstein barycenters [62], if $d_{\mathbb{W}}(\mathbb{P}_j, \mathbb{P}) \to 0$ and $\mu_{\mathbb{P}_j}$ are corresponding barycenters, then there exists a converging subsequence of $\{\mu_{\mathbb{P}_j}\}$, and its limit is a barycenter of $\mathbb{P}$. Moreover, on a Riemannian manifold (M, d_{g}), if $\mathbb{P}$ assigns positive mass to the set of absolutely continuous measures, the functional $\mu \mapsto \int_{\mathcal{W}_2(M)} d_W(\mu, \nu)^2 \, \mathrm{d}\, \mathbb{P}(\nu)$ is strictly convex [90, Theorem 7.19]. This strict convexity implies the uniqueness of its minimizer $\mu_{\mathbb{P}}$, the Wasserstein barycenter. With these preliminaries, we state a key result by Kim and Pass [58, Theorem 6.2] that serves as a crucial reference point for our work:

Theorem 0.1 (Kim and Pass' result on absolute continuity)**.** *Let (M, d_{g}) be a* compact *Riemannian manifold. For a positive number $L > 0$, denote by $\mathcal{A}_L$ the set of absolutely continuous probability measures (with respect to the volume measure* Vol*) on M whose density functions are* bounded from above *by L. If a measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(M))$ satisfies $\mathbb{P}(\mathcal{A}_L) > 0$ for some $L > 0$, then $\mathbb{P}$ has a unique barycenter $\mu_{\mathbb{P}}$, which is itself absolutely continuous with a bounded density function.*

The proof of Theorem 0.1 by Kim and Pass involves applying their prior results on multi-marginal optimal transport [57] to finitely supported measures. They subsequently establish a uniform upper bound on the densities of barycenters for the approximating sequence, a step where both the compactness of M and a lower Ricci curvature bound are utilized. While Theorem 0.1 marks a significant advance, its reliance on the compactness of M and the strong assumption that $\mathbb{P}$ assigns positive mass to measures with uniformly bounded densities ($\mathbb{P}(\mathcal{A}_L) > 0$) motivates our objective: to establish a more direct link between lower Ricci curvature bounds and the absolute continuity of Wasserstein barycenters, potentially under relaxed assumptions.

To address these limitations, this thesis introduces a novel approach centered on displacement functionals. These functionals assign to an absolutely continuous probability measure $\mu = f \cdot \mathrm{Vol}$ the quantity $\mathcal{G}(\mu) := \int_M G(f)\,\mathrm{d}\,\mathrm{Vol}$. The utility of $\mathcal{G}$ is tied to the properties of the function $G : \mathbb{R}^+ \to \mathbb{R}$, particularly its convexity or growth conditions. Prominent examples, such as those with $G(\rho) = \rho \log \rho$ or $G(\rho) = -n\,\rho^{1-1/n}$ ($n \in \mathbb{N}^*$), are intrinsically linked to lower Ricci curvature bounds; indeed, synthetic definitions of these bounds on metric measure spaces often rely on the convexity properties of such functionals along Wasserstein geodesics [69, 96, 97]. Drawing inspiration from their role in encoding geometric information like lower Ricci curvature bounds, our strategy involves choosing a specific class of functions G. Since a finite value of $\mathcal{G}(\mu)$ implies the absolute continuity of μ, the core element of our method is to establish an effective upper bound for $\mathcal{G}(\mu_{\mathbb{P}})$, the functional evaluated at the Wasserstein barycenter $\mu_{\mathbb{P}}$. Precisely, for a finitely supported measure $\mathbb{P} = \sum_{i=1}^n \lambda_i\,\delta_{\mu_i}$ where each μ_i has compact support and only a subset (say, the first k) are absolutely continuous, our approach yields an upper bound for $\mathcal{G}(\mu_{\mathbb{P}})$ in terms of the convex combination $\sum_{i=1}^k \lambda_i\,\mathcal{G}(\mu_i)$ plus some additional, well-controlled terms (Proposition 4.3). This result is notably different from a similar inequality by Kim and Pass [58, Theorem 7.11], where a finite upper bound requires all measures μ_i in the support of $\mathbb{P}$ to be absolutely continuous. Consequently, our capacity to handle mixtures that include potentially singular measures significantly expands the applicability of using such functionals to deduce barycenter regularity. This refined control is vital for establishing the absolute continuity of Wasserstein barycenters, since it ensures $\mathcal{G}(\mu_{\mathbb{P}})$ remains bounded when transitioning from finitely supported $\mathbb{P}$ to general measures via approximation arguments. Achieving this bound relies on two of our key technical contributions: the derivation of a Hessian equality for Wasserstein barycenters (Theorem 4.1) and the application of new estimates in proving the aforementioned upper bounds of $\mathcal{G}(\mu_{\mathbb{P}})$.

These considerations lead to the following intermediate proposition (Proposition 4.9). Here, the set $\mathsf{B}(G, L)$ comprises absolutely continuous measures $\mu = f \cdot \mathrm{Vol}$ such that $\mathcal{G}(\mu) \le L$, for some $L > 0$ and a function G specified by Definition 4.7.

Proposition 0.2. *Let (M, d_{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 $\mathsf{B}(G, L)$ defined in Definition 4.7 with respect to the volume measure on M, then the unique barycenter of $\mathbb{P}$ is absolutely continuous.*

To refine this result further, we leverage tools from functional analysis and Souslin space theory. We revisit a modified de la Vallée Poussin criterion (Theorem 4.13) to connect the condition $\mathbb{P}(\mathsf{B}(G, L)) > 0$ with conditions involving the $\sigma(L^1, L^\infty)$ weak topology on the densities of absolutely continuous measures. Precisely, the closed set $\mathsf{B}(G, L)$ in Proposition 0.2 can be replaced by a compact set with respect to the weak topology. This shift towards topological properties of sets of absolutely continuous measures (or their densities) leads us to employ the Souslin space theory. Though not commonly seen in the literature of optimal transport, the Souslin space theory provides helpful tools to find connections between different topologies from a measure theoretical viewpoint. Note that, since $\mathbb{P}$ is a Radon measure on the Polish space $\mathcal{W}_2(M)$, if $\mathbb{P}$ assigns positive mass to the (Borel) set of absolutely continuous measures, it must assign positive mass to some compact subset of these absolutely continuous measures (in the $\mathcal{W}_2(M)$ topology). The aforementioned tools help demonstrate that such a compactness result also holds for the weak topology. This line of argument culminates in the first main result of this thesis:

Theorem 0.3. *Let (M, d_{g}) be a complete Riemannian manifold with a lower Ricci curvature bound. If a measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(M))$ assigns positive mass to the set of absolutely continuous probability* *measures with respect to the volume measure on M, then the unique barycenter of $\mathbb{P}$ is absolutely continuous.*

This result significantly extends Theorem 0.1 in two key aspects. Firstly, it establishes the absolute continuity of the barycenter $\mu_{\mathbb{P}}$ under the considerably weaker condition that $\mathbb{P}$ merely assigns positive mass to the set of absolutely continuous measures. This requirement, often a natural one for ensuring the uniqueness of $\mu_{\mathbb{P}}$, represents a crucial relaxation from Theorem 0.1, which demands that $\mathbb{P}$ gives mass to measures with uniformly bounded densities (i.e., $\mathbb{P}(\mathcal{A}_L) > 0$). Secondly, our result holds for general complete manifolds with a lower Ricci curvature bound, thereby relaxing the compactness assumption of the prior theorem. Consequently, even when applied to the compact setting originally considered by Kim and Pass (as compact manifolds are indeed complete and possess a lower Ricci curvature bound), our proposition provides a stronger statement due to this less restrictive condition on $\mathbb{P}$.

Singular Wasserstein barycenters on metric trees

Having established that a lower Ricci curvature bound ensures the absolute continuity of Wasserstein barycenters, a natural subsequent inquiry concerns how their singularity relates to other geometric structures. Inspired by [50], we begin with a concrete example on the tripod formed by attaching three copies of the unit interval $[0, 1]$ at a common endpoint 0 (Figure 1). Let ν_1, ν_2, ν_3 be three

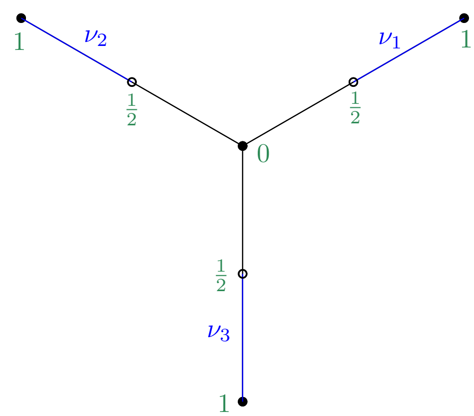


Figure 1: $\mathbb{P} = \sum_{i=1}^{3} \frac{1}{3}\, \delta_{\nu_i}$ on the tripod

probability measures, each supported in the sub-interval $[\frac{1}{2}, 1]$ of a distinct branch of this tripod. As detailed in Proposition 6.59, a calculation involving the barycenter selection map reveals that the unique Wasserstein barycenter of $\mathbb{P} := \sum_{i=1}^{3} \frac{1}{3}\, \delta_{\nu_i}$ is $\mu_{\mathbb{P}} = \delta_0$, a Dirac measure concentrated at the central vertex 0.

To contextualize this example and relate it to our previous findings on absolute continuity, we briefly introduce the setting of metric trees. A metric tree is a geodesic metric space $\Gamma = (V, E, d_l)$, where V is the set of vertices, E is the set of edges, and $d_l(x, y)$ is the length of the unique shortest path connecting any two points $x, y \in \Gamma$. The canonical reference measure on Γ is the one-dimensional Hausdorff measure $\mathcal{H}$, which coincides with the Lebesgue measure on each edge and assigns zero measure to the vertices V. The tripod in our example is a metric tree with three edges of length 1 and four vertices.

Consider the specific case where each ν_i in the tripod example is the uniform probability measure on the interval $[\frac{1}{2}, 1]$ of its respective branch. Each ν_i is then absolutely continuous with respect to $\mathcal{H}$. Consequently, the measure $\mathbb{P} = \frac{1}{3}\sum \delta_{\nu_i}$ is supported entirely on absolutely continuous measures. Nevertheless, its barycenter $\mu_{\mathbb{P}} = \delta_0$ is singular with respect to $\mathcal{H}$. While a tripod is not a smooth manifold, we can still infer from various generalized notions of curvature bounds that around the common vertex 0, the curvature of the tripod is not bounded from below. This observation suggests a link between the failure of such curvature bounds and the emergence of singular barycenters. The flexibility in choosing the measures ν_i further hints at a rich, yet potentially tractable, structure for singular Wasserstein barycenters on metric trees, motivating our focused study in this setting.

Auxiliary techniques for metric trees

To systematically investigate this phenomenon, we develop and utilize two key auxiliary tools. The first one is a localization principle for Wasserstein barycenters, which we term the restriction property (Corollary 5.4):

Theorem 0.4 (Restriction property of Wasserstein barycenters)**.** *Let (E, d) be a proper metric space, and let $\mu_{\mathbb{P}}$ be a Wasserstein barycenter of a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(E))$. Given an equality $\mu_{\mathbb{P}} = \lambda\,\mu^1 + (1-\lambda)\mu^2$ with $\mu^i \in \mathcal{W}_2(E)$ for $i = 1, 2$ and $\lambda \in (0, 1)$, there exist two probability measures $\mathbb{Q}^1, \mathbb{Q}^2$ such that μ^i is a barycenter of $\mathbb{Q}^i$ for $i = 1, 2$. Furthermore, $\mathbb{Q}_1$ and $\mathbb{Q}_2$ inherit the following properties from $\mathbb{P}$ concerning absolute continuity with respect to any given reference measure η on E:*

1. *If $\mathbb{P}$ assigns positive mass to the set of measures absolutely continuous with respect to η, then $\mathbb{Q}_1$ and $\mathbb{Q}_2$ also assign positive mass to this set.*
2. *If $\mathbb{P}$ is supported entirely in the set of measures absolutely continuous with respect to η, then $\mathbb{Q}_1$ and $\mathbb{Q}_2$ are also supported entirely on this set.*

This restriction property (Theorem 0.4), particularly when η is a canonical reference measure (e.g., the volume measure on a Riemannian manifold or $\mathcal{H}$ on a metric tree), provides a foundation for local-to-global arguments, facilitating the extension of results like Theorem 0.3 to more complex scenarios.

The second auxiliary tool is a reduction technique specifically designed for optimal transport problems on metric trees. For any oriented edge $\vec{\mathbf{e}}$ of the metric tree Γ, we define a reduction map $T^{\vec{\mathbf{e}}} : \Gamma \to \mathbb{R}$. This map effectively "flattens" the tree into the real line by identifying $\vec{e}$ with an interval and mapping the rest of Γ accordingly (see Figure 2 and Definition 6.20).

Thanks to the c-cyclical monotonicity characterization of optimal transport plans, we establish that these reduction maps preserve Wasserstein distances under certain conditions:

Theorem 0.5 (Reduction property of optimal transport on metric trees)**.** *Let $\Gamma = (V, E, d_l)$ be a metric tree. Fix an oriented edge $\vec{\mathbf{e}}$ of Γ and let $T^{\vec{\mathbf{e}}} : \Gamma \to \mathbb{R}$ be the reduction map associated to $\vec{\mathbf{e}}$ (Definition 6.20). For two given probability measures $\mu, \nu \in \mathcal{W}_2(\Gamma)$, if μ is supported in the edge $\vec{\mathbf{e}}$, then*

$$d_W(\mu, \nu) = d_W({T^{\vec{\mathbf{e}}}}_{\#}\mu, {T^{\vec{\mathbf{e}}}}_{\#}\nu),$$

where d_W denotes both the Wasserstein metrics on $\mathcal{W}_2(\Gamma)$ and on $\mathcal{W}_2(\mathbb{R})$.

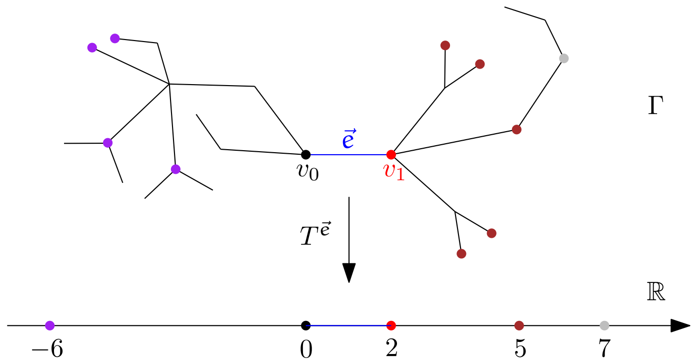


Figure 2: Illustrative example of the reduction map $T^{\vec{e}}$.

Wasserstein barycenters on the real line

The combination of the reduction technique (Theorem 0.5) and the restriction property (Theorem 0.4) guides our strategy: first, investigate properties of Wasserstein barycenters on the real line $\mathbb{R}$, and then apply the auxiliary techniques to extend the findings to metric trees.

On the real line, the Wasserstein space $(\mathcal{W}_2(\mathbb{R}), d_W)$ possesses a well-known linear structure. It can be isometrically embedded into the Hilbert space $L^2([0,1])$ by mapping a measure $\mu \in \mathcal{W}_2(\mathbb{R})$ to its quantile function $f_\mu^{-1} : [0,1] \to \overline{\mathbb{R}}$. The quantile function here is the (generalized) right-continuous inverse of the distribution function $f_\mu(t) := \mu((-\infty, t])$. This linear structure leads to an explicit formula for the Wasserstein barycenter $\mu_\mathbb{P}$ of any $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$: its quantile function is the $\mathbb{P}$-average of the quantile functions of the measures in its support:

$$f_{\mu_\mathbb{P}}^{-1}(t) = \int_{\mathcal{W}_2(\mathbb{R})} f_\nu^{-1}(t)\, \mathrm{d}\,\mathbb{P}(\nu), \quad \forall\, t \in [0,1]. \tag{3}$$

To analyze singularity properties using this formula, particularly for measures supported in $[0,1]$, we introduce the concept of dual measure $\widetilde{\mu}$.

Definition 0.6. Let μ be a probability measure supported in $[0,1]$. Its dual measure $\widetilde{\mu}$ is the probability measure whose distribution function $f_{\widetilde{\mu}}$ is given by the quantile function of μ:

$$f_{\widetilde{\mu}}(t) = f_\mu^{-1}(t), \quad \text{for } t \in (0,1). \tag{4}$$

Dual measures share many regularity and singularity properties. It can be verified that $\widetilde{\mu}$ is also supported in $[0,1]$ and that the duality is involutive, i.e., $\widetilde{\widetilde{\mu}} = \mu$. Crucially, as shown in Proposition 6.33 and Theorem 6.40, μ exhibits certain types of singularities if and only if $\widetilde{\mu}$ does. These include being finitely supported, countably supported, or singular with respect to the Lebesgue measure $\mathcal{L}^1$. By combining the properties of dual measures with the barycenter formula (3), we derive the following rigidity properties for Wasserstein barycenters on $\mathbb{R}$.

Theorem 0.7 (Rigid Properties of Wasserstein Barycenters on $\mathbb{R}$)**.** *Let $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$ be a probability measure, and let $\mu_\mathbb{P}$ be its Wasserstein barycenter. We say a property $\mathcal{Q}$ is a rigid property of $\mu_\mathbb{P}$ if the following implication holds,*

$$\mu_\mathbb{P} \textit{ satisfies property } \mathcal{Q} \implies \nu \textit{ satisfies property } \mathcal{Q} \textit{ for } \mathbb{P}\textit{-almost every } \nu.$$

The following properties of $\mu_{\mathbb{P}}$ are rigid:

1. *Being a Dirac measure.*
2. *Having compact support.*
3. *Being singular (with respect to the Lebesgue measure).*
4. *Having support of Lebesgue measure zero (i.e., being supported in a negligible set).*
5. *Being not absolutely continuous.*

These rigid properties mean that if the barycenter $\mu_{\mathbb{P}}$ is, for example, singular, then almost all measures ν in the support of $\mathbb{P}$ must also be singular. This is a powerful constraint on the measures being averaged.

Characterizing singular Wasserstein barycenters on metric trees

Equipped with these tools, we can now describe the nature of singular Wasserstein barycenters on metric trees. As an illustration of how properties of Wasserstein barycenters are extended from $\mathbb{R}$ to trees, recall that Theorem 0.3, when applied to $\mathbb{R}$ (which has zero Ricci curvature), states that if $\mathbb{Q} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$ gives mass to absolutely continuous measures, then its barycenter $\mu_{\mathbb{Q}}$ is also absolutely continuous. By applying the restriction property and the reduction technique, we generalize this to obtain the following partial regularity result on metric trees (Theorem 6.28):

Proposition 0.8. *Let $\Gamma = (V, E, d_l)$ be a metric tree. Let $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$ be a measure that assigns positive mass to the set of measures on Γ that are absolutely continuous (with respect to $\mathcal{H}$). If $\mu_{\mathbb{P}}$ is a barycenter of $\mathbb{P}$, then the restriction of $\mu_{\mathbb{P}}$ to the interior of any edge $e \in E$ is absolutely continuous. Consequently, if $\mu_{\mathbb{P}}$ is not absolutely continuous, its singular part must be supported in the set of vertices V, which is a weighted sum of Dirac measures at these vertices.*

Proposition 0.8 reveals a general principle for characterizing Wasserstein barycenters $\mu_{\mathbb{P}}$ on metric trees: the behavior of $\mu_{\mathbb{P}}$ on the interior of edges often mirrors that of barycenters on $\mathbb{R}$, while its behavior at vertices requires separate analysis, confirming the observation from the tripod example. In Section 6.5, we present a method for determining the mass of $\mu_{\mathbb{P}}$ at vertices, which involves applying the reduction technique to all edges incident to a given vertex.

Our general strategy for describing Wasserstein barycenters on metric trees then follows two steps:

1. Analyze the mass distribution at vertices using reduction techniques applied to edges incident to each vertex (Section 6.5).
2. Analyze the barycenter's restriction to the interior of each edge by reducing the problem to the real line, utilizing formula (3) and the rigid properties (Theorem 0.7) to characterize the possible types of measures.

Several examples illustrating distinctive features of Wasserstein barycenters on metric trees, derived using this approach, are presented in Section 6.6.

Summary and future research directions

This thesis advances the geometric study of optimal transport and Wasserstein barycenters through several key contributions.

1. *Hessian Equality for Wasserstein Barycenters*: In Theorem 4.1, we establish a novel Hessian equality for Wasserstein barycenters of finitely supported measures. This result provides a rigorous geometric underpinning for the intuition that the weighted sum of "tangent vectors" from the barycenter to the supporting measures vanishes, reinforcing the analogy of Wasserstein spaces as infinite-dimensional Riemannian manifolds. The introduction of the "approximate Hessian" (Section 1.3.2) in its derivation also offers a versatile tool for future differential investigations of Wasserstein barycenters.

2. *Absolute Continuity of Wasserstein Barycenters under Relaxed Assumptions*: Our main result on absolute continuity, Theorem 4.5 (restated as Theorem 0.3), significantly extends the work of Kim and Pass (Theorem 0.1). It demonstrates that a lower Ricci curvature bound is sufficient for the absolute continuity of the barycenter $\mu_{\mathbb{P}}$ if the measure $\mathbb{P}$ merely assigns positive mass to the set of absolutely continuous measures. This clarifies the crucial role of Ricci curvature, independent of the assumptions regarding compactness or the uniform boundedness of densities. The proof introduces innovative displacement functionals and incorporates Souslin space theory, enriching the analytical toolkit for optimal transport research.

3. *Restriction Property for Wasserstein Barycenters*: The restriction property (Proposition 5.2 and Corollary 5.4, summarized in Theorem 0.4) establishes a powerful localization principle for Wasserstein barycenters on proper metric spaces. As demonstrated by its application in various proofs (e.g., Theorem 6.28), this technique enables local-to-global arguments, offering a way to tackle problems in singular settings like metric trees by breaking them down into more manageable components.

4. *Systematic Study of Wasserstein Barycenters on Metric Trees*: This work initiates a systematic investigation into singular Wasserstein barycenters on metric trees (Chapter 6). For the real line, we introduce concepts such as dual measures and rigid properties, shedding new light on the fine structure of barycenters in this fundamental setting. Our novel reduction technique provides an intuitive and effective approach to optimal transport on metric trees, with promising implications for extensions to general metric graphs. The illustrative examples in Section 6.6 reveal intricate behaviors, including the non-uniqueness of Wasserstein barycenters. This phenomenon sharply distinguishes metric trees from smoother settings, where uniqueness is guaranteed under analogous conditions, thereby paving the way for deeper explorations.

The research presented in this thesis naturally opens up several promising avenues for future investigation.

Extending displacement functional arguments to metric measure spaces

Our proof of Theorem 0.3 introduces a novel approach using displacement functionals that encapsulate the lower Ricci curvature bound within an inequality for Wasserstein barycenters. This mirrors the standard definition of lower Ricci curvature bounds for metric measure spaces (MMS), which often involves analogous inequalities (e.g., convexity of entropy functionals) along Wasserstein geodesics. This structural parallel suggests that many of our arguments could be adapted to the general MMS setting. Recent advancements have indeed seen the proposal and study of barycenter curvature-dimension conditions for MMS [46, 47]. However, these developments are primarily based on the Wasserstein Jensen's inequality from [58, Theorem 7.11], which, as discussed, essentially requires all measures in the support to be absolutely continuous. Our approach, which circumvents this limitation, could therefore offer a valuable alternative. Given that the measure-theoretic components of our arguments, particularly those involving Souslin space theory, readily apply to general Polish spaces (the underlying framework for many MMS), the principal hurdle in such an extension appears to be the establishment of our core displacement functional inequality (Proposition 4.3) for barycenters of finitely supported measures $\mathbb{P}$ on an MMS.

In Chapter 3, we reformulate Kim and Pass' proof that $\mu_{\mathbb{P}}$ is absolutely continuous if $\mathbb{P} = \sum_{i=1}^{n} \lambda_i \, \delta_{\mu_i}$ with μ_1 being absolutely continuous, to clarify its validity for non-compact manifolds and its dependence on Riemannian structure. This proof relies significantly on the Brenier–McCann theorem, which characterizes optimal transport maps via gradients of c-concave potentials ϕ (e.g., as $\exp(-\nabla\phi)$, c.f. Theorem 1.27). In particular, the Hessian equality for Wasserstein barycenters (Theorem 4.1), expressed in terms of ϕ, is vital for deriving our displacement functional inequality. A significant challenge, therefore, is to extend these arguments to MMS that lack such smooth Riemannian structures and direct analogues of these tools.

Quantitative estimates for barycenter densities

The framework of displacement functionals offers the potential to derive qualitative estimates for the density functions of absolutely continuous Wasserstein barycenters. By selecting appropriate functionals, one might obtain bounds on these densities, thus providing information beyond mere absolute continuity. This approach is exemplified by results such as the L^∞ density bound for displacement interpolations in $\mathrm{CD}(0, N)$ spaces (c.f. [105, Theorem 30.20]), suggesting similar estimates could be attainable for general barycenters under suitable curvature conditions.

Necessity of curvature bounds and the role of branching

While this thesis establishes that a lower Ricci curvature bound is a sufficient condition for the absolute continuity of Wasserstein barycenters on Riemannian manifolds, its necessity remains an open question. Despite efforts, we have not identified a Riemannian manifold lacking a global lower Ricci curvature bound where barycenters of absolutely continuous measures exhibit singularity. This raises the possibility that a weaker condition, perhaps related to the non-branching property of manifolds, might suffice.

Specifically, we conjecture that on non-branching metric spaces, the absolute continuity of $\mu_{\mathbb{P}}$ might hold even without a global lower Ricci curvature bound, potentially provable by contradiction using an enhanced version of our restriction property (Theorem 0.4). If true, this would imply that the emergence of singular Wasserstein barycenters on metric trees is fundamentally linked to their branching structure, distinguishing them from (non-branching) Riemannian manifolds.

Generalizing reduction techniques to metric graphs

Optimal transport problems rarely admit explicit solutions, rendering techniques that simplify them highly valuable. Our reduction technique for metric trees (Theorem 0.5) proved effective. A promising avenue for future work is to generalize this technique to broader classes of spaces, such as general metric graphs (which may contain cycles). Developing such a generalization would not only provide tools for solving concrete optimal transport problems on graphs in practice, but would also deepen our understanding of optimal transport by expanding the repertoire of settings where (at least partial) computations are feasible. Our detailed exposition of the reduction map $T^{\vec{e}}$ and its properties in Section 6.2 was intentionally presented to aid such future extensions.

Chapter 1

Prerequisites and notation

Due to varying implicit assumptions in different references, concepts in metric geometry and measure theory are often defined with subtle differences. In this chapter, we aim to clarify the usage of these terminologies and establish the notational conventions that will be consistently followed in this document.

We begin with some definitions in set theory. Symbols $\mathbb{Z}, \mathbb{N}, \mathbb{N}^*, \mathbb{R}, \mathbb{C}$ are reserved to denote respectively the set of integers, natural numbers with 0 included, natural numbers with 0 excluded, real numbers and complex numbers. A *map* f from a set X to a set Y is an assignment of one element $f(x)$ of Y to each element x of X. The set X is called the *domain* of f, the set Y is called the *codomain* of f, and $f(X) := \cup_{x \in X} f(x) \subset Y$ is called the *image* of f. We shall use $\mathrm{Id} : X \to X$ to denote the identity map. When the codomain Y of f is a subset of the Euclidean spaces $\mathbb{R}^m$ ($m \in \mathbb{N}^*$), we also call f a *function*. In certain instances, which will be clearly indicated, functions are permitted to take the values $+\infty$ and $-\infty$.

For a real number $x \in \mathbb{R}$, we define x as *positive* if $x > 0$ and *negative* if $x < 0$. To explicitly stress that x is not equal to 0, we also use the terms strictly positive or strictly negative. $\mathbb{R}_+ := [0, +\infty)$ is the set of non-negative numbers, $\mathbb{R}_- := (-\infty, 0]$ is the set of non-positive numbers. Given two real numbers $x, y \in \mathbb{R}$, we say x is smaller than or less than y if $x \leq y$, and x is bigger than or larger than y if $x \geq y$. For a real-valued function f defined on a subset $X \subset \mathbb{R}$ of the real numbers, f is defined as *increasing* (or *non-decreasing*) if $x_1, x_2 \in X$ with $x_1 \leq x_2$ implies $f(x_1) \leq f(x_2)$. Similarly, f is *decreasing* (or *non-increasing*) if $x_1 \leq x_2$ implies $f(x_1) \geq f(x_2)$. The definitions of strictly increasing and strictly decreasing functions are obtained by replacing the non-strict inequalities with their corresponding strict counterparts.

For a function f defined on a subset $A \subset \mathbb{R}$ of the real line, its *right limit* at some point $x \in \mathbb{R}$, denoted by $\lim_{y \downarrow x} f(y)$, is defined as the limit of $f(y)$ as y converges to x through values in A that are strictly greater than x. Such limits are only defined for x in the set $\{x \in \mathbb{R} \mid \exists\, y_i \in A, y_i > x, \text{ for } i \in \mathbb{N}^* \text{ s.t. } \lim_{i \to \infty} y_i = x\}$. The *left limit* $\lim_{y \uparrow x} f(y)$ of f at x is defined analogously.

A set is *countable* if it can be mapped bijectively to $\mathbb{N}$, and is thus always infinite. A set is *uncountable* if it is infinite but not countable. For a given set X, we denote by $2^X := \{A \mid A \subset X\}$ the set of all subsets of X.

For clarity of notation, we employ brackets [] and parentheses () to divide cluster of symbols into meaningful sub-groups. Unless explicitly stated otherwise, these symbols carry no additional semantic meaning.

1.1 Length spaces and Riemannian manifolds

Given a set E, a *metric* d on E is a function $d : E \times E \to \mathbb{R}$ satisfying the following three conditions for all points $x, y, z \in E$: a) positiveness: $d(x, y) > 0$ if $x \neq y$, and $d(x, x) = 0$; b) symmetry: $d(x, y) = d(y, x)$; c) triangle inequality: $d(x, z) \leq d(y, x) + d(y, z)$. For $x, y \in E$, the non-negative number $d(x, y)$ is the called *distance* between x and y. The ordered pair (E, d) is referred to as a *metric space*. For a point $x \in E$ and a non-empty subset $A \subset E$, we define $d(x, A) := \inf_{y \in A} d(x, y)$. Denote by $\overline{B}(x, r) := \{y \in E \mid d(x, y) \leq r\}$ the closed metric ball center at x with radius r.

Consider a metric space (E, d). We recall that E is *proper* if every closed and bounded subset of E is compact, *separable* if it contains a countable dense subset, and *complete* if every Cauchy sequence in E converges to a point within E. Since compact metric spaces are complete and separable [20, Theorem 9.4], proper metric spaces inherit these properties [93, Corollary 2.3.32]. A *Polish space* is defined as a topological space that is homeomorphic to a complete and separable metric space. When we write (for example, in assumptions) that (E, d) is a Polish metric space we specifically mean that the metric d on E makes E a complete and separable metric space. In particular, proper metric spaces are Polish metric spaces.

Length spaces

Let (E, d) be a metric space. A *curve* in E is a continuous map γ from a compact interval $[a, b] \subset \mathbb{R}$ to E. We say that γ *joins* (or *connect*s) its *endpoint*s $\gamma(a), \gamma(b) \in E$, or γ is a curve *from* $\gamma(a)$ *to* $\gamma(b)$.

Definition 1.1 (Length of curves in metric spaces)**.** Let (E, d) be a metric space. The length $L_d(\gamma)$ of a curve $\gamma : [a, b] \mapsto E$ is

$$L_d(\gamma) := \sup_{a = t_0 \leq t_1 \leq \cdots \leq t_n = b} \sum_{i=1}^{n} d(\gamma(t_{i-1}), \gamma(t_i)), \tag{1.1}$$

where the supremum is taken over all possible partitions (no bound on n) with $a = t_0 \leq t_1 \leq \cdots \leq t_n = b$.

The length of γ is either a non-negative number or it is infinite. The curve γ is said to be *rectifiable* if its length is finite. Length spaces and geodesic spaces are defined via the *intrinsic metric*, which associates two points with the infimum of the lengths of all curves joining them.

Definition 1.2 (Length spaces and geodesic spaces)**.** Let (E, d) be a metric space. E is a length space (or an intrinsic metric space) if for any two points $x, y \in E$,

$$d(x, y) = \inf_{\gamma \text{ from } x \text{ to } y} L_d(\gamma), \tag{1.2}$$

where the infimum is taking over all curves γ joining x, y, and $L_d(\gamma)$ denotes the length of γ. E is a geodesic space (or a strictly intrinsic metric space) if the infimum in (1.2) is always reached by some rectifiable curve joining x and y.

For length spaces, we recall the following two notable properties: a) a complete locally compact length space is geodesic [23, Theorem 2.5.23]; b) given a locally compact length space, the Hopf–Rinow–Cohn-Vossen theorem states that it is complete if and only if it is proper [23, Theorem 2.5.28]. Moreover, we distinguish shortest paths and geodesics.

Definition 1.3 (Shortest paths and geodesics)**.** Let (E, d) be a length space. A curve $\gamma : [a, b] \to E$ is a shortest path if its length is equal to the distance $d(\gamma(a), \gamma(b))$ between its endpoints. The curve γ is a geodesic if it is a locally a shortest path, i.e., for any $t \in [a, b]$, there exists an interval $J := [c, d] \subset [a, b]$ such that $c < t < d$ and the restricted map $\gamma|_J$ is a shortest path.

Riemannian Manifolds

A Riemannian manifold, denoted by (M, g), is composed of a smooth manifold M, which is Hausdorff and second-countable (as defined in [65, Chapter 1]), and a Euclidean inner product g_x defined on each tangent space T_xM at every point $x \in M$. In accordance with McCann's work [73] of optimal transport on manifolds, we shall adhere to the following assumption throughout this document.

Assumption. All Riemannian manifolds (M, g) are assumed to satisfy the following properties:

1. M is an m-dimensional ($m \in \mathbb{N}^*$), connected and smooth manifold without boundary;
2. given a local coordinate system $\{x^i\}_{i=1,2,\ldots,m}$, the metric tensor components $\mathsf{g}_{ij} := \mathsf{g}(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j})$ are $\mathcal{C}^\infty$ smooth (differentiable for all degrees of differentiation) functions of local coordinates.

Assumptions related to compactness and completeness will be explicitly stated.

For a Riemannian manifold (M, g), we introduce the following notation. Denote by d_g the Riemannian distance function of M determined by g [80, §5.3]. As the Riemannian metric tensor g can be reconstructed from the Riemannian distance d_g [80, §5.6.3], we can alternatively denote the Riemannian manifold as (M, d_g), thus highlighting its metric structure. The metric space (M, d_g) is always a locally compact length space. For $x \in M$, we introduce the squared distance function $d_x^2 : M \to \mathbb{R}$, i.e., $d_x^2(y) = d_\mathsf{g}(x, y)^2$.

Denote by Vol the volume measure of M. Given a local chart (φ, U) with coordinate system $\{x^i\}_{i=1,2,\ldots,m}$, the integral of a Vol-integrable function $f : U \to \mathbb{R}$ is defined as [88, §5 of Chapter II],

$$\int_U f \,\mathrm{d}\,\mathrm{Vol} := \int_{\varphi(U)} f \circ \varphi^{-1} \sqrt{\det(\mathsf{g}_{ij})} \circ \varphi^{-1} \,\mathrm{d}\,\mathcal{L}^m,$$

where (g_{ij}) denotes the $m \times m$ matrix with components g_{ij} as previously introduced, and $\mathcal{L}^m$ denotes the Lebesgue measure on $\mathbb{R}^m$. The volume measure Vol coincides with the m-dimensional Hausdorff measure on (M, d_g) [100, Proposition 12.6].

For a tangent vector $u \in T_xM$, denote by $\|u\| := \sqrt{\mathsf{g}(u, u)}$ its norm of the Riemannian metric. Denote by $\exp_x : T_xM \to M$ the exponential map defined on T_xM and by $\exp : TM \to M$ the exponential map defined on the tangent bundle TM. In the rest of this paragraph, suppose in addition that (M, d_g) is a complete Riemannian manifold. The exponential maps are $\mathcal{C}^\infty$ smooth since g is $\mathcal{C}^\infty$ smooth [88, §2 of Chapter 2] [73, Proof of Proposition 6]. For a point $x \in M$, its *tangent cut locus* is the boundary of the set

$$\{u \in T_xM \mid d_\mathsf{g}(\exp_x u, x) = \|u\|\},$$

and its *cut locus* is the image of its tangent cut locus under the exponential map $\exp_x$. The *injectivity domain of* x is the following subset of T_xM,

$$\{t\, u \in T_xM \mid t \in [0, 1), u \in T_xM \text{ is in the tangent cut locus of } x\}.$$

We shall denote by $\mathrm{Cut}(x)$ the cut locus of x. The set $\mathrm{Cut}(x)$ is closed and negligible with respect to the volume measure Vol [66, (a) of Theorem 10.34]. The exponential map $\exp_x$ is a $\mathcal{C}^\infty$-diffeomorphism from the injectivity domain of x to $M \setminus \mathrm{Cut}(x)$ [66, (c) of Theorem 10.34]. Since the gradient of d_y^2 at x is $\nabla d_y^2(x) = -2\exp_x^{-1}(y)$ for $y \notin \mathrm{Cut}(x)$ [73, Proposition 6], d_x^2 is a $\mathcal{C}^\infty$ function on $M \setminus \mathrm{Cut}(x)$. For two points $x, y \in M$, x is in the cut locus of y if and only if y is in the cut locus of x [41, Scholium 3.78].

1.2 Tools from measure theory

A *measurable space* $(\Omega, \mathcal{F})$ is an ordered pair of set Ω and a σ-algebra $\mathcal{F}$ of Ω. The elements of $\mathcal{F}$ are called measurable sets of Ω. A measure on $(\Omega, \mathcal{F})$, or simply a measure on Ω when $\mathcal{F}$ is clearly given, is a set function $\mu : \mathcal{F} \to [0, +\infty]$ that satisfies the condition $\mu(\emptyset) = 0$ and is countably additive (σ-additive) [17, Definition 1.6.1]. This means that for all pairwise disjoint sets $\{A_i\}_{i\in\mathbb{N}^*} \subset \mathcal{F}$, we have $\mu(\cup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i)$, where infinite values are permitted. The trivial function that assigns value 0 to every element of $\mathcal{F}$, also known as null measure, will not be considered. Given a topological space E, we denote by $\mathcal{B}(E)$ its Borel σ-algebra, which is the σ-algebra generated by the open sets of E. Measures defined for the the measurable space $(E, \mathcal{B}(E))$ are called Borel measures.

The Lebesgue measure $\mathcal{L}^m$ on the Euclidean space $\mathbb{R}^m$ is not only defined on the Borel σ-algebra $\mathcal{B}(\mathbb{R}^m)$, but also assigns zero mass to all subsets of negligible Borel sets. We shall denote by $\mathfrak{u}$ the uniform probability measure on $[0, 1]$, which by definition is the restricted Lebesgue measure $\mathcal{L}^1|_{[0,1]}$. In Chapter 6, by singular measures μ, we mean measures on $\mathbb{R}$ that singular with respect to $\mathcal{L}^1$, i.e., μ and $\mathcal{L}^1$ are mutually singular. Apart from the Lebesgue measures, which are defined for Lebesgue measurable sets [17, Definition 1.5.1], we exclusively consider measures without completions. The volume measure Vol on a Riemannian manifold is treated as a Borel measure.

Fix a topological space E and a Borel measure μ on it. Given a measurable subset $A \subset E$, we define that: μ *gives mass to* A if $\mu(A) > 0$; μ is *supported in* A (or μ *assigns full mass to* A) if $\mu(E \setminus A) = 0$; A is the *support* of μ if A is closed, $\mu(E \setminus A) = 0$, and for any open subset O of E, $\mu(O) = 0$ implies $O \subset E \setminus A$; A is an *atom set* [17, Definition 1.12.7] of μ if $\mu(A) > 0$ and for any subset $A' \subset A$, $\mu(A') = 0$ or $\mu(A') = \mu(A)$. A point $x \in E$ is called an *atom* of μ if the singleton $A := \{x\}$ is an atom set of μ. A Borel measure is *atomless* if it has no atom sets, and it is *diffused* if it has no atoms. For separable metric spaces, atomless measures coincide with diffused measures [3, Lemma 3.4, Lemma 12.18]. For any point $x \in E$, the *Dirac measure* at x, denoted by δ_x, is the probability measure with $\{x\}$ being its support. Moreover, on a separable metric space, any Borel measure μ has support [17, Proposition 7.2.9], and we denote it by $\mathrm{supp}(\mu)$.

A finite Borel measure μ on E is called a *Radon measure* if for every Borel set $A \in \mathcal{B}(X)$ and $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset A$ such that $\mu(A \setminus K_\varepsilon) < \varepsilon$. A finite Borel measure on a Polish space is always a Radon measure [17, Theorem 7.1.7] [3, Theorem 12.7].

Recall that a finite Borel measure has at most countably many atoms, which follows directly from the following lemma.

Lemma 1.4. *The sum of any uncountably many strictly positive real numbers must be infinite.*

Proof. Fix a set of strictly positive numbers $\{t_\alpha, \alpha \in A\}$, where $t_\alpha > 0$ and A is an uncountable index set. Define $A_n := \{\alpha \in A \mid t_\alpha > 1/n\}$ for integers $n \geq 1$. Since $A = \cup_{n\geq 1} A_n$, there exists an integer n_0 such that A_{n_0} is an infinite set, otherwise A becomes a countable set. It follows that the sum $\sum_{\alpha\in A} t_\alpha \geq \sum_{\alpha\in A_{n_0}} 1/n_0$ must diverge. $\square$

Measurable selections

A map $f : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ between two measurable spaces is *measurable* (with respect to the σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$) if $f^{-1}(A) \in \mathcal{F}_1$ for $A \in \mathcal{F}_2$. For a set-valued map $\Psi : \Omega_1 \to 2^{\Omega_2}$ whose values are non-empty subsets of Ω_2, $f : \Omega_1 \to \Omega_2$ is a *selection* of Ψ if for any $\omega \in \Omega$, $f(\omega) \in \Psi(w)$.

We shall apply the following widely used measurable selection theorem to construct Wasserstein barycenters. Its proof could be found in [17, Theorem 6.9.3], [3, Theorem 18.13], and [93, Theorem 5.2.1].

Theorem 1.5 (Kuratowski and Ryll-Nardzewski measurable selection theorem)**.** *Let E be a Polish space, and let Ψ be a map defined on a measurable space $(\Omega, \mathcal{F})$ with values in the set of non-empty closed subsets of E. Suppose that for every open set $U \subset E$, we have*

$$\{\omega \in \Omega \mid \Psi(\omega) \cap U \neq \emptyset\} \in \mathcal{F}. \tag{1.3}$$

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

With the help of additional metric assumptions, we can simplify (1.3) as follows [93, Lemma 5.1.2]. We remark that the complement of the set $\{\omega \in \Omega \mid \Psi(\omega) \cap U \neq \emptyset\}$ is not the set $\{\omega \in \Omega \mid \Psi(\omega) \cap (E \setminus U) \neq \emptyset\}$.

Lemma 1.6. *Let (E, d) be a proper metric space, and let Ψ be a map on a measurable space $(\Omega, \mathcal{F})$ with values in the set of subsets of E. If for every compact set $K \subset E$, we have*

$$\{\omega \in \Omega \mid \Psi(\omega) \cap K \neq \emptyset\} \in \mathcal{F},$$

then for every open set $U \subset E$, we have

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

Proof. Observe that for a sequence of subsets $\{A_j\}_{j \in \mathbb{N}^*}$, we have

$$\{\omega \in \Omega \mid \Psi(\omega) \cap A \neq \emptyset\} = \cup_{j \geq 1} \{\omega \in \Omega \mid \Psi(\omega) \cap A_j \neq \emptyset\}, \quad \text{where } A := \cup_{j \geq 1} A_j.$$

Hence, to prove the lemma, it suffices to express any open sets U as a countable union of compact sets. Fix a point $z \in E$, and define $K_j := \{x \in E \mid d(x, z) \leq j \text{ and } d(x, E \setminus U) \geq \frac{1}{j}\}$ for $j \in \mathbb{N}^*$. The equality $U = \cup_{j \geq 1} K_j$ holds for any metric space. Moreover, since (E, d) is a proper metric space, each set K_j is compact as a closed and bounded set. □

Conditional probability measures

For the definition of conditional measures, as given in [17, Definition 10.4.2], we focus on the special case of the product space $E^{n-1} \times E$. This restriction facilitates the introduction of necessary notation for Proposition 3.4 and ultimately aids in its proof.

Definition 1.7 (Conditional probability measures)**.** Let E be a Polish space and let $n \geq 2$ be a positive integer. Denote by $\mathrm{x}' = (x_2, \ldots, x_n) \in E^{n-1}$ the last $n-1$ components of a point $\mathrm{x} = (x_1, x_2, \ldots, x_n) \in E^n$. Given a probability measure γ on E^n, define the measure $\pi := p_{2\#}\gamma$ on E^{n-1}, where p_2 is the projection $\mathrm{x} \in E \times E^{n-1} \mapsto \mathrm{x}' \in E^{n-1}$. We call $\gamma(\cdot, \cdot) : \mathcal{B}(E^n) \times E^{n-1} \to \mathbb{R}$ a conditional measure for γ, written as $\mathrm{d}\,\gamma(\mathrm{x}) = \gamma(\mathrm{d}\,\mathrm{x}, \mathrm{x}')\,\mathrm{d}\,\pi(\mathrm{x}')$, if

1. for all $\mathsf{x}' \in E^{n-1}$, $\gamma(\cdot, \mathsf{x}')$ is a probability measure on E^n,
2. for π-almost every $\mathsf{x}' \in E^{n-1}$, $\gamma(\cdot, \mathsf{x}')$ is supported in $E \times \{\mathsf{x}'\}$,
3. for any Borel set $\mathsf{R} \subset E^n$, the function $\mathsf{x}' \mapsto \gamma(\mathsf{R}, \mathsf{x}')$ is measurable, and
4. for any Borel set $\mathsf{R} \subset E^n$ and $\mathsf{S} \subset E^{n-1}$, $\gamma[\mathsf{R} \cap (E \times \mathsf{S})] = \int_{\mathsf{S}} \gamma(\mathsf{R}, \mathsf{x}') \, \mathrm{d}\, \pi(\mathsf{x}')$.

Under our assumption that E is a Polish space, conditional measures always exist [17, Corollary 10.4.10]. For π-almost every x', the measure $\gamma(\cdot, \mathsf{x}')$ is unique [17, Lemma 10.4.3] and coincides with the disintegration [39, 452E] of γ that is consistent with the projection p_2.

Souslin spaces

Souslin space theory is vital for proving one of our main results, Theorem 4.5, presented in Chapter 4. Also known as Suslin spaces or analytic sets, this theory's application in measure theory is mainly referenced in Bogachev [17, Sections 1.10, 6.6, 6.7, 7.4]. For historical context and introductory material, see also [39, Chapter 42] and [92, p.28].

Definition 1.8 (Souslin spaces)**.** A subset of a Hausdorff space is called Souslin if it is the image of a Polish space under a continuous map. The empty set is considered as Souslin as well. A Souslin space is a Hausdorff space that is a Souslin set.

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 [17, 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 [39, Proposition 423F] or [17, Theorem 6.7.3];
3. If E is a Souslin space, then every finite Borel measure μ on E is Radon [17, Theorem 7.4.3].

For a Polish space E, such as the Euclidean space $\mathbb{R}^m$, a subset A of E is a Souslin set if and only if it is the projection of a Borel subset of the product space $E \times \mathbb{R}$ [17, Theorem 6.7.2]. Nevertheless, every uncountable Polish space contains a Souslin subset that is not a Borel set [55, Theorem (14.2)]. For concrete examples of such sets, see [17, Theorem 6.7.10] and [3, Examples 12.33, 12.34].

Functional analysis

In this subsection, we recall a few results selected from functional analysis that will be used in Section 4.3, especially in the proof of Proposition 4.12.

For vector spaces, we fix the scalar field to be $\mathbb{R}$. Let $(E, \|\cdot\|)$ be a normed vector space. The dual space of E is the space E^* of all continuous linear functionals $f : E \to \mathbb{R}$, and it is equipped with the following operator norm,

$$\forall\, f \in E^*, \quad \|f\| := \sup_{\|x\| \leq 1, x \in E} |f(x)|.$$

A Banach space is a complete normed vector space. Since $\mathbb{R}$ is a Banach space, the dual space of a norm space is always a Banach space [102, Proposition 1.16, Definition 1.17]. For example, consider a measurable space $(\Omega, \mathcal{F})$ with a σ-finite measure μ on it. Then the space $L^1(\mu)$ of μ-integrable functions on Ω is a Banach space [21, Theorem 4.8], and $L^\infty(\mu)$ is the dual space of $L^1(\mu)$ [21, Theorem 4.14] [17, Theorem 4.4.1]. The weak topology of E, usually denoted by $\sigma(E, E^*)$, is the coarsest topology such that for any $f \in E^*$, the function $x \in E \mapsto f(x)$ is continuous. In other words, $\{x_n\}_{n\in\mathbb{N}^*} \subset E$ converges weakly to $x \in E$ if and only if $\lim_{n\to\infty} f(x_n) = f(x)$ for all $f \in E^*$.

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 [68, Theorem II.3].

Theorem 1.9 (Eberlein–Šmulian theorem). *A subset K of a Banach space E is pre-compact with respect to the weak topology if and only if, from each sequence of elements of K, we can extract a weakly convergent subsequence.*

The following Banach–Steinhaus theorem is also known as the uniform boundedness principle. For its proof, see [17, Theorem 4.4.3] or [21, Theorem 2.2].

Theorem 1.10 (Banach–Steinhaus theorem). *Let E be a Banach space. Let $\mathsf{F} \subset E^*$ be a set of continuous linear functional on E. If for any $x \in E$,*

$$\sup_{f\in\mathsf{F}} |f(x)| < +\infty,$$

then F is unfiormly bounded with respect to the operator norm,

$$\sup_{f\in\mathsf{F}} \|f\| < +\infty.$$

Remark 1.11. Thanks to the isometric embedding of a normed vector space E into the dual space of E^* [21, §1.3], Theorem 1.10 applied to the Banach space E^* implies that every weakly converging sequence of E is bounded in norm.

Given a measurable space $(\Omega, \mathcal{F})$, a real-valued countably additive set function $\nu : \mathcal{F} \to \mathbb{R}$ is also referred to as a finite signed measure on Ω. To deal with the set-wise convergence of countably additive set functions, we introduce the following Vitali–Hahn–Saks theorem. For its proof, see [98, §3.14], [17, Theorem 4.6.3] or [4, Theorem A8.15].

Theorem 1.12 (Vitali–Hahn–Saks theorem). *Let $(\Omega, \mathcal{F})$ be a measurable space with a probability measure μ on it. Let $\nu_n : \mathcal{F} \to \mathbb{R}$, $n \in \mathbb{N}$ 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 $A \in \mathcal{F}$;*
2. *each ν_n is absolutely continuous with respect to μ, i.e., for $A \in \mathcal{F}$, $\mu(A) = 0$ 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 1.12 with a σ-finite measure μ.

Corollary 1.13. *Let $(\Omega, \mathcal{F})$ 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*

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

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

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

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

$$\forall\, A \in \mathcal{F}, \quad \eta(A) = \int_A \sum_{j\in J} \frac{\lambda_j}{\mu(E_j)} \mathbb{1}_{E_j} \,\mathrm{d}\,\mu. \tag{1.4}$$

As $\mu(\Omega \setminus \cup_{j\in J} E_j) = 0$, (1.4) implies that $\eta(A) = 0$ if and only if $\mu(A) = 0$.

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

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

As $f \in L^1(\mu)$, the limit $\lim_{n\to\infty} \nu_n(A) = \int_A f \,\mathrm{d}\,\mu$ always exists and is finite. Since $\eta(A) = 0$ implies $\mu(A) = 0$ and thus $\nu_n(A) = 0$, Theorem 1.12 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 $\eta(A) \to 0$ [29, Lemma 4.2.1]. Hence, $\mu(A) \to 0$ implies $\sup_{n\in\mathbb{N}} |\nu_n(A)| \to 0$, which concludes the proof. □

1.3 Analysis on manifolds

In this section, we establish a rigorous framework and develop the necessary technical tools to differentiate optimal transport maps in the subsequent section. We introduce the concept of approximate Hessian for Riemannian manifolds, which we define as approximate derivative of the gradient expressed in normal coordinates. To achieve this, we first define the approximate derivative on Riemannian manifolds, using the notion of density points.

1.3.1 Approximate differentiability

We justify the definition of *density point* for Riemannian manifolds by comparing it to its usual Euclidean counterpart.

Lemma 1.14 (Density points)**.** *Let (M, d_{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*

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

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 $c_1^m \mathcal{L}^m(\varphi(N)) \leq \mathrm{Vol}(N) \leq c_2^m \mathcal{L}^m(\varphi(N))$ for any measurable subset $N \subset U$ [100, Proposition 12.6 and 12.7]. Hence, x is a density point of A if and only if

$$\lim_{r \downarrow 0} \frac{\mathcal{L}^m[\varphi(\overline{B}(x,r)) \setminus \varphi(A \cap U)]}{\mathcal{L}^m[\varphi(\overline{B}(x,r))]} = 0. \tag{1.5}$$

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, (1.5) is equivalent to that $\varphi(x)$ is a density point of $\varphi(A)$ with respect to $\mathcal{L}^m$. □

We now recall the definition of *approximate derivatives* first on Euclidean space (see [17, 5.8(v)] and [37, 3.1.2] for more detailed discussions), then on manifolds.

Definition 1.15 (Approximate derivatives on Euclidean spaces)**.** Let $m, n \geq 1$ 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 $\lim\limits_{y \in \Omega_x, y \to x} F(y) = l$. 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

$$\operatorname{ap}\lim_{y \to x} \frac{|F(y) - F(x) - L(y-x)|}{|y - x|} = 0. \tag{1.6}$$

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

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

Lemma 1.16 (Approximate derivatives on manifolds)**.** *Let (M, d_{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 $\operatorname{ap} \mathsf{D}_{\varphi(x)}[f \circ \varphi^{-1}|_{\varphi(A \cap U)}]$ exists. The approximate derivative of f at x is then defined as*

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

where $\mathsf{D}_x \varphi : T_x M \to T_{\varphi(x)} \mathbb{R}^m$ 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 [35, Theorem 6.3]. Since density points are well-defined for Riemannian manifolds by Lemma 1.14 and coordinate changes for M are smooth diffeomorphisms, it follows from (1.6) 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 (1.6) whenever $F := f \circ \varphi^{-1}$ is a constant function. □

1.3.2 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, d_{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 [106, 60], and such a ball is referred to as a geodesically convex ball.

Definition 1.17 (Semi-concavity)**.** Let (M, d_{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 [9, (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 [103, Proposition 4.3, Proposition 4.8] and [27, Theorem 5.1]. In [36, 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 1.18 (Characterization of local semi-concavity, [105, Proposition 10.12])**.** *Let (M, d_{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 $\forall\, \widetilde{x} \in \varphi(U)$, $\exists\, l_{\widetilde{x}} \in \mathbb{R}^m$, $\forall\, \widetilde{z} \in \varphi(U)$,*

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

Hence, a function is locally semi-concave if and only if it is so when expressed in local charts [36, 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 [105, Theorem 14.1] and [76, 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 1.19 (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:*

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

For $\mathcal{L}^m$-almost everywhere on A, the function $\nabla^E f$ is approximately differentiable and its approximate derivative $(\partial^2_{ij} f)_{1\leq i,j\leq m}$ 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:

$$f(z) = f(x) + \langle \nabla^E f(x), z - x\rangle + \frac{1}{2}\langle \mathrm{ap}\,\mathrm{D}_x \nabla^E f(z-x), z-x\rangle + o(\|z-x\|_2^2). \tag{1.7}$$

Remark 1.20. 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 [76, Theorem D.2.1, Theorem D.2.2]. Their equivalence to (1.7) is proven in [105, Theorem 14.25]. Compared to these equivalent formulations, our Theorem 1.19 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 $(M, \mathfrak{g})$, the Hessian at a point $x \in M$ can be interpreted either as a self-adjoint linear map from the tangent space T_xM to itself or as a symmetric bilinear form on $T_xM \times T_xM$. These two interpretations are related by duality through the Riemannian metric $\mathfrak{g}$ at x [80, 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 $\forall\, \widetilde{x} \in \varphi(U) \subset \mathbb{R}^m$, $\forall\, u, v \in \mathbb{R}^m$,

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

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

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

where ∂_i are the coordinate vectors associated with the given coordinate system [65, p.60 of Chapter 3], and Γ^k_{ij} are the Christoffel symbols of the chart, see [80, Chapter 2] for more details.

In the particular case of a chart φ inducing a normal coordinate system at $x_0 \in M$ [88, §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 $\widetilde{x}_0$ [41, 2.89 bis]. Hence, the above formula at the point $\widetilde{x}_0$ is simplified into

$$\mathrm{Hess}_{\widetilde{x}_0}(f \circ \varphi^{-1})(\partial_i, \partial_j) = \partial^2_{ij}(f \circ \varphi^{-1})(\widetilde{x}_0). \tag{1.8}$$

Since the metric matrix $(g_{ij})_{1\leq i,j\leq m}$ 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 1.21 (Hessian of semi-concave functions)**.** Let (M, 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

$$\mathrm{Hess}_x f(u) := (\mathsf{D}_x\varphi)^{-1} \circ \mathrm{ap}\, \mathsf{D}_{\varphi(x)} \nabla^E(f \circ \varphi^{-1}) \circ \mathsf{D}_x\varphi(u), \quad \forall\, u \in T_xM. \tag{1.9}$$

Remark 1.22. To justify Definition 1.21, 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

$${}^t(\mathsf{D}_{\psi(z)}T) \cdot [\nabla^E(f \circ \varphi^{-1})(\varphi(z))] = \nabla^E(f \circ \psi^{-1})(\psi(z)), \tag{1.10}$$

where z is close to x, $T := \varphi \circ \psi^{-1}$ is a $\mathcal{C}^\infty$ diffeomorphism defined around $\psi(x)$ and ${}^t(\mathsf{D}_{\psi(z)}T)$ is the transpose of T's differential at $\psi(z)$. See the proof of Lemma 1.16 for a similar argument. Moreover, in our definition (1.9) 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 (1.8) 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 (1.9) for the chart (ψ, V), the independence follows from the linearity of $\mathsf{D}_{\psi(z)}T = T$ and the equality (1.10).

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 1.23. *Let (M, 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 $\mathrm{Hess}_x f : T_xM \to T_xM$, called the Hessian of f at x, such that*

- *$\mathrm{Hess}_x f$ is a self-adjoint operator on T_xM;*
- *the function f satisfies the following second-order expansion at x,*

$$f(\exp_x u) = f(x) + \mathsf{D}_x f(u) + \frac{1}{2}\mathsf{g}_x(\mathrm{Hess}_x f(u), u) + o(\|u\|^2), \tag{1.11}$$

for $u \in T_xM$.

1.4 Optimal transport and Wasserstein spaces

Let (E, d) be a Polish metric space. We consider the (2-)Wasserstein space $(\mathcal{W}_2(E), d_W)$ of probability measures on E with

$$\begin{aligned} \mathcal{W}_2(E) :&= \left\{ \mu \text{ is a probability measure on } E \,\middle|\, \exists\, x_0 \in E, \int_E d(x_0, y)^2 \,\mathrm{d}\,\mu(y) < \infty \right\}, \\ d_W(\mu, \nu)^2 :&= \inf_{\gamma \in \Pi(\mu,\nu)} \int_{E\times E} d(x, y)^2 \,\mathrm{d}\,\gamma(x, y), \end{aligned} \tag{1.12}$$

where $\Pi(\mu,\nu)$ is the set of probability measures on $E\times E$ with marginals μ and ν respectively, i.e.,

$$\gamma\in\Pi(\mu,\nu)\Longleftrightarrow\forall\, A\in\mathcal{B}(E),\quad \gamma(A\times E)=\mu(A)\text{ and }\gamma(E\times A)=\nu(A).$$

The infimum in (1.12) is always attained by some measure $\gamma\in\Pi(\mu,\nu)$, and we call it an *optimal transport plan* between μ and ν. Wasserstein spaces enjoy the following well-known topological properties.

Theorem 1.24 (Topology of the Wasserstein spaces, [105, Theorem 6.18])**.** *Let (E,d) be a Polish metric space. Then the Wasserstein space $(\mathcal{W}_2(E),d_W)$ is a Polish metric space.*

The fact that $(\mathcal{W}_2(E),d_W)$ is a Polish space allows for an iterative construction. Thus, we can define the Wasserstein space $(\mathcal{W}_2(\mathcal{W}_2(E)),d_{\mathbb{W}})$ over the Polish space $(\mathcal{W}_2(E),d_W)$, where $d_{\mathbb{W}}$ is the 2-Wasserstein distance on $\mathcal{W}_2(\mathcal{W}_2(E))$. Convergence with respect to the Wasserstein metric is characterized as follows.

Proposition 1.25 (Convergence with respect to the Wasserstein metric, [105, Theorem 6.9])**.** *Let (E,d) be a Polish metric space. Given a sequence of probability measures $\{\mu_n\}_{n\in\mathbb{N}}$ in the Wasserstein space $(\mathcal{W}_2(E),d_W)$ and a probability measure $\mu\in\mathcal{W}_2(E)$, the limit $\lim_{n\to\infty}d_W(\mu_n,\mu)=0$ holds if and only if there exist a point $x_0\in E$ and a positive constant $C>0$ such that for all continuous functions $\phi:E\to\mathbb{R}$ with $|\phi(x)|\le C\left(1+d(x_0,x)^2\right)$, we have*

$$\lim_{n\to\infty}\int_E\phi\,\mathrm{d}\,\mu_n=\int_E\phi\,\mathrm{d}\,\mu.$$

In particular, it follows from Proposition 1.25 that convergence of probability measures with respect to the Wasserstein metric implies weak convergence.

The Wasserstein space $\mathcal{W}_2(E)$ is not proper unless the base space E is compact [8, Remark 7.19]. If (E,d) is a Polish and geodesic space, then $(\mathcal{W}_2(E),d_W)$ is geodesic as well [7, Theorem 2.10]. We refer the reader to the classic references [105, 104, 90] for a comprehensive treatment of optimal transport theory and Wasserstein spaces.

Optimal transport on Riemannian manifolds

Let us first recall the definition of c-concave functions on Riemannian manifolds.

Definition 1.26 (c-transforms and c-concave functions)**.** Let (M,d_{g}) be a Riemannian manifold. Define the function $c:M\times M\to\mathbb{R}$ as the half of the squared distance function, i.e., for $x,y\in M$,

$$c(x,y):=\frac{1}{2}d_{\mathrm{g}}(x,y)^2. \tag{1.13}$$

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

$$\phi(x)=\inf_{y\in Y}c(x,y)-\psi(y),\quad\forall x\in X. \tag{1.14}$$

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 [73], which extends Brenier's seminal theorem [104, 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 [73, Lemma 4].

Theorem 1.27 (Optimal transport on manifolds, [30, Theorem 3.2])**.** *Let (M, d_{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 (with respect to the volume measure* Vol*). 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.*

1.4.1 Optimal transport on the real line

The real line provides a notable setting where the optimal transport problem admits an explicit solution, expressed via quantile functions (Theorem 1.37). In this subsection, we shall first review interesting basic properties of quantile functions. As many of these properties will be used repeatedly in Chapter 6, we also provide detailed proofs for most of them.

Quantile functions

The definition of quantile functions involves taking the infimum of a given subset of $\mathbb{R}$. A subtlety arises when this subset is empty. To address this, we adopt the following convention for the infimum of an empty set with a specified domain $(y, z) \subset \mathbb{R}$:

$$\inf_{x \in (y,z)} \emptyset = z, \tag{1.15}$$

where y is allowed to be $-\infty$ and z is allowed to be $+\infty$. In contrast, when we are certain that we are not taking the infimum of an empty set, we shall use the notation $\inf_x$ or $\inf\limits_x$, which omits the specified domain.

Definition 1.28 (Distribution functions and quantile functions)**.** Let μ be a probability measure on $\mathbb{R}$. Its distribution function $f_\mu : \mathbb{R} \to [0, 1]$ is defined by $f_\mu(x) := \mu((-\infty, x])$, and its quantile function $f_\mu^{-1} : [0, 1] \to \overline{\mathbb{R}}$ is defined by

$$\begin{aligned} & f_\mu^{-1}(t) := \inf_x \{x \in \mathbb{R} \mid f_\mu(x) > t\} \text{ for } 0 < t < 1 \\ \text{and} \qquad & f_\mu^{-1}(0) := \lim_{t \downarrow 0} f_\mu^{-1}(t), \quad f_\mu^{-1}(1) := \lim_{t \uparrow 1} f_\mu^{-1}(t), \end{aligned}$$

where the extended real line $\overline{\mathbb{R}}$ is the set of real numbers plus two infinite values $\{-\infty, +\infty\}$.

In the literature, there exist different definitions of distribution functions and quantile functions. Our choice ensures that they share common properties such as right-continuity and monotonicity. To justify this point, it is helpful to recall the general definition of right-continuous inverse.

Lemma 1.29 (Right-continuous inverses)**.** *Let $f : (y, z) \to \mathbb{R}$ be function defined on a possibly unbounded interval $(y, z) \subset \mathbb{R}$. Define its right-continuous inverse $f^{-1} : \mathbb{R} \to \overline{\mathbb{R}}$ as follows,*

$$f^{-1}(t) := \inf_{x \in (y,z)} \{x \in (y, z) \mid f(x) > t\}, \qquad t \in \mathbb{R}.$$

If the function f^{-1} is finite on an open interval $I := (a, b)$, then its restriction $f^{-1}|_I$ is right-continuous and non-decreasing.

Proof. Assuming that $f^{-1}|_I : (a, b) \to \mathbb{R}$ is a finite function, we show that it is right-continuous. Observe that if $f(x) > t$ with $t \in (a, b)$, then there exists $\varepsilon > 0$ such that $f(x) > t + \varepsilon$. Hence, the set $\{x \mid f(x) > t\}$ is the union of $\{x \mid f(x) > t + \varepsilon\}$ for $\varepsilon > 0$, which shows that f^{-1} is right-continuous and non-decreasing on (a, b). □

According to Definition 1.28, quantile functions are completely determined by their values on the open interval $(0, 1)$. Moreover, on this interval, Lemma 1.29 guarantees their right-continuity and monotonicity, as shown in the following lemma.

Lemma 1.30. *Fix a probability measure μ on $\mathbb{R}$. Its distribution function f_μ is right-continuous and non-decreasing on $\mathbb{R}$. Its quantile function f_μ^{-1} is finite on the open interval $(0, 1)$, and the real-valued function $f_\mu^{-1}|_{(0,1)} : (0, 1) \to \mathbb{R}$ is right-continuous and non-decreasing.*

Proof. By definition, for $x \in \mathbb{R}$, $f_\mu(x) := \mu((-\infty, x])$. Thanks to the relation $(-\infty, x] = \cap_{y>x}(-\infty, y]$, this function is right-continuous [17, Proposition 1.3.3] and non-decreasing. Moreover, [17, Proposition 1.3.3] also implies the basic properties that $\lim_{x\to-\infty} f_\mu(x) = 0$ and $\lim_{x\to+\infty} f_\mu(x) = 1$.

It follows that for any $0 < t < 1$, the set $\{x \in \mathbb{R} \mid f_\mu(x) > t\}$ is non-empty with a finite infimum, and $f_\mu^{-1}(t)$ is thus finite. Hence, Lemma 1.29 applies to f_μ^{-1} with the interval $I = (0, 1)$. □

Remark 1.31. According to Definition 1.28, the quantile function f_μ^{-1} is right-continuous at 0. Furthermore, it follows from the proof of Lemma 1.30 that $f_\mu^{-1}(0)$ can be equivalently defined as $\inf_x\{x \in \mathbb{R} \mid f_\mu(x) > 0\}$. However, since the set $\{x \in \mathbb{R} \mid f_\mu(x) > 1\}$ is always empty, if we define $f_\mu^{-1}(t)$ uniformly as $\inf_{x\in\mathbb{R}}\{x \mid f_\mu(x) > t\}$ for all $t \in [0, 1]$, then $f_\mu^{-1}(1) = +\infty$ for any probability measure μ on $\mathbb{R}$. This is not convenient to express some properties of quantile functions (c.f. Lemma 1.33) compared to Definition 1.28.

For the discontinuity points of quantile functions, we characterize them as follows.

Lemma 1.32. *Fix a probability measure μ on $\mathbb{R}$. Denote by $f_\mu^{-1}(t_-) := \lim_{s\uparrow t} f_\mu^{-1}(s)$ the left limit of the quantile function f_μ^{-1} at $t \in (0, 1)$. Fix $t \in (0, 1)$. If $y = f_\mu^{-1}(t_-)$, $z = f_\mu^{-1}(t)$ with $y < z$, then*

$$f_\mu(\theta) = t \text{ for } \theta \in (y, z) \quad \text{and} \quad f_\mu(y - \epsilon) < t < f_\mu(z + \epsilon) \text{ for } \epsilon > 0. \tag{1.16}$$

Conversely, if (1.16) holds for $y < z$, then $y = f_\mu^{-1}(t_-)$, $z = f_\mu^{-1}(t)$. Therefore, $t \in (0, 1)$ is a discontinuity point of f_μ^{-1} if and only if the interval $(f_\mu^{-1}(t_-), f_\mu^{-1}(t))$ is a connected component of the complement of the support of μ.

Proof. Note that if $0 < t < 1$, then both $f_\mu^{-1}(t_-)$ and $f_\mu^{-1}(t)$ are finite according to Lemma 1.30. Moreover, since f_μ is non-decreasing, the set $\{x \in \mathbb{R} \mid f_\mu(x) > t\}$ is an interval unbounded from above. This interval could be possibly closed or open, but must have $f_\mu^{-1}(t)$ as its left endpoint by definition of $f_\mu^{-1}(t)$.

(Proof of ⇒) Assuming $y = f_\mu^{-1}(t_-)$, $z = f_\mu^{-1}(t)$ with $y < z$, we prove (1.16) by considering the value $f_\mu(w)$ for $w \in \mathbb{R}$ in different cases as follows.

1. If $w > z = f_\mu^{-1}(t)$, then by our preceding description of the set $\{x \in \mathbb{R} \mid f_\mu(x) > t\}$, it contains the point w, which implies $f_\mu(w) > t$.

2. If $w < y = f_\mu^{-1}(t_-) = \lim_{\epsilon\downarrow 0} f_\mu^{-1}(t-\epsilon)$, then there exists $\epsilon > 0$ such that $w < f_\mu^{-1}(t-\epsilon)$, which implies $f_\mu(w) \le t - \epsilon$ by definition of $f_\mu^{-1}(t-\epsilon)$ and thus $f_\mu(w) < t$.

3. If $f_\mu^{-1}(t_-) = y < w < z = f_\mu^{-1}(t)$, then $w \notin \{x \in \mathbb{R} \mid f_\mu(x) > t\}$ and thus $f_\mu(w) \le t$. Since $w > f_\mu^{-1}(t-\epsilon)$ for any $\epsilon > 0$, $f_\mu(w) > t - \epsilon$, which further implies $f_\mu(w) = t$ by the preceding inequality $f_\mu(w) \le t$.

(Proof of $\Leftarrow$) Now assume that (1.16) holds for $y < z$. Since $f_\mu(z-\delta) = t < f_\mu(z+\delta)$ holds for $\delta \in (0, z-y)$, we have $f_\mu^{-1}(t) = \inf_x\{x \in \mathbb{R} \mid f_\mu(x) > t\} = z$. By the right-continuity of f_μ, $f_\mu(y) = \lim_{\theta\downarrow y} f_\mu(\theta) = t$. If $0 < s < t = f_\mu(y)$, then $f_\mu^{-1}(s) \le y$ and thus $f_\mu^{-1}(t_-) = \lim_{s\uparrow t} f_\mu^{-1}(s) \le y$. We prove by contradiction that $f_\mu^{-1}(t_-) = y$. Indeed, if $w := f_\mu^{-1}(t_-) < y$, then for any $w' > w$ and $0 < s < t$, $f_\mu^{-1}(s) \le w < w'$ and thus $f_\mu(w') > s$ (c.f. Case 1 in the previous paragraph), which further implies $f_\mu(w) \ge t$ by the right-continuity of f_μ. However, this is a contradiction since $f_\mu(w) \le f_\mu(y-\epsilon)$ for $0 < \epsilon < y - w$ by the monotonicity of f_μ and $f_\mu(y-\epsilon) < t$ for any $\epsilon > 0$ by assumption.

For the last part, note that the open set $\mathbb{R} \setminus \mathrm{supp}(\mu)$ is a disjoint union of open intervals, with each of them being a connected component of $\mathbb{R} \setminus \mathrm{supp}(\mu)$. By definition of support, a bounded interval (y, z) $(y < z)$ is one of these connected component if and only if the distribution function f_μ is constant on the interval (y, z) but not constant on any interval $(y-\delta, z+\delta)$ for $\delta > 0$. $\square$

Moreover, for a probability measure μ on $\mathbb{R}$ with compact support, we can describe its support with the two values $f_\mu^{-1}(0)$ and $f_\mu^{-1}(1)$ as follows.

Lemma 1.33. *Let μ be a probability measure on $\mathbb{R}$. The infimum and supremum of the support of μ are related to its quantile function as follows,*

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

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

Proof. We first prove the following two inequalities,

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

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

We now prove the inequalities,

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

The case that $\inf \operatorname{supp}(\mu) = -\infty$ or $\sup \operatorname{supp}(\mu) = +\infty$ is trivial, we are left to consider the case where they are finite. If $y < \inf \operatorname{supp}(\mu)$, then $f_\mu(y) = 0$ and thus $y \le f_\mu^{-1}(t)$ for all $t \in (0, 1)$, which implies $y \le f_\mu^{-1}(0)$. As y is arbitrarily chosen, we must have $f_\mu^{-1}(0) \ge \inf \operatorname{supp}(\mu)$ since the opposite inequality $f_\mu^{-1}(0) < \inf \operatorname{supp}(\mu)$ and $y := \frac{1}{2} f_\mu^{-1}(0) + \frac{1}{2} \inf \operatorname{supp}(\mu)$ would lead to a contradiction. If $z > \sup \operatorname{supp}(\mu)$, then $f_\mu(z) = 1$ and thus $z \ge f_\mu^{-1}(t)$ for all $t \in (0, 1)$, which implies $z \ge f_\mu^{-1}(1)$. As z is arbitrarily chosen, we must have $f_\mu^{-1}(1) \le \inf \operatorname{supp}(\mu)$ since the opposite inequality $f_\mu^{-1}(1) > \inf \operatorname{supp}(\mu)$ and $z := \frac{1}{2} f_\mu^{-1}(1) + \frac{1}{2} \sup \operatorname{supp}(\mu)$ would lead to a contradiction.

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

In Definition 1.28, we define quantile functions as the right-continuous inverses (Lemma 1.29) of distribution functions. The following technical lemma [83, Lemma (4.8) of Chapter 0] holds for general right-continuous and non-decreasing functions defined on properly chosen intervals. It implies that quantile functions fully characterize probability measures, a property to be used later.

Lemma 1.34. *Let $f : \mathbb{R} \to [0, 1]$ and $g : (0, 1) \to \mathbb{R}$ be two right-continuous and non-decreasing functions. Then*

$$g(t) = \inf_{x \in \mathbb{R}} \{x \mid f(x) > t\} \text{ for } t \in (0,1) \quad \iff \quad f(x) = \inf_{t \in (0,1)} \{t \mid g(t) > x\} \text{ for } x \in \mathbb{R},$$

where we followed the convention (1.15), i.e., $\inf_{x\in\mathbb{R}} \emptyset := +\infty$ and $\inf_{t\in(0,1)} \emptyset := 1$. In particular, for a probability measure μ on $\mathbb{R}$, its distribution function f_μ is the right-continuous inverse (defined in Lemma 1.29) of $f_\mu^{-1}|_{(0,1)}$, i.e., $[f_\mu^{-1}|_{(0,1)}]^{-1} = f_\mu$.

Proof. For simplicity, we write $\{f > t\}$ and $\{g > x\}$ to denote the sets $\{x \in \mathbb{R} \mid f(x) > t\}$ and $\{t \in (0, 1) \mid g(t) > x\}$ respectively. We drop the subscripts of inf in symbols $\inf_{x\in\mathbb{R}}\{f > t\}$ and $\inf_{t\in(0,1)}\{g > x\}$ when the sets are shown to be non-empty.

Let us prove the implication from left to right. Assume that the left-hand side is true. For $x \in \mathbb{R}$, define $h(x) := \inf_{t\in(0,1)}\{g > x\}$. Fix an arbitrary real number $x \in \mathbb{R}$, we prove the equality $f(x) = h(x)$ by showing the following two inequalities.

1. We first prove the inequality $f(x) \le h(x)$. It holds trivially when $h(x) = 1$. We are left to prove the case that $h(x) < 1$, i.e., the set $\{g > x\}$ is non-empty. For any $t \in \{g > x\}$, since $g(t) = \inf\{f > t\} > x$, we have $x \notin \{f > t\}$ and thus $f(x) \le t$. It follows that $f(x) \le \inf\{g > x\} = h(x)$.

2. We then prove the inequality $f(x) \ge h(x)$. Again, this inequality is trivial when $f(x) = 1$. Hence, we proceed with case that $f(x) < 1$. As f is right-continuous at x, for $\delta > 0$ sufficiently small, we have $f(x) \le f(x + \delta) < 1$. For such a δ, since $g(f(x + \delta)) = \inf\{f > f(x + \delta)\} \ge x + \delta > x$, we have $f(x + \delta) \ge \inf\{g > x\} = h(x)$. Therefore, $f(x) = \lim_{\delta\downarrow 0} f(x + \delta) \ge h(x)$.

The implication from right to left can be proven similarly. For the last statement, it suffices to set $f := f_\mu$ and $g := f_\mu^{-1}|_{(0,1)}$. $\square$

The following lemma is analogous to the characterization of weak convergence using distribution functions. It helps to deal with the measurability issues of maps related to $\mu \mapsto f_\mu^{-1}(t)$ with t fixed,

Lemma 1.35. *Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathbb{R}$. The sequence $\{\mu_n\}_{n\in\mathbb{N}}$ converges weakly to a probability measure μ on $\mathbb{R}$ if and only if $f_{\mu_n}^{-1}(t)$ converges to $f_\mu^{-1}(t)$ for any $0<t<1$ such that f_μ^{-1} is continuous at t. Moreover, if the convergence holds, then*

$$\limsup_{n\to+\infty} f_{\mu_n}^{-1}(t) \le f_\mu^{-1}(t), \ \forall\, t\in[0,1) \quad \textit{and} \quad \liminf_{n\to+\infty} f_{\mu_n}^{-1}(1) \ge f_\mu^{-1}(1). \tag{1.17}$$

Proof. The characterization of weak convergence in terms of convergence of quantile functions at continuity points is proven in references such as [26, Proposition 5.7 of Chapter III] and [101, Lemma 21.2]. We are left to show the inequalities in (1.17). Assume that the weak convergence of $\{\mu_n\}_{n\in\mathbb{N}}$ to μ holds.

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

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

which implies $\limsup_{n\to+\infty} f_{\mu_n}^{-1}(t) \le f_\mu^{-1}(t)$ by the right-continuity of f_μ^{-1} at t.

We prove the case $t=1$ by contradiction. Assume that there exists $x\in\mathbb{R}$ such that

$$\liminf_{n\to+\infty} f_{\mu_n}^{-1}(1) = \lim_{n\to+\infty} \inf_{k>n} f_{\mu_k}^{-1}(1) < x < f_\mu^{-1}(1).$$

It follows from Definition 1.28 that $\mu((-\infty,x])<1$ and $\mu_k((-\infty,x])=1$ for infinitely many k. Since $\{\mu_n\}_{n\in\mathbb{N}}$ converges weakly to μ, the upper semi-continuity of distribution functions implies

$$1=\limsup_{n\to+\infty}\mu_n((-\infty,x]) = \limsup_{n\to+\infty} f_{\mu_n}(x) \le f_\mu(x) = \mu((-\infty,x])<1,$$

which is a contradiction. □

We provide an example showing that the inequalities in (1.17) can be strict.

Example 1.36. For $n=1,2,\ldots$, denote by $\mu_n := \mathcal{N}(0,1/n)$ the normal distribution on the real line with mean 0 and variance $1/n$. By the convergence of their quantile functions on the interval $(0,1)$, the sequence μ_n with $n\ge 1$ converges weakly to the Dirac measure $\mu:=\delta_0$ at 0. However, we have $f_{\mu_n}^{-1}(0)=-\infty$ and $f_{\mu_n}^{-1}(1)=+\infty$ for any $n\ge1$ while $f_\mu^{-1}(0)=f_\mu^{-1}(1)=0$.

The importance of quantile functions in the optimal transport theory is highlighted by the following theorem. We refer to [104, Theorem 2.18] for a proof.

Theorem 1.37. *Let μ,ν be two probability measures in the Wasserstein space $(\mathcal{W}_2(\mathbb{R}),d_W)$. Then their quantile functions $f_\mu^{-1}, f_\nu^{-1}\in L^2([0,1])$ are squared integrable and*

$$d_W(\mu,\nu)^2 = \int_0^1 [f_\mu^{-1}(t) - f_\nu^{-1}(t)]^2 \,\mathrm{d}\, t. \tag{1.18}$$

To further derive optimal transport maps between probability measures on $\mathbb{R}$, we first prove the following well-known lemma related to the uniform probability $\mathfrak{u}$ restricted on $[0,1]$.

Lemma 1.38. *Let μ be a probability measure on $\mathbb{R}$ and let $\mathfrak{u} := \mathcal{L}^1|_{[0,1]}$ be the uniform measure on $\mathbb{R}$. The quantile function f_μ^{-1} of μ pushes forward $\mathfrak{u}$ to μ, i.e., $[f_\mu^{-1}]_\#\mathfrak{u}=\mu$. If μ is atomless, then the distribution function f_μ of μ pushes forward μ to $\mathfrak{u}$, i.e., $[f_\mu]_\#\mu=\mathfrak{u}$.*

Proof. We first prove the equality $[f_\mu^{-1}]_\# \mathfrak{u} = \mu$. It suffices to show that, for any $x \in \mathbb{R}$,

$$\mu((-\infty, x]) = [f_\mu^{-1}]_\# \mathfrak{u}((-\infty, x]),$$
$$\text{or equivalently,} \quad \inf_{t\in(0,1)} \{t \mid f_\mu^{-1}(t) > x\} = \mathfrak{u}(\{t \in [0,1] \mid f_\mu^{-1}(t) \le x\}), \tag{1.19}$$

where we applied Lemma 1.34 with the definition $f_\mu(x) := \mu((-\infty, x])$ in the left-hand side. If the set $\{f_\mu^{-1} > x\}$ is empty, then both sides of (1.19) are equal to 1. If the set $\{f_\mu^{-1} > x\}$ is non-empty, then it is a sub-interval of $[0,1]$. Moreover, this sub-interval has 1 as its right endpoint, and shares a common endpoint with its complement $\{f_\mu^{-1} \le x\}$, which has 0 as its left endpoint. Hence, in this case, both sides of (1.19) are equal to the common endpoint.

Now we assume that μ is atomless and prove the equality $[f_\mu]_\# \mu = \mathfrak{u}$, which is equivalent to the following statement,

$$\forall\, t \in [0,1], \qquad \mu(\{x \in \mathbb{R} \mid f_\mu(x) \le t\}) = t. \tag{1.20}$$

For $t = 0$ or $t = 1$, the equality (1.20) holds trivially. As μ is atomless, its distribution function f_μ is continuous, which implies that the image set $f_\mu(\mathbb{R})$ is connected and thus contains the interval $(0,1)$. Hence, for any given $t \in (0,1)$, there exists $y \in \mathbb{R}$ such that $f_\mu(y) = t$. Since f_μ is non-decreasing, the set $\{f_\mu \le t\} \setminus (-\infty, y]$ is contained in the set $\{f_\mu = t\}$. As f_μ is continuous and non-decreasing, $\{f_\mu = t\}$ is either a singleton or a closed interval, and in both cases, the set is μ-negligible since μ is atomless. Therefore, $\mu(\{f_\mu \le t\}) = \mu((-\infty, y]) = f_\mu(y) = t$, which is the equality (1.20) to prove. ☐

As a corollary, we obtain the following equalities when compositing distribution functions and quantile functions.

Corollary 1.39. *Let μ be a probability measure on $\mathbb{R}$ and let $\mathfrak{u} := \mathcal{L}^1|_{[0,1]}$. If μ is atomless, then*

$$f_\mu \circ f_\mu^{-1}(t) = t, \qquad \textit{for every } t \in (0,1), \tag{1.21}$$
$$f_\mu^{-1} \circ f_\mu(x) = x, \qquad \textit{for } \mu\textit{-almost every } x \in \mathbb{R}. \tag{1.22}$$

Proof. Let us first prove (1.21) for $\mathfrak{u}$-almost everywhere:

$$f_\mu \circ f_\mu^{-1}(t) = t, \qquad \text{for } \mathfrak{u}\text{-almost every } t \in (0,1), \tag{1.23}$$

Since $f_\mu^{-1}(t)$ is finite for $t \in (0,1)$ by Lemma 1.30, Lemma 1.34 implies that

$$f_\mu \circ f_\mu^{-1}(t) = \inf_{s\in(0,1)} \{s \mid f_\mu^{-1}(s) > f_\mu^{-1}(t)\}.$$

As t is smaller than any possible element in $\{f_\mu^{-1} > f_\mu^{-1}(t)\}$, we have $f_\mu \circ f_\mu^{-1}(t) \ge t$. Hence,

$$\begin{aligned}\int_0^1 |f_\mu \circ f_\mu^{-1}(t) - t| \,\mathrm{d}\, t &= \int_{\mathbb{R}} f_\mu \circ f_\mu^{-1} \,\mathrm{d}\, \mathfrak{u} - \int_{\mathbb{R}} \mathrm{Id} \,\mathrm{d}\, \mathfrak{u} \\ &= \int_{\mathbb{R}} \mathrm{Id} \,\mathrm{d}[f_\mu \circ f_\mu^{-1}]_\# \mathfrak{u} - \int_{\mathbb{R}} \mathrm{Id} \,\mathrm{d}\, \mathfrak{u} = 0,\end{aligned}$$

where we applied the equality $[f_\mu \circ f_\mu^{-1}]_\# \mathfrak{u} = [f_\mu]_\# \mu = \mathfrak{u}$ implied by Lemma 1.38. It follows that $|f_\mu \circ f_\mu^{-1}(t) - t| = 0$ for $\mathfrak{u}$-almost every $t \in (0,1)$, which implies (1.23).

By considering the integral $\int_{\mathbb{R}} |f_\mu^{-1} \circ f_\mu(x) - x| \,\mathrm{d}\,\mu(x)$, we can prove the μ-almost everywhere equality (1.22) similarly thanks to the inequality $f_\mu^{-1} \circ f_\mu(x) \geq x$ and the equality $[f_\mu^{-1} \circ f_\mu]_{\#}\mu = \mu$.

Finally, let us deduce (1.21) from (1.23). Define $A := \{t \in (0,1) \mid f_\mu \circ f_\mu^{-1}(t) = t\}$. Since $\mathfrak{u}(A) = \mathfrak{u}([0,1]) = 1$ and any open subset of $[0,1]$ has strictly positive $\mathfrak{u}$-measure, for any $t \in (0,1)$, there exists a sequence $\{t_n\}_{n\geq 1} \subset A$ such that $t_n \geq t_{n+1}$ and $\lim_{n\to\infty} t_n = t$. Hence, by the right-continuity of f_μ and f_μ^{-1},

$$t = \lim_{n\to\infty} t_n = \lim f_\mu \circ f_\mu^{-1}(t_n) = f_\mu \circ \left(\lim_{n\to\infty} f_\mu^{-1}(t_n)\right) = f_\mu \circ f_\mu^{-1}(t),$$

which implies $t \in A$ and thus the statement (1.21). $\square$

We are thus able to deduce the following corollary of Theorem 1.37.

Corollary 1.40. *Let μ, ν be two probability measures in the Wasserstein space $(\mathcal{W}_2(\mathbb{R}), d_W)$. If μ is atomless, then $f_\nu^{-1} \circ f_\mu$ is an optimal transport map pushing forward μ to ν.*

Proof. Thanks to the equality (1.22) and $[f_\mu]_{\#}\mu = \mathfrak{u}$, we have

$$\int_{\mathbb{R}} [f_\nu^{-1} \circ f_\mu(x) - x]^2 \,\mathrm{d}\,\mu = \int_{\mathbb{R}} [f_\nu^{-1} \circ f_\mu(x) - f_\mu^{-1} \circ f_\mu(x)]^2 \,\mathrm{d}\,\mu = \int_0^1 [f_\mu^{-1}(t) - f_\nu^{-1}(t)]^2 \,\mathrm{d}\,t.$$

Hence, it follows from Theorem 1.37 that

$$d_W(\mu,\nu)^2 = \int_{\mathbb{R}} [f_\nu^{-1} \circ f_\mu(x) - x]^2 \,\mathrm{d}\,\mu,$$

which implies that $f_\nu^{-1} \circ f_\mu$ is an optimal transport map. $\square$

1.4.2 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 [30, Sections 4 & 5]. These properties will be used in Chapter 4. To justify them, we remark that our definition of Hessian enjoys the second-order expansion (1.11), which allows us to apply properties proven for the Hessian defined in [30, Definition 3.9]. See Remark 1.20 and [30, Discussion after Definition 3.9] for more details.

To motivate the definition of differentiating optimal transport maps, we first illustrate how to differentiate the maps $\exp(-\nabla\phi)$ with ϕ being $\mathcal{C}^2$ smooth. Let us first recall the definition of parallel transport. We denote by ∇ the Levi-Civita connection on a Riemannian manifold.

Definition 1.41 (Parallel transport)**.** Let $(M, \mathfrak{g})$ be a Riemannian manifold. Given a smooth curve $\gamma : I \to M$ on an open interval I, $t_0 \in I$ and $v \in T_{\gamma(t_0)}M$, a vector field X along γ is called the parallel transport of v along γ if

$$X_{\gamma(t_0)} = v, \quad \text{and} \quad \nabla_{\gamma'(t)}X = 0 \text{ for } t \in I.$$

When a particular point $t \in I$ is selected in the context, for example, by explicitly considering the tangent space $T_{\gamma(t)}M$, we also call the tangent vector $X_{\gamma(t)} \in T_{\gamma(t)}M$ the parallel transport of v.

For the existence and uniqueness of parallel transport, see [66, Theorem 4.32]. Parallel transport is deeply connected with the Riemannian metric. For example, it is a linear isometry along the smooth curve [66, Proposition 5.5]. Moreover, it determines the Levi-Civita connection.

Proposition 1.42 (Parallel transport determines the connection, [66, Corollary 4.35])**.** *Let (M, g) be a Riemannian manifold. Suppose X and Y are smooth vector fields on M. Fix a point $p \in M$ and a smooth curve $\gamma : I \to M$ with $t_0 \in I$ such that $\gamma(t_0) = p$ and $\gamma'(t_0) = X_p$. For $t \in I$, denote by $\Pi^{\gamma}_{t\to t_0} : T_{\gamma(t)}M \to T_{\gamma(t_0)}M$ the parallel transport map, sending tangent vectors in $T_{\gamma(t)}M$ to their parallel transports in $T_{\gamma(t_0)}M$ along the curve γ. Then*

$$\nabla_X Y|_p = \lim_{t\to t_0} \frac{\Pi^{\gamma}_{t\to t_0} Y_{\gamma(t)} - Y_p}{t - t_0}. \tag{1.24}$$

Note that in the equality (1.24), the right-hand is independent of the smooth curve γ, provided $\gamma'(t_0) = X_p$. Also, we remind that the vector $\nabla_X Y|_p$ can be also written as $\nabla_{X_p} Y$ as it only depends on X_p and the value of Y in a neighborhood of p [66, Proposition 4.5]. Since parallel transports along different curves are considered in the following proof, we introduce the symbol $\Pi_{z\to w} : T_zM \to T_wM$ without indicating the curve explicitly, to represent the map sending a tangent vector $v \in T_zM$ to its parallel transport in T_wM along the minimal geodesic γ from z to w.

Lemma 1.43. *Let (M, 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 $\mathrm{d}_y^2/2$, have the same gradient at x, then*

$$\mathsf{D}_x F = [\mathsf{D}_{-\nabla\phi(x)} \exp_x] \circ (\mathrm{Hess}_x\, d_y^2/2 - \mathrm{Hess}_x\, \phi). \tag{1.25}$$

In the above formula,

1. *$\mathsf{D}_{-\nabla\phi(x)} \exp_x : T_{-\nabla\phi(x)}T_xM \to T_yM$ denotes the differential of the exponential map $\exp_x : T_xM \to M$ at $-\nabla\phi(x)$;*
2. *the composition is defined via the canonical identification of $T_{-\nabla\phi(x)}T_xM$ with T_xM.*

Proof. The formula (1.25) is already proven in [30, 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 [88, (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 $\Pi_{z\to w} : T_zM \to T_wM$ 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_zM, $g(z, z) = F(z)$. 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. Let us verify that the differential at x of the map $z \in M \mapsto \Pi_{z\to x}\nabla\phi(z) \in T_xM$ is $\mathrm{Hess}_x\, \phi$, i.e.,

$$\mathsf{D}_x G = \mathrm{Hess}_x\, \phi \quad \text{with} \quad G(z) := \Pi_{z\to x}\nabla\phi(z), \tag{1.26}$$

where the tangent space $T_{G(x)}T_xM$ is identified with T_xM so that $\mathsf{D}_xG : T_xM \to T_{G(x)}T_xM$ is regarded as map from the space T_xM to itself. Indeed, given a vector $v \in T_xM$, by introducing the minimal geodesic $\gamma : t \in (-\delta, \delta) \mapsto \exp_x tv$ for small $\delta > 0$, we have $\Pi^{\gamma}_{t\to 0} = \Pi_{\gamma(t)\to x}$ according to Definition 1.41. Hence, Proposition 1.42 implies

$$\mathrm{Hess}_x\, \phi(v) = \nabla_v \nabla\phi = \lim_{t\to 0} \frac{\Pi^{\gamma}_{t\to 0}\nabla\phi(\gamma(t)) - \nabla\phi(x)}{t} = \lim_{t\to 0} \frac{G(\gamma(t)) - G(\gamma(0))}{t} = \mathsf{D}_x G(v),$$

where we used the definition of Hessian as the covariant derivative of gradients [80, Proposition 2.2.6]. Since $v \in T_x M$ is arbitrarily chosen, (1.26) is thus proven. Therefore,

$$\begin{aligned} \mathsf{D}_x F &= \partial_w g(x,x) + \partial_z g(x,x) = \partial_z g(x,x) && (1.27)\\ &= [\mathsf{D}_{-\nabla\phi(x)} \circ \exp_x] \circ (\mathrm{Hess}_x\, d_y^2/2 - \mathrm{Hess}_x\, \phi), && (1.28) \end{aligned}$$

where we applied $F(z) = g(z,z)$, the chain rule and $g(w,x) \equiv y$ for the line (1.27), and applied the relation between Hessians and differential of parallel transports, as illustrated by (1.26), for the line (1.28). $\square$

We are now ready to import the definition of the (weak) differential of optimal transport maps from [30], with which we can then state the change of variables formula.

Proposition 1.44 (Differentiating optimal transport maps, [30, Proposition 4.1])**.** *Let (M, d_{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* $\mathrm{Hess}_x\, \phi$ *exists (1.9). Then the point $y := F(x)$ is not in the cut locus of x, $\nabla\phi(x) = \nabla d_y^2/2(x)$, and* $\mathrm{Hess}_x\, d_y^2/2 - \mathrm{Hess}_x\, \phi$ *is positive semi-definite. Define the differential $\mathsf{D}_x F : T_x M \to T_y M$ of F at x as*

$$\mathsf{D}_x F := [\mathsf{D}_{-\nabla\phi(x)} \exp_x] \circ (\mathrm{Hess}_x\, d_y^2/2 - \mathrm{Hess}_x\, \phi), \tag{1.29}$$

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

The Jacobian determinant of the differential $\mathsf{D}_x F$, as defined in Proposition 1.44, is calculated with respect to normal coordinate systems of the tangent spaces $T_x M$ and $T_y M$ [30, Lemma 2.1]. By [30, 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 [105, p.364 of Chapter 14].

Proposition 1.45 (Interpolation and change of variables formula)**.** *Let (M, d_{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 1.27.*

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

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

satisfies the following properties:

1. *$\mu(\Omega) = 1$;*
2. *defining $F^t := \exp(-t\nabla\phi)$ for $0 \le t \le 1$, we have* $\mathrm{Jac}\, F^t > 0$ *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 N as in Property 3,*

$$\int_M A(g) \,\mathrm{d}\operatorname{Vol} = \int_N A\left(\frac{f}{\operatorname{Jac} F}\right) \operatorname{Jac} F \,\mathrm{d}\operatorname{Vol}. \tag{1.30}$$

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

Proof. All the statements follow from [30, Claim 4.4, Theorem 4.2, Corollary 4.7] except Property 2 for $t \in (0, 1)$. To justify this proposition, we fix $t \in (0, 1)$ and deduce Property 2 from the following known results:

(a) $\det[\mathsf{D}_{-t\nabla\phi(x)} \exp_x] > 0$ since $\exp_x(-t\nabla\phi(x))$ is not in the cut locus of x [66, (c) of Theorem 10.34].

(b) $t\phi$ is c-concave [30, Lemma 5.1].

(c) $\operatorname{Hess}_x d^2_{F^t(x)}/2 - t \operatorname{Hess}_x d^2_{F(x)}/2$ is positive semi-definite [30, Lemma 2.3].

(d) $\operatorname{Hess}_x d^2_{F(x)}/2 - \operatorname{Hess}_x \phi$ is positive definite since it is positive semi-definite [30, Proposition 4.1] and $\det \mathsf{D}_x F = \det[\mathsf{D}_{-\nabla\phi(x)} \exp_x] \cdot \det[\operatorname{Hess}_x d^2_{F(x)}/2 - \operatorname{Hess}_x \phi] > 0$ [30, Claim 3.4].

Since $\det \mathsf{D}_x F^t = \det[\mathsf{D}_{-t\nabla\phi(x)} \exp_x] \cdot \det[\operatorname{Hess}_x d^2_{F^t(x)}/2 - t \operatorname{Hess}_x \phi]$ according to Proposition 1.44 and Result (b), it suffices to show $\det[\operatorname{Hess}_x d^2_{F^t(x)}/2 - t \operatorname{Hess}_x \phi] > 0$ by Result (a). Denote by m the dimension of M. Recall that the Minkowski's determinant inequality [104, (5.23)] states, if A, B are two symmetric $m \times m$ matrices such that A is positive semi-definite and B is positive definite, then

$$\det[A + B]^{\frac{1}{m}} \geq \det A^{\frac{1}{m}} + \det B^{\frac{1}{m}}.$$

Considering the equality

$$\operatorname{Hess}_x d^2_{F^t(x)}/2 - t \operatorname{Hess}_x \phi = \left[\operatorname{Hess}_x d^2_{F^t(x)}/2 - t \operatorname{Hess}_x d^2_{F(x)}/2\right] + \left[t \operatorname{Hess}_x d^2_{F(x)}/2 - t \operatorname{Hess}_x \phi\right],$$

it follows from Result (c) and Result (d) that

$$\begin{aligned} \det\left[\operatorname{Hess}_x d^2_{F^t(x)}/2 - t \operatorname{Hess}_x \phi\right]^{\frac{1}{m}} &\geq \det\left[\operatorname{Hess}_x d^2_{F^t(x)}/2 - t \operatorname{Hess}_x d^2_{F(x)}/2\right]^{\frac{1}{m}} \\ &\quad + t \det\left[\operatorname{Hess}_x d^2_{F(x)}/2 - \operatorname{Hess}_x \phi\right]^{\frac{1}{m}} > 0, \end{aligned}$$

which concludes the proof. ∎

Chapter 2

General framework for barycenters

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 second moment is finite, then its mean $\int_{\mathbb{R}^m} x \, \mathrm{d}\, \mu(x)$ can be equivalently defined as the unique point where the infimum

$$\inf_{y \in \mathbb{R}^m} \int_{\mathbb{R}^m} \|y - x\|_2^2 \, \mathrm{d}\, \mu(x)$$

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).

For barycenters in proper metric spaces, their existence is a consequence of the compactness property. Furthermore, in Section 2.1, we also demonstrate the existence of measurable barycenter selection maps, a crucial element for the construction of Wasserstein barycenters (Proposition 2.12). Consequently, the framework for barycenters in this chapter involves considering a proper metric space (E, d) and studying barycenters within E or Wasserstein barycenters in $\mathcal{W}_2(E)$. To provide partial justification for this framework, we also include Section 2.2, which presents counter-examples illustrating the failure of barycenter's existence in metric spaces that are not proper. Finally, Section 2.3 reviews established results concerning the existence and uniqueness of Wasserstein barycenters.

2.1 Barycenters on proper metric spaces

Given that Wasserstein spaces are composed of probability measures with finite second moments, the definition of barycenters for these measures is sufficient for our development. For a slightly more general definition, we refer to [94, Proposition 4.3]. For clarity, we shall use the symbol z_μ to represent a chosen barycenter of the measure μ, but it is important to note that this does not imply the uniqueness of z_μ.

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 $z_\mu \in E$ a barycenter of μ if

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

Recall that 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 [77] for more details and some other properties of barycenters in a proper space.

We can readily construct counter-examples demonstrating the lack of uniqueness for barycenters.

Example 2.2. Let $\mathbb{S}^2$ be the two-dimensional sphere and fix two antipodal points x, y of $\mathbb{S}^2$. Consider the measure $\mu = \frac{1}{2}\delta_x + \frac{1}{2}\delta_y$ on $\mathbb{S}^2$. Then all points in the equator, i.e., the set of all points with equal distances to x and y, are barycenters of μ.

Given the prevalence of the phenomenon illustrated by Example 2.2, we prioritize investigating measurable selections of barycenters, which are necessary for our subsequent development, rather than identifying conditions ensuring uniqueness.

2.1.1 Measurable selection of barycenters

Let us recall the following topological property of projection maps. 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 (i.e., mapping open sets to open sets). The map p_1 (respectively p_2) is closed if E_2 (respectively E_1) is compact [20, Proposition 8.2].

Proposition 2.3 (Measurable selections of barycenters)**.** *Let (E, d) be a proper metric space. The function $f : \mathcal{W}_2(E) \to \mathbb{R}$ defined by*

$$f(\mu) := \min_{x\in E} d_W(\mu, \delta_x)$$

is continuous. There exists a measurable map $Z : \mathcal{W}_2(E) \to E$ such that for $\mu \in \mathcal{W}_2(E)$, $Z(\mu)$ is a barycenter of μ. Moreover, if $A \subset \mathcal{W}_2(E)$ is a compact set, then the set of all barycenters of μ for μ running through A is compact.

Proof. Observe that $d_W(\mu, \delta_x)^2 = \int_{x\in E} d(x, y)^2 \,\mathrm{d}\,\mu(y)$ is exactly the term to be minimized when we define the barycenters of μ. As (E, d) is a proper metric space, the minimum in the definition of $f(\mu) = \min_{x\in E} d_W(\mu, \delta_x)$ is reached by the barycenters of μ, which shows that f is well-defined. We now prove the continuity of f. For $\mu, \nu \in \mathcal{W}_2(E)$ and $y \in E$, thanks to the triangle inequality of the Wasserstein metric d_W, we have

$$f(\mu) = \min_{x\in E} d_W(\mu, \delta_x) \le d_W(\mu, \delta_y) \le d_W(\mu, \nu) + d_W(\nu, \delta_y). \tag{2.1}$$

By taking the infimum of the right-hand side of (2.1) over all $y \in E$, we obtain $f(\mu) \le d_W(\mu, \nu) + f(\nu)$. After exchanging the roles of μ and ν, it follows that $|f(\mu) - f(\nu)| \le d_W(\mu, \nu)$, which implies the continuity of f. Hence, the following set

$$\Gamma := \{(\mu, z) \in \mathcal{W}_2(E) \times E \mid d_W(\mu, \delta_z) = f(\mu)\},$$

is closed. Furthermore, $(\mu, z) \in \Gamma$ if and only if z is a barycenter of μ.

We then prove the existence of measurable selection of barycenters. Fix a compact subset K of E. Consider the following set

$$\Gamma_K := p_1[\Gamma \cap (\mathcal{W}_2(E) \times K)] \subset \mathcal{W}_2(E),$$

where $p_1 : \mathcal{W}_2(E) \times K \to \mathcal{W}_2(E)$ is the first projection map. Note that Γ_K is set of all measures in $\mathcal{W}_2(E)$ with one barycenter located at K. Since K is a compact set, p_1 is a closed map [20, Proposition 8.2], which implies that Γ_K is a closed set and thus measurable. Consider the map $\Psi : \mathcal{W}_2(E) \to 2^E$ sending $\mu \in \mathcal{W}_2(E)$ to the set of barycenters of μ. By definition of Γ and Γ_K, we have

$$\Psi(\mu) = p_2[\Gamma \cap (\{\mu\} \times E)] \quad \text{and} \quad \{\mu \in \mathcal{W}_2(E) \mid \Psi(\mu) \cap K \neq \emptyset\} = \Gamma_K.$$

Therefore, according to Theorem 1.5 and Lemma 1.6, to obtain a measurable selection map $Z : \mathcal{W}_2(E) \to E$ of Ψ, we are left to show that $\Psi(\mu)$ is closed for any $\mu \in \mathcal{W}_2(E)$. However, this property follows from the fact that $\Gamma \cap (\{\mu\} \times E)$ is a closed set as an intersection of two closed sets. It remains to prove the "moreover" part of the proposition, which is a generalization of the previous property.

For a compact set $A \subset \mathcal{W}_2(E)$, the set of all barycenters of μ for μ running through A can be equivalently expressed as

$$\mathrm{bary}(A) := p_2[\Gamma \cap (A \times E)] \subset E,$$

where $p_2 : A \times E \to E$ is the second projection map. Since A is compact, p_2 is then a closed map [20, Proposition 8.2], which further implies that $\mathrm{bary}(A)$ is a closed set. We claim that the set $\mathrm{bary}(A)$ is bounded. Fix two arbitrarily chosen points $x, y \in \mathrm{bary}(A)$, and suppose that they are respectively barycenters of $\mu_x, \mu_y \in A$. By the triangle inequality of d_W, we have

$$\begin{aligned} d(x,y) = d_W(\delta_x, \delta_y) &\leq d_W(\mu_x, \delta_x) + d_W(\mu_y, \delta_y) + d_W(\mu_x, \mu_y) \\ &= f(\mu_x) + f(\mu_y) + d_W(\mu_x, \mu_y) \\ &\leq 2 \sup_{\mu \in A} f(\mu) + \sup_{\mu,\nu \in A} d_W(\mu, \nu). \end{aligned}$$

Thanks to the continuity of f and the compactness of A, the term $2\sup_{\mu \in A} f(\mu) < +\infty$ is bounded. By the continuity of the distance function $d_W : \mathcal{W}_2(E) \times \mathcal{W}_2(E) \to \mathbb{R}$ and compactness of the set $A \times A$, the term $\sup_{\mu,\nu \in A} d_W(\mu,\nu) < +\infty$ is also bounded. Hence, our preceding claim is proven as points $x, y \in \mathrm{bary}(A)$ are arbitrarily chosen. It follows that $\mathrm{bary}(A) \subset E$ is compact since E is a proper metric space. ☐

We shall apply Proposition 2.3 mainly with the following type of measures that are supported in finitely many points.

Corollary 2.4. *Let (E, d) be a proper metric space. Fix a positive integer $n \geq 1$ and n positive real numbers $\lambda_i > 0$ for $i = 1, \dots, n$ such that $\sum_{i=1}^n \lambda_i = 1$. The function $f : E^n \to \mathbb{R}^n$ defined by*

$$f(x_1, \dots, x_n) := \min_{y \in E} \sum_{i=1}^n \lambda_i \, d(y, x_i)^2 \tag{2.2}$$

is continuous. There exists a measurable map $B : E^n \to E$ such that for $\mathsf{x} = (x_1, \dots, x_n) \in E^n$, $B(\mathsf{x})$ is a barycenter of the probability $\sum_{i=1}^n \lambda_i \, \delta_{x_i}$. Moreover, if $\mathsf{A} \subset E^n$ is a compact set, then the set of all barycenters of $\sum_{i=1}^n \lambda_i \, \delta_{x_i}$ for $\mathsf{x} = (x_1, \dots, x_n)$ running through A is compact.

Proof. Consider the following map from E^n to $\mathcal{W}_2(E)$:

$$\theta : (x_1, \dots, x_n) \in E^n \mapsto \sum_{i=1}^n \lambda_i \, \delta_{x_i} \in \mathcal{W}_2(E).$$

We first prove that θ is a continuous map. For two points $\mathtt{x} = (x_1, \ldots, x_n) \in E^n$ and $\mathtt{y} = (y_1, \ldots, y_n) \in E^n$, by considering the transport plan sending x_i to y_i for each $1 \leq i \leq n$, we have

$$d_W(\theta(\mathtt{x}), \theta(\mathtt{y}))^2 = d_W(\sum_{i=1}^n \lambda_i\, \delta_{x_i}, \sum_{i=1}^n \lambda_i\, \delta_{y_i})^2 \leq \sum_{i=i}^n \lambda_i\, d(x_i, y_i)^2.$$

It follows that if $\mathtt{y}$ converges to $\mathtt{x}$ in E^n, then $\theta(\mathtt{y})$ converges to $\theta(\mathtt{x})$ in $\mathcal{W}_2(E)$.

By definition of f in (2.2), for $\mathtt{x} = (x_1, \ldots, x_n) \in E^n$, we have

$$f(\mathtt{x}) = \min_{y \in E} \sum_{i=1}^n \lambda_i\, d(x_i, y)^2 = \min_{y \in E} d_W(\sum_{i=1}^n \lambda_i\, \delta_{x_i}, \delta_y)^2 = \min_{y \in E} d_W(\theta(\mathtt{x}), \delta_y)^2.$$

It follows from Proposition 2.3 and the continuity of θ that f is also continuous. Moreover, according to Proposition 2.3, there exists a measurable barycenter selection map $Z : \mathcal{W}_2(E) \to E$. Hence, the map $B := Z \circ \theta$ is measurable and sends $(x_1, \ldots, x_n)$ to a barycenter of $\sum_{i=1}^n \lambda_i\, \delta_{x_i}$. Moreover, since θ is continuous, the set $A := \theta(\mathsf{A}) \subset \mathcal{W}_2(E)$ is compact if $\mathsf{A} \subset E^n$ is compact, which implies the last part of the corollary by Proposition 2.3. □

2.1.2 Barycenters and cut-loci

In the context of Riemannian manifolds, the problem of finding barycenters for a finite set of points exhibits a close relationship with optimal transport problems through the presence of c-concave functions. This connection will be leveraged in Chapter 3.

Lemma 2.5. *Let (M, d_{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 $\sum_{i=1}^n \lambda_i = 1$. With the function c given in (1.13), we define*

$$f : (x_1, x_2, \ldots, x_n) \in M^n \mapsto \min_{w \in M} \sum_{i=1}^n \lambda_i\, c(w, x_i) = \frac{1}{2} \min_{w \in M} \sum_{i=1}^n \lambda_i\, d_{\mathrm{g}}(w, x_i)^2. \tag{2.3}$$

Fix a non-empty compact subset $X \subset M$ and $n-1$ points $x_i \in M$ for $2 \leq i \leq n$. 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 Corollary 2.4. Using the given definition of Y, we can replace the minimum over M in (2.3) by the minimum over X, 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

$$g_1(y) \leq f_1^c(y) := \inf_{x \in X} c(x, y) - f_1(x). \tag{2.4}$$

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 two inequalities, $f_1(x_1) + f_1^c(y) \leq c(x_1, y) = f_1(x_1) + g_1(y)$ and (2.4), that $g_1(y) = f_1^c(y)$. Since y is arbitrarily chosen, we conclude that $g_1 = f_1^c \in \mathcal{I}^c(Y, X)$. □

The c-concave function $g_1 \in \mathcal{I}^c(Y, X)$ defined in Lemma 2.5 has simple expression unlike its c-transform f_1. Furthermore, thanks to the following lemma by Kim and Pass [57, Lemma 3.1], we conclude that g_1 is $\mathcal{C}^2$ smooth since squared distance functions are $\mathcal{C}^2$ smooth out of cut-locus. This differential property of g_1 (to be used in Lemma 3.2) is crucial to prove the absolute continuity of Wasserstein barycenters.

Lemma 2.6 (Barycenters and cut loci, [57, Lemma 3.1 and proof of Theorem 6.1])**.** *Let (M, d_{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 $\sum_{i=1}^n \lambda_i \, \delta_{x_i}$.*

2.2 Counter-examples of barycenter's existence

Recall that the Wasserstein space $\mathcal{W}_2(E)$ over a metric space E is not proper unless the space E is compact [8, Remark 7.19]. Consequently, the existence of barycenter in Wasserstein spaces is not guaranteed *a priori*. To better illustrate the obstacles toward barycenter's existence, we dedicate this section to examining the existence of barycenters in general metric spaces that are not proper.

As recalled in Definition 1.2, a length space [23, Chapter 2] is a metric space where the distance between two points is the infimum of the lengths of all rectifiable curves joining them. Here, curves are continuous maps from compact intervals $[a, b] \subset \mathbb{R}$ to the metric space. For example, Riemannian manifolds and Wasserstein spaces over them are length spaces. We shall provide some counter-examples of barycenters' existence in length spaces. The following lemma facilitates determining whether a point is a barycenter.

Lemma 2.7. *Given two points x, y in a length space (E, d), z_μ is a barycenter of $\mu := \frac{1}{2}\delta_x + \frac{1}{2}\delta_y$ if and only if it is a midpoint between x and y, i.e., $d(x, z_\mu) = d(z_\mu, y) = \frac{1}{2}d(x, y)$.*

Proof. A midpoint z between x and y reaches the two equalities in the following long inequality,

$$d(x, y)^2 \leq (d(x, z) + d(z, y))^2 \leq 2\left(d(x, z)^2 + d(z, y)^2\right) = 4\int_E d(z, w)^2 \, \mathrm{d}\, \mu(w),$$

which implies that z is a barycenter of μ if z is a midpoint.

Assume that z_μ is a barycenter of $\mu := \frac{1}{2}\delta_x + \frac{1}{2}\delta_y$. For a rectifiable curve $\gamma : [0, 1] \to E$, denote by $\gamma|_{[s,t]}$ its restriction on $[s, t] \subset [0, 1]$ and by $\tau_\gamma \in [0, 1]$ a "midway position" such that $L_d(\gamma_{[0,\tau_\gamma]}) = L_d(\gamma_{[\tau_\gamma,1]})$ (see (1.1) for the definition of L_d), whose existence follows from the continuity of length structure with respect to concatenation [23, §2.2.1]. For a rectifiable curve γ from x to y, since z_μ is a barycenter of μ,

$$\begin{aligned} d(x, z_\mu)^2 + d(z_\mu, y)^2 &= 2\int_E d(z_\mu, w)^2 \, \mathrm{d}\, \mu(w) \leq 2\int_E d(\gamma(\tau_\gamma), w)^2 \, \mathrm{d}\, \mu(w) \\ &\leq L_d(\gamma_{[0,\tau_\gamma]})^2 + L_d(\gamma_{[\tau_\gamma,1]})^2 = \frac{1}{2}L_d(\gamma)^2. \end{aligned}$$

Taking the infimum over all possible γ on the right-hand side, we obtain $d(x, z_\mu)^2 + d(z_\mu, y)^2 \leq \frac{1}{2}d(x, y)^2$, which shows that z_μ is a midpoint between x and y. ☐

With Lemma 2.7, we can construct counter-examples in length spaces as follows.

Example 2.8 (No existence of barycenters in some length spaces)**.** Recall that a locally compact complete length space is proper [23, Theorem 2.5.28] and thus guarantees the existence of barycenters. Here are two counter-examples of barycenter's existence when the space is not proper.

1. For a locally compact but not complete length space, consider the unit disk without origin. From physical intuition there is no barycenter for its uniform measure. Alternatively, we can pick two center-symmetric points $x = -y$ and consider the measure $\frac{1}{2}\delta_x + \frac{1}{2}\delta_y$ as an example.

2. For a complete but not locally compact length space, we shall prove Lemma 2.9 as an example.

To justify the second example above, it suffices to show that the given space is not geodesic, since in a complete length space, shortest paths always exist if midpoints always exist [23, Theorem 2.4.16]. Recall that a length space is called geodesic (Definition 1.3) if the distance between two points is equal to the length of some rectifiable curve connecting them.

In the following example inspired by [45, Example 5.1] (see also [91, Example 4.43]), we express the induced lengths of Lipschitz curves using integrals of their derivatives, similar to the case of Riemannian manifolds. Since energy variation shares the same solutions as arc-length variation, we can disprove the existence of shortest paths between two selected points by showing that the corresponding energy variation has no solution. Note that starting from an arbitrary metric space, we can always define an induced length structure (Definition 1.1) on it, and further turn the space into a length metric space by equipping it with the metric induced by the previous length structure. See [23, §2.3.3] for more details of this construction.

Lemma 2.9 (Infinite dimensional ellipsoids in the Hilbert space $\mathbb{R}^\infty$)**.** *Let $(c_n)_{n\in\mathbb{N}^*}$ be a strictly decreasing sequence with a positive lower bound. We define*

$$E := \left\{ (x_1, x_2, \ldots) \in \mathbb{R}^\infty \mid \sum_{n=1}^{\infty} \frac{x_n^2}{c_n^2} = 1 \right\}.$$

Let d be the metric on E inherited from the Hilbert space $\mathbb{R}^\infty$ and let L_d be the length structure on E induced by d. Then there is no curve γ connecting two poles $e := (c_1, 0, \ldots)$ and $-e := (-c_1, 0, \ldots)$ that reaches the infimum length $L_d(\gamma)$ between them.

Moreover, there exists a length space $(E, \hat{d})$ defined via L_d that is complete but not geodesic.

Proof. Denote by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ respectively the inner product and norm of the Hilbert space $\mathbb{R}^\infty$. Let $\mathfrak{u} := \mathcal{L}^1|_{[0,1]}$ the uniform measure on $[0,1]$. A rectifiable curve γ always admits an arc-length proportional parametrization on $[0,1]$ and its length does not depend on its parametrization [23, Proposition 2.5.9]. For a Lipschitz curve γ, one can define its derivative $\gamma' : [0,1] \to \mathbb{R}^\infty$ almost everywhere since it has *countably many* components and each component of γ is a Lipschitz function from $[0,1]$ to $\mathbb{R}$. It follows from the Newton–Leibniz formula [17, Theorem 5.4.2] that for $0 < s \le t < 1$, $\gamma(t) - \gamma(s) = \int_s^t \gamma' \,\mathrm{d}\,\mathfrak{u}$. Hence, $\|\gamma'\|$ is equal to the metric derivative (speed) of γ for $\mathfrak{u}$-almost everywhere [8, Remark 1.1.3]. We thus have the following arc-length integral formula [23, Theorem 2.7.6],

$$L_d(\gamma|_{[s,t]}) = \int_s^t \|\gamma'\| \,\mathrm{d}\,\mathfrak{u} \quad \text{for } 0 \le s \le t \le 1.$$

We show that there is no curve connecting e and $-e$ with infimum length. Indeed, if there is one, then it could be realized by a Lipschitz curve. We claim that arc-length variation shares the same solution as energy variation over Lipschitz curves on $[0,1]$:

$$\arg\min_{\gamma} L(\gamma) := \arg\min_{\gamma} \int_0^1 \|\gamma'\| \,\mathrm{d}\,\mathfrak{u} = \arg\min_{\gamma} \int_0^1 \|\gamma'\|^2 \,\mathrm{d}\,\mathfrak{u},$$

where the minimum is taken over all Lipschitz curves $\gamma : [0,1] \to E$ with given endpoints. For this claim, the Cauchy-Schwarz inequality implies that solutions to the energy variation should have arc-length proportional parametrizations. In this case, the energy is exactly the square of the arc-length, so they attain their minima simultaneously.

It remains to show that there is no solution to the corresponding energy variation. Given a Lipschitz curve $\gamma = (\gamma_1, \gamma_2, \ldots) : [0,1] \to E$ connecting e and $-e$, we shall construct a Lipschitz curve $\eta : [0,1] \to E$ whose energy is strictly smaller than γ. Since γ is continuous, it is impossible to have all functions γ_i with $i \geq 2$ being zero. Fix an arbitrary integer $n \geq 3$ such that γ_{n-1} is not a zero function. We modify γ leaving γ_1 and γ_k for $k > n$ unchanged to lower the energy of γ as follows. Define the continuous function $q := \|(\frac{\gamma_2}{c_2}, \ldots, \frac{\gamma_n}{c_n})\|$ and the open subset $A := q^{-1}(0, \infty) \subset [0,1]$. Note that A is not empty since $\gamma_{n-1}([0,1]) \neq \{0\}$ by our choice of n. We modify γ only on A to define a new curve $\eta : [0,1] \to E$ connecting e and $-e$,

$$\eta(t) := (\gamma_1(t), 0, \ldots, 0, c_n q(t), \gamma_{n+1}(t), \ldots), \quad t \in A.$$

For $t \in A$, we have $q(t) > 0$ and thus $q'(t) = \frac{[q^2]'}{2q} = \frac{1}{q} \sum_{i=2}^{n} \frac{\gamma_i(t) \cdot \gamma_i'(t)}{c_i}$. Hence, we obtain the following inequality on A,

$$\begin{aligned}
\|\gamma'\|^2 - \|\eta'\|^2 &= \|(\gamma_2', \ldots, \gamma_n')\|^2 - (c_n q')^2 \\
&= \|(\gamma_2', \ldots, \gamma_n')\|^2 - \langle(\gamma_2', \ldots, \gamma_n'),\, \frac{1}{q}(\frac{c_n}{c_2}\frac{\gamma_2}{c_2}, \ldots, \frac{c_n}{c_n}\frac{\gamma_n}{c_n})\rangle^2 \\
&\geq \|(\gamma_2', \ldots, \gamma_n')\|^2 - \|(\gamma_2', \ldots, \gamma_n')\|^2 \cdot \frac{1}{q^2} \cdot \|(\frac{c_n}{c_2}\frac{\gamma_2}{c_2}, \ldots, \frac{c_n}{c_n}\frac{\gamma_n}{c_n})\|^2 \\
&\geq \|(\gamma_2', \ldots, \gamma_n')\|^2 - \|(\gamma_2', \ldots, \gamma_n')\|^2 \cdot \frac{q^2}{q^2} = 0,
\end{aligned}$$

where in the above two inequalities, we applied respectively the Cauchy-Schwarz inequality and the assumption that $\frac{c_n}{c_i} < 1$ for $1 \leq i \leq n$. However, the obtained inequality $\|\gamma'\|^2 - \|\eta'\|^2 \geq 0$ becomes strict on the set where $\|(\gamma_2'(t), \ldots, \gamma_{n-1}'(t))\| \neq 0$, which by our choice of n is not negligible. It follows that the curve η has strictly lower energy than γ. Since the curve γ is arbitrarily chosen, the energy variation has no solution.

The induced length space $(E, \hat{d})$ is defined via the length structure L_d such that $\hat{d}(x,y)$ is the infimum of $L_d(\gamma)$ for all rectifiable curves γ connecting x and y [23, §2.3.3]. It is shown above that $(E, \hat{d})$ is not geodesic. We are left to show that it is complete. Since $d(x,y) \leq \hat{d}(x,y)$ and (E,d) is complete, it suffices to show the claim that d and $\hat{d}$ induce the same topology on E. To prove this claim, we first consider the case where E is replaced by the unit sphere B in the Hilbert space $\mathbb{R}^\infty$. Denote by d_B and $\hat{d}_B$ the metrics constructed from B in the same way how d and $\hat{d}$ are constructed for E. The distance formula $\hat{d}_B(x,y) = \arccos\langle x, y\rangle$ for $x, y \in B$ holds since we can approximate x, y by points with finitely many non-zero components and apply the distance formulae for finite-dimensional unit spheres. It follows from the distance formula that (B, d_B) and $(B, \hat{d}_B)$ share the same topology. Now we argue that the general case of E for the claim can be reduced to the previous case. Consider the map $f : B \to E$ that sends $(x_1, \cdots, x_n, \ldots) \in B$ to $(c_1 x_1, \ldots, c_n x_n, \ldots) \in E$. Since both the sequence $\{c_i\}_{i \in \mathbb{N}^*}$ and the sequence $\{c_i^{-1}\}_{i \in \mathbb{N}^*}$ are bounded, the map f and its inverse f^{-1} are Lipschitz continuous, with respect to the pair of metric spaces (E,d) and (B, d_B) or the pair of metric spaces $(E, \hat{d})$ and $(B, \hat{d}_B)$. Therefore, (E,d) is homeomorphic to (B, d_B) and $(E, \hat{d})$ is homeomorphic to $(B, \hat{d}_B)$, which thus proves the claim. □

As a complement to Example 2.8, we also present the following counter-example demonstrating the non-existence of barycenters on a metric space that is locally compact and complete but not a length space.

Example 2.10. We endow $\mathbb{R}$ with the metric function $d(x, y) = \phi(|x - y|)$ for $x, y \in \mathbb{R}$, where ϕ is a sub-additive piece-wisely linear function defined as: $\phi(0) = 0$, $\phi(x) = x + 0.5$ for $0 < x < 1$ and $\phi(x) = x + 1$ for $x \geq 1$. This metric space is locally compact since all singletons are open, closed and compact. It is not a proper space since a closed ball with radius 1 contains infinitely many points while each point is an open set, so this ball is not compact. It is a complete space since if $d(x, y) < 0.5$ then $x = y$. We consider the probability measure $\mu := \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ on $(\mathbb{R}, d)$.

Define $f(x) := \int_{\mathbb{R}} d^2(x, y) \,\mathrm{d}\,\mu(y) = \int_{\mathbb{R}} \phi^2(|x - y|) \,\mathrm{d}\,\mu(y)$. We plot these two functions ϕ and f below. Red points are values of functions where they are discontinuous.

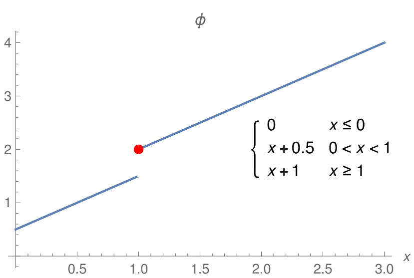


Figure 2.1: $d(x, y) := \phi(|x - y|)$

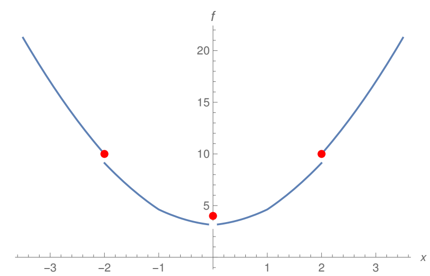


Figure 2.2: $f(x) := \int_{\mathbb{R}} d^2(x, y) \,\mathrm{d}\,\mu(y)$

Then $f(-x) = f(x)$ and f is increasing on $(0, +\infty)$. Moreover, f is continuous on $(0, 1)$ but $\lim_{x \to 0} f(x) < f(0) < f(1)$. This shows that f has no minimum value and thus μ has no barycenter.

2.3 Known properties of Wasserstein barycenters

Wasserstein barycenters are barycenters of probability measures on Wasserstein spaces. In this section, we outline some established properties related to the existence and uniqueness of Wasserstein barycenters. Following the development of [62], to prove the existence, which is not obvious given the conter-examples presented in the last section, we shall begin by constructing Wasserstein barycenters of finitely many measures. This construction will rely on a specific type of multi-marginal optimal transport plans, which we now introduce.

Definition 2.11 (Multi-marginal optimal transport plans)**.** Let (E, d) be a proper metric 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 Θ the set of probability measures on E^n with marginals $\mu_1, \ldots, \mu_n$ in this order. We call $\gamma \in \Theta$ a multi-marginal optimal transport plan (of its marginals) if

$$\int_{E^n} \min_{y \in E} \sum_{i=1}^n \lambda_i \, d(y, x_i)^2 \,\mathrm{d}\,\gamma(x_1, \ldots, x_n) = \min_{\theta \in \Theta} \int_{E^n} \min_{y \in E} \sum_{i=1}^n \lambda_i \, d(y, x_i)^2 \,\mathrm{d}\,\theta(x_1, \ldots, x_n). \tag{2.5}$$

In what follows, the marginal measures μ_i and constants λ_i will be clear from the context, and Definition 2.11 is the sole type of multi-marginal optimal transport problems we shall consider. By

Corollary 2.4, the cost function $\inf_{y\in E}\sum_{i=1}^n \lambda_i\, d(x_i,y)^2$ is continuous with respect to $(x_1,\dots,x_n)\in E^n$. Hence, we can prove the existence of a multi-marginal optimal transport plan γ in the same way as the classic proof for the existence of optimal couplings between two measures [105, Theorem 4.1]. Now we are ready to construct Wasserstein barycenters. It is important to note that although we shall use the notation $\mu_{\mathbb{P}}$ to indicate that a measure is a barycenter of $\mathbb{P}$, this notation should not be interpreted as implying the uniqueness of such a barycenter.

Proposition 2.12 (Construction of Wasserstein barycenters of $\sum_{i=1}^n \lambda_i\,\delta_{\mu_i}$)**.** *Let (E,d) be a proper metric 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 $\mu_1,\dots,\mu_n\in\mathcal{W}_2(E)$ be n probability measures and let γ be a multi-marginal optimal transport plan of them, i.e., satisfying (2.5). If $B: E^n\to E$ is a measurable map such that $B(x_1,\dots,x_n)$ is a barycenter of $\sum_{i=1}^n \lambda_i\,\delta_{x_i}$, then*

1. *$\mu_{\mathbb{P}} := B_{\#}\gamma$ 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 $\mu_{\mathbb{P}}$ and μ_i, where p_i denotes the canonical projection $(x_1,\dots,x_n)\in E^n\mapsto x_i\in E$;*
3. *if $X, X_1,\dots,X_n$ are $n+1$ random variables from a probability space $(\Omega,\mathcal{F},P)$ to (E,d) with law $\mu_{\mathbb{P}},\mu_1,\dots,\mu_n$ such that $\mathbb{E}\, d(X,X_i)^2 = d_W(\mu_{\mathbb{P}},\mu_i)^2$, i.e., (X,X_i) is an optimal transport coupling between $\mu_{\mathbb{P}}$ 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 [104, Lemma 7.1], there are $n+1$ random variables $X, X_1,\dots X_n$ valued in E with laws $\nu,\mu_1,\dots\mu_n$ such that $\mathbb{E}\, d(X,X_i)^2 = d_W(\nu,\mu_i)^2$. We introduce the symbol $\mathtt{x} := (x_1,\dots,x_n)$ to represent a general point in E^n with components $x_1,\dots,x_n$ in this order. Since $\mu_i = p_{i\#}\gamma$, we have

$$
\begin{aligned}
\sum_{i=1}^n \lambda_i\, d_W(\mu_{\mathbb{P}},\mu_i)^2 &\leq \sum_{i=1}^n \int_{E^n} \lambda_i\, d(B(\mathtt{x}),x_i)^2 \,\mathrm{d}\,\gamma(\mathtt{x}) = \int_{E^n} \min_{y\in E}\sum_{i=1}^n \lambda_i\, d(y,x_i)^2\,\mathrm{d}\,\gamma(\mathtt{x})\\
&\leq \mathbb{E}\,\min_{y\in E}\sum_{i=1}^n \lambda_i\, d(y,X_i)^2 \leq \mathbb{E}\sum_{i=1}^n \lambda_i\, d(X,X_i)^2\\
&= \sum_{i=1}^n \lambda_i\, d_W(\nu,\mu_i)^2,
\end{aligned}
$$

where we sequentially applied the definitions of $\mu_{\mathbb{P}} = B_{\#}\gamma$, $d_W(\mu_{\mathbb{P}},\mu_i)$, γ, B and $X, X_1,\dots,X_n$. Since ν is arbitrarily chosen, it follows that $\mu_{\mathbb{P}}$ is a Wasserstein barycenter. By setting $\nu=\mu_{\mathbb{P}}$ in the above inequality, we actually obtain an equality. Our last two statements follow from this inequality. Firstly, this equality implies that $\sum_{i=1}^n \lambda_i\, d_W(\mu_{\mathbb{P}},\mu_i)^2 \leq \sum_{i=1}^n\int_{E^n}\lambda_i\, d(B(\mathtt{x}),x_i)^2\,\mathrm{d}\,\gamma(\mathtt{x})$ is indeed always an equality, which proves the second statement. Secondly, it also implies that the law of $(X_1,\dots,X_n)$ 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. □

The general existence of Wasserstein barycenters was first established in [62], which is based on the consistency of Wasserstein barycenters.

Theorem 2.13 (Consistency of Wasserstein barycenters, [62])**.** *Let (E,d) be a proper metric space. Fix a probability measure $\mathbb{P} \in (\mathcal{W}_2(\mathcal{W}_2(E)), d_{\mathbb{W}})$ on $(\mathcal{W}_2(E), d_W)$. Given a sequence of measures $\mathbb{P}_j \in \mathcal{W}_2(\mathcal{W}_2(E))$ with their corresponding barycenters $\mu_{\mathbb{P}_j} \in \mathcal{W}_2(E)$, if $d_{\mathbb{W}}(\mathbb{P}_j, \mathbb{P}) \to 0$ as j goes to $+\infty$, then $d_W(\mu_{\mathbb{P}_j}, \mu_{\mathbb{P}}) \to 0$ for some barycenter $\mu_{\mathbb{P}}$ of $\mathbb{P}$ up to extracting a subsequence of $\mu_{\mathbb{P}_j}$.*

Recall that finitely supported probability measures are dense in Wasserstein spaces [105, Theorem 6.18], so the consistency of Wasserstein barycenters together with the previous construction of Wasserstein barycenters (Proposition 2.12) implies the following theorem.

Theorem 2.14 (Existence of Wasserstein barycenters, [62])**.** *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 Theorem 2.13, we may need to pass to a subsequence of Wasserstein barycenters $\mu_{\mathbb{P}_j}$ and the limit barycenter $\mu_{\mathbb{P}}$ is not known in advance. Hence, Theorem 2.13 will be enhanced if we can assert the uniqueness of barycenters under some additional assumptions, as follows.

Proposition 2.15 (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 $d_W(\mu,\nu)^2 = \int_E d(x,T(x))^2 \,\mathrm{d}\,\mu$, 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 given point in $\mathcal{W}_2(E)$, as shown by [90, Theorem 7.19] and [58, 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 $d_W(\mu,\cdot)^2$ 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

$$d_W(\mu, \lambda\,\nu_1 + (1-\lambda)\nu_2)^2 \le \lambda\, d_W(\mu,\nu_1)^2 + (1-\lambda) d_W(\mu,\nu_2)^2. \tag{2.6}$$

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 (2.6) cannot be an equality unless $\lambda = 0$ or $\lambda = 1$. Indeed, if (2.6) is an equality for some $0 < \lambda < 1$, then by setting $\gamma := \lambda\,\gamma_1 + (1-\lambda)\,\gamma_2$ we have

$$\begin{aligned}
\lambda\, d_W(\mu,\nu_1)^2 + (1-\lambda) d_W(\mu,\nu_2)^2 &= d_W(\mu, \lambda\,\nu_1 + (1-\lambda)\nu_2)^2 \\
&\le \int_{E\times E} d(x,y)^2 \,\mathrm{d}\,\gamma(x,y) \\
&= \lambda\, d_W(\mu,\nu_1)^2 + (1-\lambda) d_W(\mu,\nu_2)^2,
\end{aligned}$$

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

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

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 2.14 is unique. □

Remark 2.16. Under the assumptions of Proposition 2.15, the optimal transport plan between $\mu \in \mathcal{A}$ and $\nu \in \mathcal{W}_2(M)$ is unique. Indeed, if we set $\nu_1 = \nu_2 = \nu$, then (2.6) becomes an equality for any $\lambda \in [0,1]$. Hence, given any two optimal transport plans γ_1 and γ_2 between measures μ and ν, our arguments on the measure $\gamma := \lambda\,\gamma_1 + (1-\lambda)\,\gamma_2$ imply that they must coincide.

There are many setups in which we can apply Proposition 2.15. We typically choose $\mathcal{A}$ as the set of absolutely continuous measures with respect to some given reference measure. The following lemma will be applied in Section 4.3.2, whose particular case, ensuring that $\mathcal{A}$ is a Borel set of $(\mathcal{W}_2(E), d_W)$, is now needed in this section.

Lemma 2.17. *Let (E,d) be a metric space equipped with a σ-finite Borel measure μ on E. Assume that μ is outer regular, i.e., for any Borel set $N \in \mathcal{B}(E)$,*

$$\mu(N) = \inf\{\mu(O) \mid O \text{ open neighborhood of } N\}.$$

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

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

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 (possibly null) mass to some open neighborhood of every point in E, then with respect to the Wasserstein metric topology, $\mathcal{E}_{\epsilon,\delta}$ is a closed set and $\mathcal{A}$ is a Borel set.

Proof. Our proof is based on [58, 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 [17, Corollary 8.2.10], we have

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

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: $\forall\, N \in \mathcal{B}(E),\, \mu(N) = 0 \implies \nu(N) = 0$. We now prove the reverse inclusion. 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 [17, Theorem 2.5.7, Proposition 2.6.4].

Given a proper space E and a locally finite Borel measure μ on E, observe that μ gives finite mass to compact sets, and every open subset of E is σ-compact. It follows that μ is outer regular [98, Theorem 6 of §2.7] and also σ-finite. Since Wasserstein convergence implies weak convergence, the set $\mathcal{E}_{\epsilon,\delta}$ is closed with respect to the Wasserstein metric. It follows that $\mathcal{A}$ is a Borel set of $\mathcal{W}_2(E)$. $\square$

Remark 2.18. On a metric space, any finite Borel measure is outer regular, see [17, Definition 7.1.5, Theorem 7.1.7] or [16, 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 [98, 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 $E = \cup_{i\geq 1} O_i$. We also mention that there exists a σ-finite and locally finite but not outer regular Borel measure on a locally compact Hausdorff space [19, problem 5 of Exercise §1, INT IV.119].

Thanks to Proposition 2.15 and Lemma 2.17, 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 [105, Theorem 10.41] or Gigli [42, Theorem 7.4];
2. compact finite dimensional Alexandrov spaces, see Bertrand [13, Theorem 1.1];
3. for $K \in \mathbb{R}$ and $N \geq 1$, non-branching $\mathrm{CD}(K, N)$ spaces, see Gigli [43, Theorem 3.3];
4. for $K \in \mathbb{R}$ and $N \geq 1$, $\mathrm{RCD}^*(K, N)$ spaces, see Gigli, Rajala and Sturm [44, 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 [25, Theorem 1.1];
6. (2-)essentially non-branching spaces with qualitatively non-degenerate reference measures, see Kell [56, 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 2.16) between an absolutely continuous probability measure and a given probability measure is induced by a measurable map, allowing us to apply Proposition 2.15.

Chapter 3

Absolutely continuous barycenters of finitely many measures

Given the established existence and uniqueness of Wasserstein barycenters (under mild assumptions) on Riemannian manifolds (M, d_{g}), we are now prepared to demonstrate the absolute continuity of these barycenters with respect to the volume measure Vol. In this chapter, we turn our attention to Wasserstein spaces $\mathcal{W}_2(M)$ defined over Riemannian manifolds and aim to prove that the unique Wasserstein barycenter of a finite collection of measures is absolutely continuous with respect to Vol, provided that at least one measure in the collection possesses this property.

To achieve this, we adapt the proof of absolute continuity developed by Kim and Pass [58] for the specific case of compact manifolds. Our adaptation will incorporate a geometric perspective and will address both the case of finitely many measures with compact support and the more general scenario where compactness assumptions are relaxed.

3.1 Lipschitz continuity of optimal transport maps

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 3.1 (Regularity of displacement interpolations, [105, Theorem 8.5, Theorem 8.7])**.** *Let (M, d_{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.*

In [105, Chapter 8], the Lipschitz continuity presented in Proposition 3.1 is demonstrated as a consequence of Mather's shortening lemma. Furthermore, the subsequent statement regarding absolute continuity follows from the property that Lipschitz maps preserve sets of Lebesgue measure zero. An alternative approach to establishing Lipschitz continuity is provided by Bernard and Buffoni [12] through the Hamilton-Jacobi equation, a method extended to non-compact settings by Fathi and Figalli [36]. For the specific case of absolutely continuous measures on Euclidean spaces, McCann [72, Proposition 1.3] presented a concise proof of Lipschitz continuity. Further relevant references can be found in the bibliographical notes of Chapter 8 in Villani [105]. The objective of this subsection is to generalize the result stated in Proposition 3.1.

We deduce the following Lipschitz continuity from the c-concave functions defined in Lemma 2.5, which are related to barycenter selection maps and thus Wasserstein barycenters of the probability measures $\lambda_1\delta_{\mu_1} + \sum_{i=2}^n \lambda_i\, \delta_{\delta_{x_i}}$. Recall that a measurable barycenter selection map $B : M^n \to M$ (Corollary 2.4) sends $\mathrm{x} = (x_1, \dots, x_n) \in M^n$ to a barycenter of $\sum_{i=1}^n \lambda_i\, \delta_{x_i}$. In the following results, the constants $\lambda_i, 1 \le i \le n$ for B are given in the context.

As a convention to simplify the notation, we denote by $\mathrm{x}' = (x_2, \dots, x_n) \in M^{n-1}$ the last $n-1$ components of $\mathrm{x} \in M^n$, and identify the pair (x_1, x') with x. Introduce the following two projection maps,

$$\begin{aligned} p_1 : M \times M^{n-1} \to M, &\quad p_1(x_1, \mathrm{x}') = x_1; \\ p_2 : M \times M^{n-1} \to M^{n-1}, &\quad p_2(x_1, \mathrm{x}') = \mathrm{x}'. \end{aligned}$$

Lemma 3.2 (Lipschitz continuous maps $F = \exp(-\nabla g_1)$)**.** *Let (M, d_{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 $\mathrm{x}' = (x_2, \dots, x_n) \in M^{n-1}$. 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 $g_1 : y \in M \mapsto -1/\lambda_1 \sum_{i=2}^n \lambda_i\, c(y, x_i)$ (the function c is defined in (1.13)). 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:*

$$z \in Y \text{ and } x_1 = F(z) \iff x_1 \in X \text{ and } z \text{ is a barycenter of } \sum_{i=1}^n \lambda_i\, \delta_{x_i}. \tag{3.1}$$

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

Proof. According to Lemma 2.6, g_1 is smooth in a neighborhood of Y, which implies that F is continuously differentiable on this neighborhood. Since g_1 restricted to Y is a c-concave function (Lemma 2.5) 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 [73, Lemma 7] shows that

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

Note that though the cited lemma of McCann is proven for compact manifolds, the arguments of its proof only depend on the existence of the 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 2.5) and thus

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

which implies the characterization (3.1). $F(Y) = X$ follows from (3.1) 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 2.12 shows that $\mu_{\mathbb{P}} = B_\#\gamma$ is a Wasserstein barycenter of $\mathbb{P}$. Moreover, since $p_1(x_1, \mathrm{x}') = x_1 = F(B(x_1, \mathrm{x}'))$ for $x_1 \in X$ by (3.1), Proposition 2.12 shows that $(B, p_1)_\#\gamma = (B, F \circ B)_\#\gamma = (\mathrm{Id}, F)_\#\mu_{\mathbb{P}}$ is an optimal transport plan between $\mu_{\mathbb{P}}$ and μ_1. ◻

3.2 Divide-and-conquer via conditional measures

Lemma 3.2 implies that any barycenter selection map on $X \times \{\mathrm{x}'\}$ is injective (note that this is different from being unique). The following lemma by Kim and Pass [57, Lemma 3.5] generalizes this injectivity, and it will help us to generalize Lemma 3.2.

Lemma 3.3. *Let (M, d_{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 γ is a multi-marginal optimal transport plan with marginals $\mu_1, \ldots, \mu_n$, then*

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

where $\operatorname{bary}(\{(x_1, \ldots, x_n)\})$ *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 [57, Lemma 3.5], which is based on c-cyclical monotonicity and Lemma 2.6. 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 [57, proof of Theorem 6.1]. The following proposition constructs an optimal transport map pushing forward $\mu_{\mathbb{P}} := B_{\#}\gamma$ to μ_1 when $\mu_i, 2 \leq i \leq n$, are discrete measures and thus generalizes Lemma 3.2. The optimal transport map may fail to be a Lipschitz map, but it is a disjoint union of Lipschitz maps defined as follows. 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 1.7) to deduce further conclusions from F_j's Lipschitz continuity.

Proposition 3.4. *Let (M, d_{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 probability 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., probability measures supported in 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 $\mu_{\mathbb{P}} := B_{\#}\gamma$ is a barycenter of $\mathbb{P} := \sum_{i=1}^n \lambda_i \, \delta_{\mu_i}$. The measure $\mu_{\mathbb{P}}$ is supported in a disjoint union of at most countably many compact sets, and on each of them Lemma 3.2 defines a Lipschitz continuous map with a compact subset $X \subset M$ and a point $\mathrm{x}' \in M^{n-1}$ such that $X \times \{\mathrm{x}'\}$ is contained in the support of γ. Denote by F the disjoint union of these Lipschitz maps.* $(\mathrm{Id}, F)_{\#}\mu_{\mathbb{P}}$ *is an optimal transport plan between $\mu_{\mathbb{P}}$ and μ_1.*

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

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

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

Proof. Proposition 2.12 shows that $\mu_{\mathbb{P}}$ is a Wasserstein barycenter. Let us reveal more details of γ. The measure $\pi := p_{2\#}\gamma$ on M^{n-1} is discrete since its marginals $\mu_2, \ldots, \mu_n$ are so. Denote by $\{\mathrm{x}'_j\}_{j \in J}$ the set of all atoms of π, where $J \subset \mathbb{N}$ is an at most countable set (Lemma 1.4). For each $j \in J$, we introduce the following definitions. Define $\pi_j := \pi(\{\mathrm{x}'_j\}) > 0$ and define $X_j := p_1(\operatorname{supp}\gamma \cap (M \times \{\mathrm{x}'_j\}))$. Applying Lemma 3.2 to X_j and $\mathrm{x}'_j \in M^{n-1}$, 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 is supported in the union $\bigcup_{j\in J}\{\mathsf{x}_j\}$, γ is supported in the union $\bigcup_{j\in J} X_j \times \{\mathsf{x}'_j\}$. As in Lemma 3.3, we denote by $\mathrm{bary}(\{(x_1,\ldots,x_n)\})$ 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 3.2, $z \in \mathrm{bary}(\{(F_i(z),\mathsf{x}'_i)\}) \cap \mathrm{bary}(\{(F_k(z),\mathsf{x}'_k)\})$. Since $\bigcup_{j\in J} X_j\times\{\mathsf{x}'_j\} \subset \operatorname{supp}\gamma$ and $F_j(Y_j) = X_j$, Lemma 3.3 forces that $\mathsf{x}'_i = \mathsf{x}'_k$ and thus $i = k$, which implies our claim. Define F as the disjoint union of $F_j, j\in J$, i.e., $F|_{Y_j} = F_j$. Since $p_1(x,\mathsf{x}'_j) = x = F(B(x,\mathsf{x}'_j))$ for $x\in X_j$ and γ is supported in the union $\bigcup_{j\in J} X_j\times\{\mathsf{x}'_j\}$, it follows from Proposition 2.12 that $(B,p_1)_\#\gamma = (B, F\circ B)_\#\gamma = (\mathrm{Id},F)_\#\mu_{\mathbb{P}}$ is an optimal transport plan between $\mu_{\mathbb{P}}$ and μ_1. Since the union $\bigcup_{j\in J} Y_j$ is the domain of F and $F_\#\mu_{\mathbb{P}} = \mu_1$, $\mu_{\mathbb{P}}$ is supported in 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 $\mathrm{d}\,\gamma(\mathsf{x}) = \gamma(\mathrm{d}\,\mathsf{x},\mathsf{x}')\,\mathrm{d}\,\pi(\mathsf{x}')$. In accordance with the notation in Lemma 3.2, for each $j\in J$, denote by $(y_j^2,\ldots,y_j^n) =: \mathsf{x}'_j \in M^{n-1}$ the re-writing of x'_j in components and introduce

$$\nu_j := \frac{1}{\pi_j}\mu_1|_{X_j}, \quad \mathbb{Q}_j := \lambda_1\,\delta_{\nu_j} + \sum_{l=2}^n \lambda_l\,\delta_{\delta_{y_j^l}}, \quad \nu_{\mathbb{Q}_j} := B_\#\gamma(\cdot,\mathsf{x}'_j).$$

Lemma 3.2 implies that $F_{j\,\#}\nu_{\mathbb{Q}_j} = \nu_j$ and $\nu_{\mathbb{Q}_j}$ is a barycenter of $\mathbb{Q}_j$. For $\mathsf{R}\in\mathcal{B}(M^n)$ and $j\in J$, thanks to the property $\pi_j = \pi(\{\mathsf{x}'_j\}) > 0$, we have $\gamma[\mathsf{R}\cap(M\times\{\mathsf{x}'_j\})] = \gamma(\mathsf{R},\mathsf{x}'_j)\,\pi_j$ by Definition 1.7, which implies that $\gamma(\cdot,\mathsf{x}'_j)$ is supported in $X_j\times\{\mathsf{x}'_j\}$. Since γ is supported in the union $\bigcup_{j\in J} X_j \times \{\mathsf{x}'_j\}$, we obtain the following equality by choosing R of the form $A\times M^{n-1}$ with $A\in\mathcal{B}(M)$,

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

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

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

By setting $f = p_1$ and $f = B$, we obtain

$$\mu_1 = \sum_{j\in J}\pi_j\,\nu_j \quad \text{and} \quad \mu_{\mathbb{P}} = \sum_{j\in J}\pi_j\,\nu_{\mathbb{Q}_j}. \tag{3.2}$$

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.

We now assume the existence of a common Lipschitz constant C of all $F_j, j\in J$. As the images of a Borel set under the Lipschitz maps F_j are not necessarily Borel sets, we state the regularity of the volume measure as follows (c.f. [105, Proof of Theorem 8.7]) to simplify the subsequent arguments. 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 $\mathrm{Vol}(W_j)\le C^m\,\mathrm{Vol}(N\cap Y_j)$ [100, Proposition 12.6, Proposition 12.12, Remark after Proposition 12.12].

For $j\in J$, with the re-writing $\mathsf{x}'_j = (y_j^2, y_j^3,\ldots,y_j^n)\in M^{n-1}$, $\gamma(\cdot,\mathsf{x}'_j)$ is equal to the product measure $\nu_j\otimes\delta_{y_j^2}\otimes\cdots\otimes\delta_{y_j^n}$. It follows from Lemma 3.2 that

$$F_{j\,\#}\nu_{\mathbb{Q}_j} = \nu_j = \frac{1}{\pi_j}\mu_1|_{X_j}. \tag{3.3}$$

As $F_j(N \cap Y_j) \subset W_j$, we have $\nu_{\mathbb{Q}_j}(N \cap Y_j) \leq \nu_{\mathbb{Q}_j}(F_j^{-1}(W_j)) = \nu_j(W_j)$. Therefore, according to (3.2) and (3.3),

$$\mu_{\mathbb{P}}(N) = \sum_{j \in J} \pi_j \, \nu_{\mathbb{Q}_j}(N \cap Y_j) \leq \sum_{j \in J} \pi_j \frac{1}{\pi_j} \mu_1|_{X_j}(W_j) = \sum_{j \in J} \mu_1(W_j) = \mu_1(\bigcup_{j \in J} W_j), \tag{3.4}$$

where we used $W_j \subset X_j$ and $\mu_1(X_i \cap X_k) = 0$ if $i \neq k \in J$. Since $Y_j, j \in J$ are disjoint, $\mathrm{Vol}(\bigcup_{j \in J} W_j) \leq C^m \sum_{j \in J} \mathrm{Vol}(N \cap Y_j) \leq C^m \,\mathrm{Vol}(N)$. 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 $\mu_{\mathbb{P}}(N) \leq \mu_1(\bigcup_{j \in J} W_j) \leq \epsilon$ by (3.4). Therefore, the implication $\mu_1 \in \mathcal{E}_{\epsilon,\delta} \implies \mu_{\mathbb{P}} \in \mathcal{E}_{\epsilon,\delta/C^m}$ is proven, which concludes our proof. $\square$

Remark 3.5. Figuratively speaking, the sets $X_j, j \in J$ create a tiling of the support of μ_1 and the points $\mathrm{x}'_j, j \in J$ pull them apart (via barycenter selection maps) into disjoint sets $Y_j, j \in J$, which contain different pieces of the support of $\mu_{\mathbb{P}}$ separately.

3.3 Absolute continuity implied by compactness

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 $2 \leq i \leq n$ with discrete measures to apply Proposition 3.4. If μ_1 is absolutely continuous, then $\mathbb{P}$ has a unique barycenter $\mu_{\mathbb{P}}$, which is approximated by the barycenters of the approximating sequence (Theorem 2.13). Recalling that the sets $\mathcal{E}_{\epsilon,\delta}$ (defined in Lemma 2.17) provide a full characterization of absolutely continuous measures and are closed under weak convergence, the remaining task to prove the absolute continuity of $\mu_{\mathbb{P}}$ is to establish the existence of a common Lipschitz constant C for the optimal transport map F (defined as in Lemma 3.2) across all elements of the approximating sequence. Proposition 3.4 will then allow us to conclude the result. It is important to note that the domain Y of the map F changes along the approximating sequence, thus the existence of the constant C is not a straightforward consequence of compactness. More precisely, our goal is to prove:

Theorem 3.6 (Absolute continuity of the barycenter of $\sum_{i=1}^n \lambda_i \, \delta_{\mu_i}$)**.** *Let (M, d_{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 $\mu_{\mathbb{P}}$ of $\mathbb{P} := \sum_{i=1}^n \lambda_i \, \delta_{\mu_i}$ is absolutely continuous with compact support.*

Proof. The uniqueness of $\mu_{\mathbb{P}}$ and the compactness of $\mathrm{supp}(\mu_{\mathbb{P}})$ follow from Proposition 2.15, Proposition 2.12, and Corollary 2.4. We are left to show the absolute continuity of $\mu_{\mathbb{P}}$. To prove it, we approximate each μ_i for $2 \leq i \leq n$ in $(\mathcal{W}_2(M), d_W)$ by a sequence of finitely supported probability 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 consistency of Wasserstein barycenters (Theorem 2.13), the unique barycenter $\mu_{\mathbb{P}_j}$ of $\mathbb{P}_j$ converges in $(\mathcal{W}_2(M), d_W)$ to the unique barycenter $\mu_{\mathbb{P}}$ 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 $\mathrm{x}' := (x_2, \ldots, x_n) \in M^{n-1}$ such that $X \times \{\mathrm{x}'\} \subset \mathrm{supp}\, \gamma_j$. Applying Lemma 3.2 to X and 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 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 the point z is a barycenter of $\sum_{i=1}^n \lambda_i\, \delta_{x_i}$ (Lemma 3.2) with $x_1 := F(z)$, it is a critical point of the following map

$$w \in M \mapsto \sum_{i=1}^n \lambda_i\, d_{\mathsf{g}}(w, x_i)^2, \tag{3.5}$$

which implies $\sum_{i=1}^n \lambda_i \nabla d_{x_i}^2(z) = 0$ thanks to Lemma 2.6. Hence, by definition of g_1, we get $\nabla d_{x_1}^2/2(z) = \nabla g_1(z)$. Moreover, Lemma 2.6 enables us to apply Lemma 1.43 to compute the differential of F at z,

$$\begin{aligned} \mathsf{D}_z F = \mathsf{D}_z \exp(-\nabla g_1) &= [\mathsf{D}_{-\nabla g_1(z)} \exp_z] \circ (\mathrm{Hess}_z\, d_{x_1}^2/2 - \mathrm{Hess}_z\, g_1) \\ &= [\mathsf{D}_{-\nabla g_1(z)} \exp_z] \circ \frac{1}{2\lambda_1} \sum_{i=1}^n \lambda_i\, \mathrm{Hess}_z\, d_{x_i}^2. \end{aligned} \tag{3.6}$$

In (3.6), $\sum_{i=1}^n \lambda_i\ \mathrm{Hess}_z\, d_{x_i}^2$ is positive semi-definite since z reaches the global minimum of the map (3.5). We now bound (3.6) via compactness as follows. Consider the compact set $\mathsf{A} := \mathrm{supp}(\mu_1) \times \cdots \times \mathrm{supp}(\mu_n) \subset M^n$. Corollary 2.4 implies that the set of all barycenters of $\sum_{i=1}^n \lambda_i\, \delta_{y_i}$ for $(y_1, \ldots, y_n)$ running through the set A is compact. Moreover, by our construction of $\mathbb{P}_j$, the union of the supports of $\mu_{\mathbb{P}}$, μ_i, $\mu_{\mathbb{P}_j}$ and μ_i^j for $1 \le i \le n$ and $j \ge 1$ is compact. Hence, independent of z, j and x', $\mathsf{D}_{-\nabla g_1(z)} \exp_z$ is uniformly bounded (in norm) and $\sum_{i=1}^n \lambda_i\ \mathrm{Hess}_z\, d_{x_i}^2$ is uniformly bounded from above by the Rauch comparison theorem for Hessians of distance functions, which is applicable here and gives a constant upper bound thanks to the compactness, see [30, Lemma 3.12 and Corollary 3.13] or [80, 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 3.4 to measures $\mu_1, \mu_2^j, \ldots, \mu_n^j$, we have for $\epsilon, \delta > 0$, $\mu_1 \in \mathcal{E}_{\epsilon,\delta} \implies \mu_{\mathbb{P}_j} \in \mathcal{E}_{\epsilon,\delta/C^m}$ since $\mu_{\mathbb{P}_j}$ is the unique barycenter of $\mathbb{P}_j$. As $\mu_{\mathbb{P}_j}$ converges to $\mu_{\mathbb{P}}$ weakly, Lemma 2.17 shows that all measures $\mu_{\mathbb{P}_j}$ for $j \ge 1$ and $\mu_{\mathbb{P}}$ are absolutely continuous since μ_1 is so. □

3.4 Absolute continuity without compactness

Theorem 3.6 can be further generalized to the case where the measures do not have compact support. This extension is achieved by decomposing the support of a multi-marginal optimal transport plan γ into a countable union of compact sets. Such a decomposition is feasible because complete Riemannian manifolds are proper metric spaces.

Theorem 3.7 (Absolute continuity of Wasserstein barycenter of finitely many measures)**.** *Let (M, d_{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 n probability measures $\mu_i \in \mathcal{W}_2(M), 1 \le i \le n$ such that μ_1 is absolutely continuous. The unique barycenter $\mu_{\mathbb{P}}$ of $\mathbb{P} := \sum_{i=1}^n \lambda_i\, \delta_{\mu_i}$ is absolutely continuous.*

Proof. According to the construction of Wasserstein barycenters in Proposition 2.12 and the uniqueness of $\mu_{\mathbb{P}}$, we can write $\mu_{\mathbb{P}} = B_{\#}\gamma$ with a measurable barycenter selection map B and a multi-marginal optimal transport plan γ of marginal measures $\mu_1, \mu_2, \ldots, \mu_n$ in this order.

Since M^n is a manifold and γ is a probability measure, there exist at most countably many compact sets $K_j \subset M^n, j \in J \subset \mathbb{N}$ such that $k_j := \gamma(K_j) > 0$, $\sum_{j \in J} k_j = 1$ and $\gamma(K_i \cap K_j) = 0$

for indices $i \neq j$. To justify the existence of these sets, we choose K_j as a closed metric annulus of M^n with a fixed center and two finite radii. To decide the radii, we require that $K_i \cap K_j$ is either empty or the boundary set of a metric ball. For our choice of the radii, it suffices to show that γ can only give non-zero mass to at most countably many boundary sets of metric balls. Recall that uncountable sum of positive real numbers must diverge (Lemma 1.4), so the previously required argument follows from the fact that γ is a probability measure.

For $j \in J$, we define the probability measure $\gamma_j := \frac{1}{k_j}\gamma|_{K_j}$ to be the normalized restriction of γ to the set K_j. By assumptions, for $N \in \mathcal{B}(M)$,

$$\mu_{\mathbb{P}}(N) = \gamma(B^{-1}(N)) = \sum_{j\in J} k_j\, \gamma_j(B^{-1}(N)) = \sum_{j\in J} k_j\, [B_{\#}\gamma_j](N). \tag{3.7}$$

For $j \in J$, denote by $\nu_1^j, \dots, \nu_n^j \in \mathcal{W}_2(M)$ the n marginals of γ_j in this order, and define $\mathbb{Q}_j := \sum_{i=1}^n \lambda_i\, \delta_{\nu_i^j}$, $\mu_{\mathbb{Q}_j} := B_{\#}\gamma_j$. We prove by contradiction that γ_j must be a multi-marginal optimal transport plan of its marginals (c.f. [105, Theorem 4.6]). Indeed, if this is not true and γ_j' is a multi-marginal optimal transport plan of the marginals of γ_j, then γ no longer satisfies our assumption of being optimal, since its cost (i.e., the integral (2.5)) becomes strictly bigger than the cost of the measure $\gamma' := \gamma|_{M^n\setminus K_j} + k_j\, \gamma_j'$. It follows from Proposition 2.12 that $\mu_{\mathbb{Q}_j} = B_{\#}\gamma_j$ is the unique Wasserstein barycenter of $\mathbb{Q}_j$. Since γ_j has compact support and the first marginal μ_1 of γ is absolutely continuous, all marginals $\nu_1^j, \dots, \nu_n^j$ have compact support and the first one ν_1^j is absolutely continuous. Hence, by Theorem 3.6, the barycenter $\mu_{\mathbb{Q}_j}$ of $\mathbb{Q}_j$ is absolutely continuous. According to (3.7), if $\mathrm{Vol}(N) = 0$, then $\mu_{\mathbb{P}}(N) = \sum_{j\in J} k_j\, \mu_{\mathbb{Q}_j}(N) = 0$. Therefore, the probability measure $\mu_{\mathbb{P}}$ is absolutely continuous. $\square$

Remark 3.8. The proof of Theorem 3.6 involves three steps:

1. Lemma 3.2 handles the case where μ_i $(2 \leq i \leq n)$ are Dirac measures, leveraging the product structure of $\gamma = \mu_1 \otimes \cdots \otimes \mu_n$.

2. Proposition 3.4 establishes an estimate based on the existence of a uniform Lipschitz constant for the case where μ_i $(2 \leq i \leq n)$ are discrete measures, building upon the previous case.

3. Compactness is used to obtain a uniform Lipschitz constant for approximation sequences of discrete measures converging to the given compactly supported measures μ_i $(2 \leq i \leq n)$.

To show absolute continuity for the discrete marginal case (step 2) alone, measure-theoretic arguments similar to the proof of Theorem 3.7 suffice, employing a “divide-and-conquer” strategy. As seen in Proposition 3.4, this corresponds to the use of conditional measures. However, this correspondence might be subtle, as Proposition 3.4 primarily prepares Lipschitz constant arguments for Theorem 3.6 to handle uncountable supports, which goes beyond the scope of the measure-theoretic approach.

Chapter 4

Absolute continuity via lower Ricci curvature bounds

In the last chapter, we have seen that for $\mathbb{P} := \sum_{i=1}^{n} \lambda_i \, \delta_{\mu_i}$ with μ_1 being absolutely continuous, Kim and Pass' proof of the absolute continuity of the (unique) barycenter $\mu_{\mathbb{P}}$ of $\mathbb{P}$ remains valid for non-compact manifolds M (Theorem 3.6). 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 $\mu_{\mathbb{P}_j}$ are already shown to be absolutely continuous. Thanks to the consistency of Wasserstein barycenters (Theorem 2.13), $\mu_{\mathbb{P}_j}$ converges to $\mu_{\mathbb{P}}$ weakly. However, this is not sufficient to ensure that $\mu_{\mathbb{P}}$ is also absolutely continuous. To overcome this difficulty, Kim and Pass [58] imposed a uniform upper density bound on $\mu_{\mathbb{P}_j}$'s, which forced them to include the assumption that $\mathbb{P}$ gives mass to a set of absolutely continuous probability measures whose density functions are uniformly bounded.

In this chapter, instead of following their quantitative approach, we seek for proper integral functionals $\mathcal{G}$ 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 [24], [105, Theorem 29.20], [90, Chapter 7], and [5, Chapter 15]. We summarize their assumptions and conclusions in Lemma 4.6. Additionally, we aim to control the value of $\mathcal{G}$ at $\mu_{\mathbb{P}_j}$ 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 [105], focus on the λ-convexity of $\mathcal{G}$, a widely studied property that would satisfy our requirements if we tolerate some independent constants in its inequality expression of convexity (Proposition 4.3). 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 [69, 96, 97]. Proposition 4.3 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 4.9 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 4.13), 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, d_{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.*

4.1 Hessian equality of Wasserstein barycenters

In this section, we prove the Hessian equality for Wasserstein barycenters of finitely many measures (Theorem 4.1). A similar property is named as the *2nd order balance* (inequality) by Kim and Pass [58, Theorem 4.4], but being an equality instead of an inequality is crucial for our proof of Proposition 4.3. Let us take a special case to illustrate this equality. Consider the reduced case in Lemma 3.2. Namely, take n (≥ 2) positive numbers $\lambda_i > 0$ such that $\sum_{i=1}^n \lambda_i = 1$ and denote by $\mu_{\mathbb{P}}$ the barycenter of $\mathbb{P} := \sum_{i=1}^n \lambda_i \, \delta_{\mu_i}$, where μ_1 is absolutely continuous with compact support and $\mu_i = \delta_{x_i}, 2 \leq i \leq n$, 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 $\phi_i(z) := c(z, x_i), 2 \leq i \leq n$. Thanks to Lemma 3.2 and Lemma 2.6, if z is in the support of $\mu_{\mathbb{P}}$, then z is not in the cut locus of any x_i, which implies $\exp(-\nabla \phi_i)_{\#}\mu_{\mathbb{P}} = \mu_i$ for $2 \leq i \leq n$. 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.

The Hessian equality (4.1) to prove is a second-order relation. We first demonstrate a first-order counterpart of this equality using the conclusion of Proposition 2.12 that relates barycenters in manifolds to Wasserstein barycenters.

Theorem 4.1 (Hessian equality for Wasserstein barycenters)**.** *Let (M, d_{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 $\mu_{\mathbb{P}}$ 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 $\mu_{\mathbb{P}}$ forward to μ_i, where ϕ_i is a c-concave function given by Theorem 1.27.*

For $\mu_{\mathbb{P}}$-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

$$\sum_{i=1}^n \lambda_i \operatorname{Hess}_x \phi_i = 0. \tag{4.1}$$

Proof. By Theorem 3.6, $\mu_{\mathbb{P}}$ is absolutely continuous with compact support. We now apply Proposition 2.12 to $\mathbb{P}$. Since $\mu_{\mathbb{P}}$ is the unique barycenter of $\mathbb{P}$, it coincides with the barycenter constructed in Proposition 2.12. Consider the identity map $\mathrm{Id} : (M, \mathcal{B}(M), \mu_{\mathbb{P}}) \to M$ as a random variable taking values in M. It has law $\mu_{\mathbb{P}}$, and the random variable $F_i = F_i \circ \mathrm{Id}$ has law μ_i for $1 \leq i \leq n$. Proposition 2.12 implies that for $\mu_{\mathbb{P}}$-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 $\mu_{\mathbb{P}}(\Omega) = 1$ 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

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

By Lemma 2.6, 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 $\nabla \phi_i(x) = \frac{1}{2} \nabla d^2_{F_i(x)}|_{w=x}$ holds as both gradients exist [30, 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 $\phi_i, 1 \le i \le n$. 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 (approximate) Hessians of f and $\phi_i, 1 \le i \le n$, all exist. Let Ω_2 be the set of density points of Ω. We have $\mathrm{Vol}(\Omega \setminus \Omega_1) = 0$ by Proposition 1.23, and $\mathrm{Vol}(\Omega \setminus \Omega_2) = 0$ by [35, 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 (1.9). Besides, noting that ∇f is constant on Ω, we infer from the last statement of Lemma 1.16 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 $\mu_{\mathbb{P}}(\Omega_1 \cap \Omega_2) = 1$ thanks to the absolute continuity of $\mu_{\mathbb{P}}$. $\square$

4.2 Displacement functionals for Wasserstein barycenters

Recall that the notion of Hessian plays a central role in differentiating optimal transport maps (Proposition 1.44). 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 [95], Lott and Villani [69, §7], Cordero-Erausquin et al. [31] and Villani [105, Chapter 14]. The function $J(t)$ defined below is actually $\mathrm{D}_x \exp(-\nabla t \, \phi)$ using (1.29). By convention, for a function f with variable $t \in \mathbb{R}$, we denote by $\dot{f}$ its derivative with respect to t.

Proposition 4.2. *Let (M, d_{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 1.23) exists. Then $t \in [0,1] \mapsto \gamma(t) = \exp(-t \nabla \phi)(x)$ is a minimal geodesic. Define*

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

Denote by $\Delta\phi(x)$ the trace of $\mathrm{Hess}_x \, \phi$ and by $\det J(t), 0 \le t \le 1$ 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

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

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

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

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

Proof. Since $\mathrm{Hess}_x \, \phi$ exists, $\gamma(1)$ is not in the cut-locus of x [30, Proposition 4.1] and thus γ is a minimal geodesic. Let $\{e_1, \ldots, e_m\} \subset T_x M$ be an orthonormal basis. Fix an index $1 \le i \le m$. Fix $\delta > 0$ such that the curve $t \in [0,1] \mapsto \exp_x(t \, \delta e_i)$ is a minimal geodesic. For $s \in (-\delta, \delta)$, define $y_s := \exp_x s \, e_i$ and consider the following family of geodesics (with parameter s)

$$\begin{aligned} \alpha : [0,1] \times (-\delta, \delta) &\to M \\ (t,s) &\mapsto \exp_{y_s} \left\{ -t \, \Pi_{x \to y_s} \left[\nabla \phi(x) + s \, \mathrm{Hess}_x \, \phi(e_i) \right] \right\}, \end{aligned}$$

where $\Pi_{x \to y_s} : T_x M \to T_{y_s} M$ (c.f. Lemma 1.43) is the parallel transport along the minimal geodesic $t \in [0,1] \mapsto \exp_x(t \, \delta e_i)$. Note that the variation field $\partial_s \alpha(t,0)$ of α satisfies the Jacobi equation along γ with initial condition $\partial^2_{t,s}\alpha(0,0) = e_i$ and $\partial^3_{t,t,s}\alpha(0,0) = -\operatorname{Hess}_x \phi(e_i)$ [66, Proposition 10.4].

We now compute the Jacobi field $\partial_s\alpha(t,0)$. If $\nabla\phi$ exists at y_s, then the infinitesimal characterization of $\operatorname{Hess}_x \phi$ [30, Definition 3.9] implies,

$$\nabla\phi(y_s) = \Pi_{x\to y_s}\left[\nabla\phi(x) + s\operatorname{Hess}_x \phi(e_i) + o(s)\right] \quad \text{as } s \to 0, \tag{4.2}$$

which is a non-smooth version of the relation (1.26) between Hessian and parallel transport. Fix a $t \in [0,1]$ and a sequence of real numbers $s_j \to 0$ with $|s_j| < \delta$ such that $\nabla\phi$ exists at $w_j := \exp_x(s_j e_i)$. By definition of α and (4.2), we have $\exp(-t\nabla\phi)(w_j) = \alpha(t,s_j) + o(s_j)$ as $j \to \infty$. Hence, using the normal coordinates around $\exp(-t\nabla\phi)(x)$, we can compute $\partial_s\alpha(t,0)$ as follows,

$$\partial_s\alpha(t,0) = \lim_{j\to\infty} \frac{\exp^{-1}_{\exp(-t\nabla\phi)(x)}[\exp(-t\nabla\phi)(w_j)]}{s_j - 0} = \mathrm{D}_x \exp(-\nabla t\,\phi)\cdot e_i = J(t)\cdot e_i,$$

where in the second equality we used the fact that $t\,\phi$ is a c-concave function [30, Lemma 5.1] and the infinitesimal justification [30, (b) of Proposition 4.1] of differentiating $\exp(-\nabla t\,\phi)$.

Since $\partial_t\alpha(t,0) = \dot\gamma(t)$, we obtain the following Jacobi equation derived from α,

$$\ddot J_i(t) + R(J_i(t),\dot\gamma(t))\cdot\dot\gamma(t) = 0, \quad J_i(0) = e_i,\ \dot J_i(0) = -\operatorname{Hess}_x \phi(e_i),$$

where $J_i(t) := J(t)\cdot e_i$ and R is the Riemannian curvature tensor on M.

Therefore, J satisfies a matrix form of Jacobi equation to which we can apply differential equation comparison theorems, and then conclude our proposition. The details are given in many references such as Villani [105, Theorem 14.8]. ☐

The following displacement functionals $f\,\mathrm{d}\operatorname{Vol} \in \mathcal{W}_2(M) \mapsto \int G(f)\,\mathrm{d}\operatorname{Vol}$ are inspired by the entropy functional, where $G(x) := x\log x$. To uniformly bound (from above) their values of the approximating sequence of barycenter measures to which the consistency of Wasserstein barycenters is applied, we add the assumption of bounded derivatives. Examples of G can be constructed according to Theorem 4.13.

Proposition 4.3 (Displacement functionals)**.** *Let (M, d_{g}) be an m-dimensional complete Riemannian 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 $\mu_{\mathbb{P}}$ the unique Wasserstein barycenter of $\mathbb{P} := \sum_{i=1}^n \lambda_i\,\delta_{\mu_i} \in (\mathcal{W}_2(\mathcal{W}_2(M)), d_{\mathbb{W}})$, which is absolutely continuous, and we denote by f its density function.*

Let G be a function on $[0,\infty)$ with $G(0) = 0$ such that 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,

$$\int_M G(f)\,\mathrm{d}\operatorname{Vol} \leq \sum_{i=1}^k \frac{\lambda_i}{\Lambda}\int_M G(g_i)\,\mathrm{d}\operatorname{Vol} + \frac{L_H K}{2\Lambda} d_{\mathbb{W}}(\mathbb{P},\delta_{\mu_{\mathbb{P}}})^2 + \frac{L_H}{2\Lambda}(m^2+2m), \tag{4.3}$$

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

Remark 4.4. The following example helps to understand (4.3). 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 $\operatorname{Ent}(f \cdot \mathrm{Vol}) := \int_M G(f)\,\mathrm{d}\,\mathrm{Vol}$. The inequality (4.3) becomes

$$\operatorname{Ent}(\mu_{\mathbb{P}}) \leq \lambda \operatorname{Ent}(\mu_1) + (1-\lambda)\operatorname{Ent}(\mu_2) + \frac{K}{2}\lambda(1-\lambda)d_W(\mu_1,\mu_2)^2 + \frac{m^2}{2} + m,$$

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 [96, §4,2] and [69, Definition 0.7].

Moreover, $L_H(m^2+2m)/(2\Lambda)$ is also the only additional term when we compare inequality (4.3) with the Wasserstein Jensen's inequality proven by Kim and Pass [58, Theorem 7.11], which corresponds to the case $k = n$. However, our inequality (4.3) for the case $k < n$ is crucial to the proof of our main result in the next section.

Proof of Proposition 4.3. For $1 \leq i \leq n$, let $F_i := \exp(-\nabla\phi_i)$ be the optimal transport map from $\mu_{\mathbb{P}}$ to μ_i with ϕ_i a c-concave function given by Theorem 1.27. According to Theorem 4.1 and Proposition 1.45, there exists a Borel set $\Omega \subset M$ with $\mu_{\mathbb{P}}(\Omega) = 1$ 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

$$\int_M G(g_i)\,\mathrm{d}\,\mathrm{Vol} = \int_{N_i} G\left(\frac{f}{\operatorname{Jac} F_i}\right)\operatorname{Jac} F_i\,\mathrm{d}\,\mathrm{Vol}, \quad 1 \leq i \leq k, \tag{4.4}$$

where $N_i \subset \Omega$ for $1 \leq i \leq k$ are Borel sets such that $\mu_{\mathbb{P}}(N_i) = 1$ 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 (4.4) that

$$\int_M G(g_i)\,\mathrm{d}\,\mathrm{Vol} = \int_{N_i} H(\log f + l_i)\,\mathrm{d}\,\mu_{\mathbb{P}}, \quad 1 \leq i \leq k. \tag{4.5}$$

Applying Proposition 4.2 to ϕ_i for $1 \leq i \leq k$, we have on Ω,

$$l_i \geq \Delta\phi_i - K\|\nabla\phi_i\|^2/2, \quad 1 \leq i \leq k. \tag{4.6}$$

For $x \in \Omega$ and $1 \leq i \leq n$, since $\operatorname{Hess}_x d^2_{F_i(x)}/2 - \operatorname{Hess}_x \phi_i$ is positive semi-definite (Proposition 1.44), we can also bound $\Delta\phi_i(x)$ from above using the upper bound of the Laplacian of distance functions, as observed by Kim and Pass [58, Lemmma 2.7]:

$$\begin{aligned}\Delta\phi_i(x) \leq \Delta d^2_{F_i(x)}/2 &\leq m\frac{\sqrt{K}d_{\mathrm{g}}(x,F_i(x))}{\tanh(\sqrt{K}d_{\mathrm{g}}(x,F_i(x)))}\\ &\leq m(1+\sqrt{K}d_{\mathrm{g}}(x,F_i(x))) \leq m + m^2/2 + K\,\|\nabla\phi_i(x)\|^2/2,\end{aligned} \tag{4.7}$$

where we used the general inequality $\alpha/\tanh\alpha \leq 1+\alpha$ for $\alpha \geq 0$ [1], applied the inequality of arithmetic and geometric means to $\sqrt{K\,d_{\mathrm{g}}(x,F_i(x))^2} \cdot \sqrt{m^2}$, and employed the equality $d_{\mathrm{g}}(x,F_i(x)) =$

[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 $\alpha \geq 0$. As $f(0) = 0$ and $f'(\alpha) = \sinh\alpha + \alpha(\cosh\alpha - \sinh\alpha) = \sinh\alpha + \alpha\,e^{-\alpha}$, we have $f'(\alpha) \geq 0$ and thus $f(\alpha) \geq f(0) = 0$.

$\|\nabla\phi_i(x)\|$ for $x \in \Omega$. With our assumptions on H, (4.6) and (4.7) imply that for $1 \le i \le k$, on the set $\cup_{i=1}^k N_i$ (where $\log f$ is well-defined),

$$\begin{aligned} H(\log f + l_i) - H(\log f) = H'(\xi)\, l_i &\ge H'(\xi)[\Delta\phi_i - K\|\nabla\phi_i\|^2/2] \\ &\ge H'(\xi)[\Delta\phi_i - K\|\nabla\phi_i\|^2/2 - m - m^2/2] \\ &\ge L_H(\Delta\phi_i - K\|\nabla\phi_i\|^2/2) - L_H(m + m^2/2), \end{aligned} \tag{4.8}$$

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 (4.8) with coefficients λ_i/Λ on the set $\cup_{i=1}^k N_i$,

$$\begin{aligned} H(\log f) &\le \sum_{i=1}^k \frac{\lambda_i}{\Lambda} H(\log f + l_i) - \frac{L_H}{\Lambda} \sum_{i=1}^k \lambda_i(\Delta\phi_i - K\|\nabla\phi_i\|^2/2) + L_H(m + m^2/2) \\ &= \sum_{i=1}^k \frac{\lambda_i}{\Lambda} H(\log f + l_i) + \frac{L_H}{\Lambda} \sum_{i>k}^n \lambda_i\, \Delta\phi_i + \frac{L_H K}{2\Lambda} \sum_{i=1}^k \lambda_i\, \|\nabla\phi_i\|^2 + L_H(m + m^2/2) \\ &\le \sum_{i=1}^k \frac{\lambda_i}{\Lambda} H(\log f + l_i) + \frac{L_H K}{2\Lambda} \sum_{i=1}^n \lambda_i\, \|\nabla\phi_i\|^2 + \frac{L_H}{2\Lambda}(m^2 + 2m), \end{aligned} \tag{4.9}$$

where we used $\sum_{i=1}^n \lambda_i\, \Delta\phi_i = 0$ derived from the Hessian equality for the first equality and used (4.7) for the last inequality. Finally, (4.3) follows from (4.5) after integrating (4.9) over $N_1 \cap \ldots \cap N_k$ against $\mu_{\mathbb{P}}$ since $\mu_{\mathbb{P}}(N_i) = 1$ for $1 \le i \le k$ and $d_W(\mu_{\mathbb{P}}, \mu_i)^2 = \int_M \|\nabla\phi_i\|^2 \,\mathrm{d}\,\mu_{\mathbb{P}}$ for $1 \le i \le n$. $\square$

4.3 Proof of absolute continuity

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

Theorem 4.5. *Let (M, d_{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 space equipped with a σ-finite Borel measure.

4.3.1 Preserving absolute continuity along approximating sequences

We first deduce an intermediate result by applying the consistency of Wasserstein barycenters to the displacement functionals introduced in Proposition 4.3.

The following lemma, taken from Santambrogio [90, Proposition 7.7, Remak 7.8], originates from Buttazzo and Freddi [24, Theorem 2.2], which was slightly generalized later in [6, Theorem 2.34]. One can find another slightly generalized version by Ambrosio et al. [5, Theorem 15.8, Theorem 15.9] with a proof for the case of Euclidean spaces.

Lemma 4.6. *Let E be a Polish space equipped with a σ-finite Borel measure μ. Let G be a function on $[0, \infty)$ such that*

1. *$G(x) \ge 0$ with $G(0) = 0$;*

2. *G is continuous and convex;*

3. $\lim_{x\to\infty} G(x)/x = \infty$.

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 $\liminf_{i\to\infty} \int_E G(f_i)\,\mathrm{d}\,\mu$ is finite, then ν is also absolutely continuous and

$$\int_E G(f)\,\mathrm{d}\,\mu \leq \liminf_{i\to\infty} \int_E G(f_i)\,\mathrm{d}\,\mu < \infty, \tag{4.10}$$

where f is the density of ν.

Since convergence in Wasserstein metric implies weak convergence (Proposition 1.25), Lemma 4.6 ensures that the set below is closed in $\mathcal{W}_2(E)$.

Definition 4.7 ($\mathsf{B}(G, L)$ sets)**.** Let E be a Polish space equipped with a σ-finite Borel measure μ. Let G be a function on $[0, \infty)$ such that

1. $G(x) \geq 0$ with $G(0) = 0$;
2. G is non-decreasing, continuous, and convex;
3. $\lim_{x\to\infty} G(x)/x = \infty$;
4. the function $H(x) := G(e^x)/e^x$ has continuous, non-negative, and bounded derivative.

Given a positive number $L > 0$, we refer to the following subset of $\mathcal{W}_2(E)$ as $\mathsf{B}(G, L)$,

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

which is a closed subset of $\mathcal{W}_2(E)$ thanks to Lemma 4.6.

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 $\mathsf{B}(G, L)$ can be approximated by elements in $\mathsf{B}(G, L + 1)$ with compact support, as shown in the following lemma.

Lemma 4.8. *Let (E, d) be a proper metric space equipped with a σ-finite Borel measure μ. Fix a $\mathsf{B}(G, L)$ set as defined in Definition 4.7. For any probability measure $\nu \in \mathsf{B}(G, L)$, there exists a sequence of probability measures in $\mathsf{B}(G, L + 1)$ 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 $(k, x) \in \mathbb{N}^* \times E$,

$$g(k, x) := f(x)\mathbb{1}_{\overline{B}(x_0,k)}(x) + \alpha_k \,\mathbb{1}_{Y\cap\overline{B}(x_0,k)}(x), \tag{4.11}$$

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, 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 d_W using the characterization Proposition 1.25. For a continuous function $\phi : E \to \mathbb{R}$ such that $|\phi(x)| \leq 1 + d(x_0, x)^2$, note that

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

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 $\lim_{n\to\infty} d_W(\nu_n, \nu) = 0$ according to Proposition 1.25.

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

$$\forall\, (n, x) \in \mathbb{N}^* \times E, \quad G(g(k_0 + n, x)) \leq G(f(x)) + G(l + \alpha_{k_0})\, \mathbb{1}_Y(x). \tag{4.12}$$

Since G is a continuous function and $\mu(Y) < 0$, we can apply the dominated convergence theorem to (4.12) 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, $\nu_n \in \mathsf{B}(G, L + 1)$, which concludes the proof. □

As the assumptions in Definition 4.7 include the ones we used to construct displacement functionals in Proposition 4.3, we obtain the following intermediate result.

Proposition 4.9. *Let $(M, d_{\mathfrak{g}})$ 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 $\mathsf{B}(G, L)$ defined in Definition 4.7 with respect to the volume measure on M, then the unique barycenter of $\mathbb{P}$ is absolutely continuous.*

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 $d_{\mathbb{W}}$ 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 $d_{\mathbb{W}}$ 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 [105, 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 $\{\mathbb{P}^1_j\}_{j\geq 1}$ and $\{\mathbb{P}^2_j\}_{j\geq 1}$ satisfying $d_{\mathbb{W}}(\mathbb{P}^1_j, \mathbb{P}^1) \to 0$, $d_{\mathbb{W}}(\mathbb{P}^2_j, \mathbb{P}^2) \to 0$ when $j \to \infty$. Furthermore, thanks to Lemma 4.8, we can further refine the two approximating sequences to ensure that all $\mathbb{P}^1_j, \mathbb{P}^2_j$ for $j \geq 1$ are supported in probability measures with compact support and $\mathbb{P}^1_j(\mathsf{B}(G, L + 1)) = 1$. Define $\mathbb{P}_j := \mathbb{P}(\mathsf{B}(G, L))\, \mathbb{P}^1_j + (1 - \mathbb{P}(\mathsf{B}(G, L))\, \mathbb{P}^2_j$. It follows that $d_{\mathbb{W}}(\mathbb{P}_j, \mathbb{P}) \to 0$ as $j \to \infty$.

Consider the displacement functional $\mathcal{G} : f \cdot \mathrm{Vol} \mapsto \int_M G(f) \,\mathrm{d}\,\mathrm{Vol}$. Proposition 4.3 implies the following estimate of $\mathcal{G}(\mu_{\mathbb{P}_j})$ at the barycenter $\mu_{\mathbb{P}_j}$ of $\mathbb{P}_j$,

$$\mathcal{G}(\mu_{\mathbb{P}_j}) \leq \int_{\mathcal{W}_2(M)} \mathcal{G}(\nu) \,\mathrm{d}\,\mathbb{P}_j^1(\nu) + \frac{L_H K}{2\Lambda} d_{\mathbb{W}}(\mathbb{P}_j, \delta_{\mu_{\mathbb{P}_j}})^2 + \frac{L_H}{2\Lambda}(m^2 + 2m), \tag{4.13}$$

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 $\mathbb{P}$, the consistency of Wasserstein barycenters (Theorem 2.13) implies that $d_W(\mu_{\mathbb{P}_j}, \mu_{\mathbb{P}}) \to 0$ and thus $d_{\mathbb{W}}(\mathbb{P}_j, \delta_{\mu_{\mathbb{P}_j}}) \to d_{\mathbb{W}}(\mathbb{P}, \delta_{\mu_{\mathbb{P}}})$ as $j \to \infty$. Since the support of $\mathbb{P}_j^1$ is a subset of $\mathsf{B}(G, L+1)$ and $d_{\mathbb{W}}(\mathbb{P}_j, \delta_{\mu_{\mathbb{P}_j}})$ is bounded for $j \geq 1$, by setting

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

we have $\mu_{\mathbb{P}_j} \in \mathsf{B}(G, L')$ for all $j \geq 1$. It follows from Lemma 4.6 that $\mu_{\mathbb{P}}$ is absolutely continuous. $\square$

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

4.3.2 Compactness via Souslin space theory

The last step towards our main result is to show that the closed subset $\mathsf{B}(G, L)$ needed in Proposition 4.9 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 [17, 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 4.7 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 4.10 (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,

$$\mathbb{A} := \left\{ f \in L^1(\mu) \,\middle|\, f \geq 0, \int_E f \,\mathrm{d}\,\mu = 1, \int_E d(x_0, x)^2 f(x) \,\mathrm{d}\,\mu(x) < \infty \right\}, \tag{4.14}$$

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 $L^\infty(\mu)$ [17, 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 2.17, $\mathbb{A}$ 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 F in $(\mathbb{A}, \tau)$ with $\mathbb{P}(\mathsf{F}) > 0$. If we can accomplish this, then F forms a family of uniformly integrable functions by the Dunford-Pettis theorem (Proposition 4.12), bringing us closer to the main result. To find such an 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 employ some well-known results from the Souslin space theory, which can fill the gap in the previous argument with Radon measures.

Lemma 4.11. *Let (E, d) be a Polish space with an outer regular and σ-finite Borel measure μ on E. Let $\mathbb{A}$ be as in (4.14). 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 [79, Lemma 2.4.2], and we recall its arguments here. By Lemma 2.17, $\mathbb{A}$ is a Borel set for both τ_w and τ_W. Since $(\mathcal{W}_2(E), d_W)$ is a Polish space, $(\mathbb{A}, \tau_W)$ is then a Souslin space as a Borel subset of $(\mathcal{W}_2(E), d_W)$ [17, Theorem 6.6.7]. Consider the identity map $\mathrm{Id} : (\mathbb{A}, \tau_W) \to (\mathbb{A}, \tau_w)$, it is continuous and bijective. According to definition 1.8, $(\mathbb{A}, \tau_w)$ 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 [17, Theorem 6.7.3].

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 [17, Theorem 4.1.3]. Its separability is asserted in Brézis [21, Theorem 4.13] and Bogachev [17, 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 $\mathbb{R}^\infty$ [17, 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 for 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 $L^1(\mu)$ [17, Theorem 6.6.7].

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 [17, Theorem 6.7.3].

$\mathbb{P}$, 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 [17, Theorem 7.4.3]. Hence, $\mathbb{P}(\mathbb{A}) > 0$ can be approximated by the $\mathbb{P}$-measure of compact subsets of $(\mathbb{A}, \tau)$. $\square$

We prove the following slightly generalized Dunford-Pettis theorem that connects uniform integrability and the weak topology τ.

Proposition 4.12 (Dunford-Pettis theorem)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space with a σ-finite measure μ on it. Let $\mathsf{F} \subset L^1(\mu)$ be a set of μ-integrable functions. If F has compact closure in the weak topology induced by the dual space $L^\infty(\mu)$ of $L^1(\mu)$, then F is* uniformly integrable*, i.e.,*

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

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 [17, Theorem 4.4.1] and [87, Exercise 6.12]. The above definition of uniform integrability is taken from Bogachev [17, 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 [17, 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 F has compact closure in the weak topology, but is not uniformly integrable. Then, there are $\epsilon > 0$ and a sequence $\{f_n\}_{n\geq 1} \subset \mathsf{F}$ such that

$$\inf_{n\geq 1} \int_{\{|f_n|>n\}} |f_n| \,\mathrm{d}\,\mu \geq \epsilon. \tag{4.15}$$

Applying the Eberlein–Šmulian theorem (Theorem 1.9) to $\{f_n\}$ and the Banach space $L^1(\mu)$ [21, 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 $A \in \mathcal{F}$ we have

$$\lim_{k\to\infty} \int_A f_{n_k} \,\mathrm{d}\,\mu = \int_A f \,\mathrm{d}\,\mu. \tag{4.16}$$

It follows from the Vitali–Hahn–Saks theorem (Corollary 1.13) 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

$$\mu(A) < \delta \implies \sup_{k\geq 1} \int_A |f_{n_k}| \,\mathrm{d}\,\mu < \epsilon. \tag{4.17}$$

Via the isometric embedding of $L^1(\mu)$ into the dual space of $L^\infty(\mu)$ [21, Corollary 1.4], the Banach–Steinhaus theorem (Theorem 1.10) 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 (4.17) for the ϵ in (4.15), 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 (4.15) and (4.17). ☐

We also generalize the de la Vallée Poussin criterion to construct the function G in Definition 4.7. In the following proposition, the σ-finiteness of μ allows us to apply Fubini's theorem.

Theorem 4.13 (De la Vallée Poussin criterion)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space with a σ-finite measure μ on it. A subset $\mathsf{F} \subset L^1(\mu)$ is uniformly integrable, i.e.,*

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

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

1. *$G(x) = 0$ for $0 \le x \le 1$;*
2. *G is a non-decreasing and convex function that is smooth on $(0, +\infty)$;*
3. *$\sup_{f\in\mathsf{F}} \int_\Omega G(|f|) \,\mathrm{d}\,\mu \le 1$;*
4. *if we define the function $H(x) := G(e^x)e^{-x}$ on $\mathbb{R}$, then $\lim\limits_{x\to+\infty} H(x) = +\infty$, and its derivative H' is smooth with $0 \le H'(x) \le 1$.*

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

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

which shows that F is uniformly integrable.

Now assume that we are given a uniformly integrable subset $\mathsf{F} \subset L^1(\mu)$. To better motivate our construction of G, we postpone the definition of a smooth function H with $H(x) = 0, x \le 0$ to (4.21) 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 $0 \le x \le 1$. Hence, we have $G(x) = \int_0^x \int_0^s G''(t) \,\mathrm{d}\,t \,\mathrm{d}\,s$ for $x > 0$ and thus

$$\begin{aligned}
\int_\Omega G(|f|) \,\mathrm{d}\,\mu &= \int_\Omega \int_0^{|f|} \int_0^s G''(t) \,\mathrm{d}\,t \,\mathrm{d}\,s \,\mathrm{d}\,\mu = \int_\Omega \int_\mathbb{R} \int_\mathbb{R} G''(t) \cdot \mathbb{1}_{0<t<s<|f|} \,\mathrm{d}\,t \,\mathrm{d}\,s \,\mathrm{d}\,\mu \\
&= \int_\mathbb{R} \int_\mathbb{R} G''(t) \cdot \mathbb{1}_{0<t<s} \cdot \mu(|f| > s) \,\mathrm{d}\,t \,\mathrm{d}\,s \\
&= \int_\mathbb{R} G''(t) \cdot \mathbb{1}_{t>0} \int_t^\infty \mu(|f| > s) \,\mathrm{d}\,s \,\mathrm{d}\,t \\
&= \int_0^\infty \frac{H'(\log t) + H''(\log t)}{t} \int_t^\infty \mu(|f| > s) \,\mathrm{d}\,s \,\mathrm{d}\,t \\
&= \int_\mathbb{R} [H'(y) + H''(y)] \int_{e^y}^\infty \mu(|f| > s) \,\mathrm{d}\,s \,\mathrm{d}\,y, \qquad (4.18)
\end{aligned}$$

where we applied Fubini's theorem twice and a change of variables $y := \log t$. According to (4.18), 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

$$\begin{aligned}
\int_{\{|f|>t\}} |f| \,\mathrm{d}\,\mu &= \int_{\{|f|>t\}} \int_\mathbb{R} \mathbb{1}_{0<s<|f|} \,\mathrm{d}\,s \,\mathrm{d}\,\mu = \int_\mathbb{R} \int_E \mathbb{1}_{|f|>t} \cdot \mathbb{1}_{0<s<|f|} \,\mathrm{d}\,\mu \,\mathrm{d}\,s \\
&= \int_\mathbb{R} \int_\Omega \mathbb{1}_{0<s<t<|f|} + \mathbb{1}_{0<t\le s<|f|} \,\mathrm{d}\,\mu \,\mathrm{d}\,s = t\,\mu(|f| > t) + \int_t^\infty \mu(|f| > s) \,\mathrm{d}\,s. \qquad (4.19)
\end{aligned}$$

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

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

where we used (4.19) for the first inequality and the uniform integrability of F for the second one. It follows that

$$\sup_{f\in\mathsf{F}} \sum_{n\geq 0} \int_{e^{\alpha(n)}}^{\infty} \mu(|f| > s) \,\mathrm{d}\, s \leq 1. \tag{4.20}$$

For the term $H' + H''$ in (4.18), 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 (4.20). 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

$$H(x) := \begin{cases} \int_0^x e^{-s} \int_0^s \gamma(t) e^t \,\mathrm{d}\, t \,\mathrm{d}\, s, & x > 0 \\ 0, & x \leq 0 \end{cases}. \tag{4.21}$$

In this way, we have $H''(x) + H'(x) = \gamma(x)$. Using this construction, (4.18) and (4.20) imply that

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

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 $0 \leq x \leq 1$. □

4.3.3 Final step of the proof

To prove Theorem 4.5, it remains to combine the previous auxiliary propositions to replace the assumption in Proposition 4.9 that $\mathbb{P}(\mathsf{B}(G, L)) > 0$ for some set $\mathsf{B}(G, L)$ (Definition 4.7).

As in Definition 4.10, we denote by $\mathbb{A}$ the set of absolutely continuous measures in $\mathcal{W}_2(M)$. If $\mathbb{P}(\mathbb{A}) > 0$, then Lemma 4.11 provides a compact subset F of $(\mathbb{A}, \tau)$ such that $\mathbb{P}(\mathsf{F}) > 0$. Applying the Dunford-Pettis theorem (Proposition 4.12) to F with $\mu := \mathrm{Vol}$, we see that F is uniformly integrable. Then the de la Vallée Poussin criterion (Theorem 4.13) asserts the existence of a smooth function G such that $\mathsf{F} \subset \mathsf{B}(G, 1) \subset \mathbb{A}$. Therefore, our theorem follows from Proposition 4.9 and the property $\mathbb{P}(\mathsf{B}(G, 1)) \geq \mathbb{P}(\mathsf{F}) > 0$.

Chapter 5

Restriction property of Wasserstein barycenters

Our main goal in this chapter is to generalize the divide-and-conquer technique used in the proofs of Proposition 3.4 and Theorem 3.7. This generalization enables us to construct a new probability measure such that one of its barycenters is a (normalized) restriction of a given Wasserstein barycenter. Consequently, we can study local properties of Wasserstein barycenters and deduce global properties via local restrictions.

In our restriction technique, we avoid operating on multi-marginal optimal transport plans (as they are not defined for general measures $\mathbb{P}$), and instead construct a push-forward map. This map modifies each element ν in the support of $\mathbb{P}$ by restricting the optimal transport plans between a fixed barycenter of $\mathbb{P}$ and ν. We begin with a technical lemma that establishes the measurability of these modifications, which is crucial for dividing couplings of two measures (one of which is fixed) according to a given bounded measurable function.

Lemma 5.1. *Let $(E, d_1), (F, d_2)$ be two Polish spaces. Consider their product space $E \times F$ endowed with the product metric $d((x_1, y_1), (x_2, y_2))^2 := d_1(x_1, x_2)^2 + d_2(x_2, y_2)^2$ for $x_1, x_2 \in E$ and $y_1, y_2 \in F$. Fix a measure $\mu \in \mathcal{W}_2(E)$, denote by Γ_μ the subset of measures in $\mathcal{W}_2(E \times F)$ whose first marginal measure is μ. Given a non-negative bounded measurable function g on E such that $g \cdot \mu \in \mathcal{W}_2(E)$, the following map*

$$\mathcal{G} : \Pi \in \Gamma_\mu \mapsto g \cdot \Pi \in \mathcal{W}_2(E \times F)$$

is continuous with respect to the Wasserstein metric d_W of $\mathcal{W}_2(E \times F)$, where $g \cdot \Pi$ stands for the measure $g(x) \cdot \Pi(\mathrm{d}\, x, \mathrm{d}\, y)$ on $E \times F$.

Proof. Let σ be an optimal transport plan between $\Pi_1, \Pi_2 \in \Gamma_\mu$. Then the following probability measure

$$g(x_1)g(x_2) \cdot \sigma(\mathrm{d}\, x_1, \mathrm{d}\, y_1, \mathrm{d}\, x_2, \mathrm{d}\, y_2) \in \mathcal{W}_2(E \times F \times E \times F)$$

is a coupling between $g \cdot \Pi_1$ and $g \cdot \Pi_2$. Hence,

$$d_W(\mathcal{G}(\Pi_1), \mathcal{G}(\Pi_2))^2 \leq \|g\|_\infty^2 d_W(\Pi_1, \Pi_2)^2,$$

where $\|g\|_\infty$ denotes the L^∞-norm of g with respect to μ. It follows that $\mathcal{G}$ is a continuous map with respect to the Wasserstein metric. ∎

The following proposition details our method of constructing new Wasserstein barycenters by restricting an existing one.

Proposition 5.2 (Restriction property of Wasserstein barycenters)**.** *Let (E,d) be a proper metric space and let $\mu \in \mathcal{W}_2(E)$ be a probability measure on E. Then there exists a measurable function $\Pi : \nu \mapsto \Pi_\nu$ from $\mathcal{W}_2(E)$ to $\mathcal{W}_2(E \times E)$ such that Π_ν is an optimal transport plan between μ and ν, where the metric on $E \times E$ is given by $d((x_1,y_1),(x_2,y_2))^2 = d(x_1,x_2)^2 + d(y_1,y_2)^2$ for $x_1,x_2,y_1,y_2 \in E$. If one can write $\mu = \lambda\,\mu^1 + (1-\lambda)\mu^2$ with $\mu^i \in \mathcal{W}_2(E)$ for $i = 1,2$ and some fixed positive number $0 < \lambda < 1$, then for any $\nu \in \mathcal{W}_2(E)$, we can write $\Pi_\nu = \lambda\,\Pi^1_\nu + (1-\lambda)\Pi^2_\nu$ and $\nu = \lambda\,\nu^1 + (1-\lambda)\nu^2$ such that Π^i_ν is an optimal transport plan between μ^i and $\nu^i \in \mathcal{W}_2(E)$, and the map $F^i : \mathcal{W}_2(E) \to \mathcal{W}_2(E)$ that sends ν to ν^i is continuous.*

Moreover, in the above rewriting, if μ is a Wasserstein barycenter of $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(E))$, then μ^i is a Wasserstein barycenter of $\mathbb{Q}^i := {F^i}_\#\mathbb{P}$.

Proof. The existence of a measurable selection $\nu \mapsto \Pi_\nu$ of optimal transport plans is proven in [105, Corollary 5.22]. By definition of the metric on $E \times E$, it follows from $\nu,\mu \in \mathcal{W}_2(E)$ that $\Pi_\nu \in \mathcal{W}_2(E \times E)$.

Since $\mu = \lambda\,\mu^1 + (1-\lambda)\mu^2$ and $0 < \lambda < 1$, measures μ^1,μ^2 are absolutely continuous with respect to μ. For $i = 1,2$, denote by g^i the density function of μ^i with respect to μ, i.e., $\mu^i = g^i \cdot \mu$. In particular, we have $\lambda\,g^1 + (1-\lambda)g^2 = 1$ for μ-almost everywhere. Define $\Pi^i_\nu := g^i \cdot \Pi_\nu$ for $\nu \in \mathcal{W}_2(E)$, where $g^i \cdot \Pi_\nu$ stands for the measure $g^i(x) \cdot \Pi_\nu(\mathrm{d}\,x, \mathrm{d}\,y)$ on $E \times E$ as in Lemma 5.1. Since $\lambda\,\Pi^1_\nu + (1-\lambda)\Pi^2_\nu = [\lambda\,g^1 + (1-\lambda)g^2]\Pi_\nu = \Pi_\nu \in \mathcal{W}_2(E \times E)$, we have $\Pi^1_\nu, \Pi^2_\nu \in \mathcal{W}_2(E \times E)$. For $i = 1,2$ and $\nu \in \mathcal{W}_2(E)$, define ν^i as the second marginal of Π^i_ν, which belongs to $\mathcal{W}_2(E)$ since $\mu^i \in \mathcal{W}_2(E)$ and $\Pi^i_\nu \in \mathcal{W}_2(E \times E)$. It follows from the restriction property of optimal transport plans [105, Theorem 4.6] that Π^i_ν is an optimal transport plan between μ^i and ν^i. Note that by definition, $\nu^i := \pi^2_\#[g^i \cdot \Pi_\nu]$, where $\pi^2 : E \times E \to E$ is the projection map sending $(x,y) \in E \times E$ to $y \in E$. We now show that the push-forward map $\pi^2_\# : \mathcal{W}_2(E \times E) \to \mathcal{W}_2(E)$ is continuous. Given $\Pi_1, \Pi_1 \in \mathcal{W}_2(E \times E)$, if $\sigma \in \mathcal{W}_2(E^2 \times E^2)$ is an optimal transport plan between Π_1 and Π_2, then

$$
\begin{aligned}
d_W(\Pi_1,\Pi_2)^2 &= \int_{E^2 \times E^2} d(x_1,x_2)^2 + d(y_1,y_2)^2 \,\mathrm{d}\,\sigma(x_1,y_1,x_2,y_2) \\
&\geq \int_{E \times E} d(y_1,y_2)^2 \,\mathrm{d}\,[\pi^2 \times \pi^2]_\#\sigma\,(y_1,y_2) \geq d_W(\pi^2_\#\Pi_1, \pi^2_\#\Pi_2)^2,
\end{aligned}
$$

which implies the continuity of the push-forward map $\pi^2_\#$. It follows from Lemma 5.1 that $F^i : \nu \mapsto \nu^i$ is a continuous map from $\mathcal{W}_2(E)$ to $\mathcal{W}_2(E)$.

Now we assume that $\mu = \mu_{\mathbb{P}}$ is a barycenter of $\mathbb{P}$. Observe that

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

where we used the fact that Π^i_ν is an optimal transport plan between μ^i and ν^i. It follows that $\mathbb{Q}^i \in \mathcal{W}_2(\mathcal{W}_2(E))$ for $i = 1, 2$. We claim that the last inequality above must be an equality. Consider the measure $\overline{\mu} := \lambda\, \mu_{\mathbb{Q}^1} + (1-\lambda)\mu_{\mathbb{Q}^2}$ with $\mu_{\mathbb{Q}^i}$ being a Wasserstein barycenter of $\mathbb{Q}^i$. According to the decomposition $\nu = \lambda\, \nu^1 + (1-\lambda)\nu^2$, if Π_i $(i = 1, 2)$ is a transport plan between $\mu_{\mathbb{Q}^i}$ and ν^i, then $\lambda\, \Pi_1 + (1-\lambda)\Pi_2$ is a transport plan between $\overline{\mu}$ and ν. Hence, it follows from the definitions of $\mu_{\mathbb{Q}^i}$, $\mathbb{Q}^i$, ν^i, and $d_W(\overline{\mu}, \nu)^2$ that

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

Hence, if our claim is false, then the expression $\int_{\mathcal{W}_2(E)} d_W(\cdot, \nu)^2 \,\mathrm{d}\,\mathbb{P}(\nu)$ admits strictly smaller value on the measure $\overline{\mu}$ than on the barycenter measure μ. Our claim is thus proven by contradiction, which implies the last assertion in the lemma. ☐

Remark 5.3. The proof of Proposition 5.2 is structurally analogous to the proof for the restriction of optimal transport plans in [105, Theorem 4.6]. Both arguments proceed by contradiction: assuming the property fails allows for the construction of a new candidate solution, which violates the presumed optimality of the original. This analogy becomes an identity in the special case of the barycenter problem over two measures, $\nu_0, \nu_1 \in \mathcal{W}_2(E)$. Recall that the McCann interpolation $\{\nu_\theta\}_{0 \leq \theta \leq 1}$ between ν_1 and ν_2 is made of barycenters of the measures $\mathbb{P}_\theta = \theta\, \delta_{\nu_1} + (1-\theta)\delta_{\nu_0}$. This interpolation is closely related to the dynamical optimal coupling in [105, Definition 7.20]. Consequently, a decomposition of the dynamical optimal coupling corresponds to decomposing a series of barycenters $\mu_{\mathbb{P}_\theta} = \nu_\theta$.

In the thesis, a key property used frequently in conjunction with Proposition 5.2 is that the map F^i sends a probability measure ν to the measure ν^i that is absolutely continuous with respect to ν. The following corollary presents two direct consequences of this property.

Corollary 5.4. *Let (E, d) be a proper metric space. Fix a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(E))$ with a barycenter $\mu_\mathbb{P} \in \mathcal{W}_2(E)$. Given an equality $\mu_\mathbb{P} = \lambda\, \mu^1 + (1-\lambda)\mu^2$ with $\mu^i \in \mathcal{W}_2(E)$ for $i = 1, 2$ and $\lambda \in (0, 1)$, there exist two probability measures $\mathbb{Q}^1, \mathbb{Q}^2$ such that μ^i is a barycenter of $\mathbb{Q}^i$ for $i = 1, 2$ and the following property holds. For any measure η on E,*

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

Proof. Proposition 5.2 provides two continuous maps $F^1, F^2 : \mathcal{W}_2(E) \to \mathcal{W}_2(E)$ such that $\lambda\, F^1(\nu) + (1-\lambda)F^2(\nu) = \nu$ for $\nu \in \mathcal{W}_2(E)$ and μ^i $(i = 1, 2)$ is a barycenter of $\mathbb{Q}^i := F^i{}_\# \mathbb{P}$. It follows that probability measures $F^1(\nu), F^2(\nu)$ are absolutely continuous with respect to ν. Hence, given a measure η on E, if ν is absolutely continuous with respect it, then so are the measures $F^1(\nu)$ and $F^2(\nu)$. The two assertions in the corollary follows directly from the definitions of $\mathbb{Q}^1$ and $\mathbb{Q}^2$ via the maps F^1, F^2. $\square$

Chapter 6

Wasserstein barycenters on metric trees

This chapter investigates the regularity of Wasserstein barycenters in the setting of metric trees, with a particular focus on characterizing their potential singularities. While the study of absolutely continuous Wasserstein barycenters on Riemannian manifolds has seen significant progress, the nature of singular barycenters remains less understood. The geometric complexity of general manifolds motivates our shift to metric trees—a simpler, yet non-trivial class of geodesic spaces that exhibit rich phenomena. Here, we introduce a novel reduction technique that provides a systematic approach to characterizing singular Wasserstein barycenters on trees by leveraging the well-developed theory on the real line.

We begin with a canonical example that illustrates the singularity (see Example 6.25, inspired by [50]). Consider the tripod in Figure 6.1, formed by three copies of the unit interval $[0,1]$ joined at a common origin. Let $\mathbb{P} := \frac{1}{3}\sum_{i=1}^{3}\delta_{\nu_i}$ be a probability measure on the space of probability

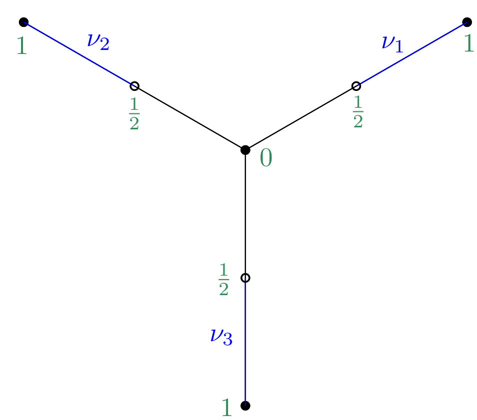


Figure 6.1: $\mathbb{P} = \sum_{i=1}^{3}\frac{1}{3}\delta_{\nu_i}$ on the tripod

measures, where each ν_i is an absolutely continuous measure supported on the outer half $[\frac{1}{2},1]$ of a distinct branch. The unique Wasserstein barycenter of $\mathbb{P}$ is the Dirac measure $\mu_{\mathbb{P}} = \delta_0$ at the central vertex. This starkly demonstrates how a collection of regular measures can collapse into a purely singular barycenter, motivating a deeper investigation into the mechanisms governing such behavior.

Our setting of a metric tree, a metric graph without cycles, treats edges as continuous intervals with prescribed lengths, making it a geodesic space. Consistent with our framework of optimal transport (Section 1.4) in this thesis, we set the squared distance function as the cost function and thus consider the 2-Wasserstein space. It is important to distinguish our work from the related literature. For instance, research on the "tree metric" or "tree-Wasserstein distance" typically considers only the vertices of a tree and benefits from a closed-form expression for the 1-Wasserstein distance [81, 61, 70, 85]. Similarly, ramified optimal transport studies transport problems between finitely support measures with branching cost structures [108, 109, 110]. While our metric graph setting aligns with that of [71], their work focuses on the 1-Wasserstein distance. To our knowledge, only a few works, such as [15, 34], have studied optimal transport with the squared distance cost on metric graphs. This highlights that, despite its apparent simplicity, the 2-Wasserstein space on a metric tree remains largely unexplored compared to its counterparts on the real line or on Riemannian manifolds.

The cornerstone of our analysis is a reduction technique introduced in Section 6.2. For any oriented edge $\vec{e} := \overrightarrow{\{v_0, v_1\}}$ of a tree Γ, we define a reduction map $T^{\vec{e}} : \Gamma \to \mathbb{R}$. As illustrated in

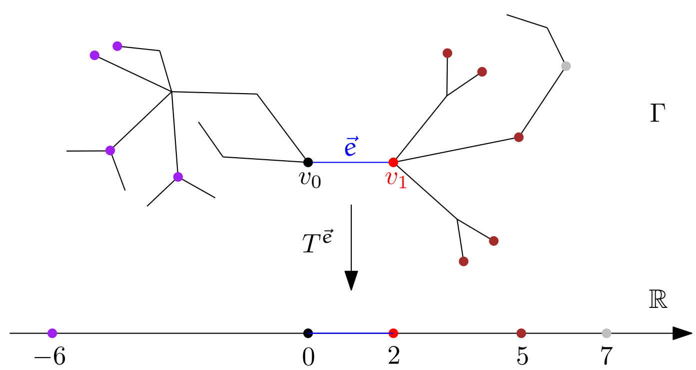


Figure 6.2: Illustrative example of the reduction map $T^{\vec{e}}$.

Figure 6.2, this map effectively "flattens" the tree into the real line by identifying the edge $\vec{e}$ with an interval and isometrically embedding the rest of the tree relative to $\vec{e}$. Our key technical result, Theorem 6.22, states that if the support of a measure $\mu \in \mathcal{W}_2(\Gamma)$ is contained within the edge $\vec{e}$, the Wasserstein distance between μ and any other measure $\nu \in \mathcal{W}_2(\Gamma)$ is preserved under this map: $d_W(\mu, \nu) = d_W(T^{\vec{e}}{}_\# \mu, T^{\vec{e}}{}_\# \nu)$. This powerful result allows us to transform certain optimal transport problems on a tree into equivalent, and more tractable, problems on the real line.

By combining this reduction technique with the restriction property of Wasserstein barycenters (Chapter 5), we develop a unified framework for extending results from $\mathbb{R}$ to metric trees. The power of this approach is demonstrated by our proof of the almost absolute continuity of Wasserstein barycenters on trees (Theorem 6.28). This theorem asserts that if $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$ gives positive mass to the set of measures that are absolutely continuous (with respect to the one-dimensional Hausdorff measure $\mathcal{H}$), then any barycenter $\mu_{\mathbb{P}}$ must be absolutely continuous everywhere except, possibly, at the vertices of the tree. In other words, singularities of the barycenter are confined to the vertex set V.

The chapter is structured as follows. We begin in Section 6.1 by formally defining metric graphs and establishing their properties as proper geodesic spaces. Our core analytical tools, the reduction technique, is introduced in Section 6.2. To build the necessary foundation, Section 6.4 explores barycenters on $\mathbb{R}$, introducing concepts like the dual measure and the rigid property to characterize singularity. For instance, we show that if the barycenter of $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$ is singular, then $\mathbb{P}$-almost every measure must also be singular (Theorem 6.51). Armed with these tools and insights, we return to metric trees. In Section 6.5, we apply our framework to rigorously characterize barycenter singularities at vertices, motivated by the almost absolute continuity theorem. Finally, Section 6.6 synthesizes our approach through several detailed examples, illustrating the unique and sometimes counter-intuitive behavior of Wasserstein barycenters on metric trees.

6.1 Definitions and preliminary properties

6.1.1 Metric (measure) graphs

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

Recall that a (undirected, simple and non-trivial) *graph* is an ordered pair $G := (\mathcal{V}, \mathcal{E})$ consisting of a non-empty set of vertices $\mathcal{V}$ and a non-empty set of edges $\mathcal{E} \subset \{\{x, y\} \mid x, y \in \mathcal{V} \text{ and } x \neq y\}$, which are unordered pairs of vertices. Note that even though our definition excludes more general graphs containing loops or parallel edges, it is not an essential restriction since we can turn them into graphs by adding vertices so that our propositions in this chapter are applicable.

We also introduce the following definitions for graphs [18, §1.1, §4.1] [32, §1.1, §1.3]. A vertex $x \in \mathcal{V}$ is *incident* with an edge $\alpha \in \mathcal{E}$ if $x \in \alpha$, in which case we also say that α is an edge at x. The two (distinct) vertices incident with an edge are its *end*s, and an edge *join*s its ends. The *degree* of a vertex is the number of edges at the vertex. Graphs are called *finite*, *infinite* or *countable* according to the number of their vertices. A *path* (respectively a *cycle*) is a finite graph whose vertices can be arranged in a linear (respectively *cyclic*) sequence, in such a way that two vertices are joined by an edge if and only if they are consecutive in the sequence. We denote a path p by $p = x_0 x_1 \dots x_k$ if $\{x_0, x_1, \dots, x_k\}$ is the set of its vertices and $\{\{x_0, x_1\}, \{x_1, x_2\}, \dots, \{x_{k-1}, x_k\}\}$ is the set of its edges, where $x_0, x_1, \dots, x_k$ are $k+1$ distinct elements, i.e., $x_i \neq x_j$ if $i \neq j$. Moreover, we say that p is a path *from* x_0 *to* x_k (as well as *between* x_0 *and* x_k). For two graphs $G = (\mathcal{V}, \mathcal{E})$ and $G' = (\mathcal{V}', \mathcal{E}')$, G' is a sub-graph of G if $\mathcal{V}' \subset \mathcal{V}$ and $\mathcal{E}' \subset \mathcal{E}$. When the sub-graph G' is a path (respectively a cycle), we also say that G' is a path (respectively a cycle) of G. A graph is *connected* if for every two different vertices, there is a path between them. A *tree* is a connected graph without cycles.

A graph is *locally finite* if the degree of each vertex is finite. It is known that a connected, locally finite and infinite graph is countable [107, Theorem 1.4]. We shall construct metric graphs from connected locally finite graphs via length functions defined on edges. For convenience, we assume in the following construction that the set of vertices is a subset of $\mathbb{N}$, which indeed imposes a global orientation of graphs.

Definition 6.1 (Metric graphs and simple paths)**.** Let $G = (\mathcal{V}, \mathcal{E})$ be a connected and locally finite graph. Without loss of generality, we assume that its vertices $\mathcal{V} \subset \mathbb{N}$ are natural numbers. For notational clarity, introduce $\mathcal{E}_{[0,1]} := \mathcal{E} \times [0, 1]$. Let $X := (\mathcal{V} \cup \mathcal{E}_{[0,1]})/\sim$ be the set of equivalent classes generated by the following relation:

$$\forall\, \alpha = \{i, j\} \in \mathcal{E} \text{ with } i < j, \qquad (\alpha, 0) \sim i \text{ and } (\alpha, 1) \sim j. \tag{6.1}$$

Elements in X are written as $[i]$ and $[(\alpha, s)]$, representing the equivalent classes of $i \in \mathcal{V}$ and $(\alpha, s) \in \mathcal{E}_{[0,1]}$ respectively. We identify $\mathcal{V}$ with $V := \{[i] \mid i \in \mathcal{V}\}$ and $\mathcal{E}$ with $E := \{[\alpha] \mid \alpha \in \mathcal{E}\}$ with $[\alpha] := \{[(\alpha, s)] \in X \mid 0 \le s \le 1\}$, which allows us to reuse definitions, such as *vertex*, *edge* and *end*, for X. For the edge α given in (6.1), we associate it with two *oriented edges*, $\overrightarrow{\{i, j\}}$ and $\overrightarrow{\{j, i\}}$, which enumerate points of the set $[\alpha]$ via the parameter s in subscript as follows,

$$\overrightarrow{\{i, j\}}_s := [(\alpha, s)], \qquad \overrightarrow{\{j, i\}}_s := [(\alpha, 1 - s)], \quad \text{ for } s \in [0, 1]. \tag{6.2}$$

In short, an oriented edge $\vec{e}$ is an edge with a given order of its two ends, satisfying $\vec{e} = \overrightarrow{\{\vec{e}_0, \vec{e}_1\}}$.

A *length function* of G is a function $l : \mathcal{E} \to \mathbb{R}$ uniformly bounded from below by a strictly positive number, i.e., $\inf_{\alpha \in \mathcal{E}} l(\alpha) > 0$. Fix such a length function l of G. Via the identification of $\mathcal{E}$ with E, we also consider l as a function defined for (oriented) edges of X.

Given two points $x, y \in X$, a *simple path* from x to y (as well as between x and y) is an *injective* map $\gamma : [a, b] \to X$ defined on a compact interval $[a, b] \subset \mathbb{R}$ with the following properties:

1. $\gamma(a) = x$, $\gamma(b) = y$.

2. If x, y are two different vertices, then there exist a path $p = v_0 v_1 \ldots v_n$ of G and a partition $a = t_0 < t_1 < \cdots < t_n = b$ such that for $k = 0, 1, \ldots, n - 1$ and $0 \le s \le 1$

$$t_{k+1} - t_k = l(\overrightarrow{\{v_k, v_{k+1}\}}) \quad \text{ and } \quad \gamma((1 - s)t_k + s\, t_{k+1}) = \overrightarrow{\{v_k, v_{k+1}\}}_s\,. \tag{6.3}$$

   Otherwise, γ is the restriction of a simple path between two vertices.

For a simple path γ defined on $[a, b]$, we define its length as $b - a$. We now define the metric d_l on X. For two given points $x, y \in X$, we set $d_l(x, y)$ to be the infimum of the lengths of all simple paths from x to y. The metric space (X, d_l) is called a *metric graph*, and we denote it by the triple $\Gamma := (V, E, d_l)$. The graph G is called the base graph of Γ. A metric graph Γ is called a *metric tree* if its base graph G is a tree.

Remark 6.2. Oriented edges of metric graphs are denoted using arrow notation, for instance, $\vec{e}$. A subscript appended to this symbol, such as $\vec{e}_s$, designates a point located within the edge $\vec{e}$, parameterized by s as detailed in (6.2). Conversely, symbols without arrows but with subscripts, such as $e_1, e_2, \ldots$, are employed to denote possibly distinct edges.

We introduce in the following terminologies for simple paths.

Definition 6.3 (Terminologies for simple paths)**.** Consider the space X constructed in Definition 6.1. Given a simple path $\gamma : [a, b] \to X$, we say

1. γ *begins at* an edge e_1 and *ends at* an edge e_2 if $\gamma(a) \in e_1$ and $\gamma(b) \in e_2$;

2. γ is a simple path *from* e_1 *to* e_2 if it begins at e_1, ends at e_2, and its image set contains e_1, e_2;

3. γ *visits* a vertex v if the vertex v is the in the image set of γ;

4. if $v_0 = \gamma(t_0), v_1 = \gamma(t_1), \ldots, v_n = \gamma(t_n)$ are *all* the vertices visited by γ and $a \leq t_0 < t_1 < \ldots < t_n \leq b$, then γ *visits* the sequence of vertices $v_0, v_1, \ldots, v_n$ *in this order*;

5. if $\gamma' : [b, c] \to X$ is another simple path such that $\gamma(b) = \gamma'(b)$, the *concatenation* of γ, γ' is the map $f : [a, c] \to X$ defined via the relations $f|_{[a,b]} = \gamma$ and $f|_{[b,c]} = \gamma'$.

The preceding construction of metric graphs in Definition 6.1 can also be found in classic references such as [22, §1.9] and [23, §3.2.2]. We now state some basic properties, especially the *geometric uniqueness* of simple paths, which are used to demonstrate that (X, d_l) is a valid metric space.

Lemma 6.4. *Consider the space X, the simple paths, and the map d_l introduced in Definition 6.1. Simple paths are geometrically unique in the following sense: given two simple paths $\gamma_1 : [a_1, b_1] \to \Gamma$ and $\gamma_2 : [a_2, b_2] \to \Gamma$ from $x = \gamma_1(a_1) = \gamma_2(a_2)$ to $y = \gamma_1(b_1) = \gamma_2(b_2)$,*

$$\text{if } \gamma_1([a_1, b_1]) = \gamma_2([a_2, b_2]), \quad \text{then } \forall\, t \in [a_1, b_1],\ \gamma_1(t) = \gamma_2(t + c),\ \text{ where } c := a_2 - a_1 = b_2 - b_1.$$

The concatenation of two simple paths results in another simple path if and only if the resulting map is injective. Simple paths between any two given points in X always exist. The function $d_l : X \times X \to \mathbb{R}$ defines a metric on X.

Proof. Given an edge $e = \{v_0, v_1\} \in E$, the requirement (6.3) forces that any two simple paths from v_0 to v_1 with e being their image set can be different at most up to a transition of the definition domain. To prove the claimed geometric uniqueness, we extend γ_1, γ_2 to be simple paths between vertices that still share the same image set, and denote by $v_0, v_1, \ldots, v_n$ the sequence of all vertices visited by them in this order. By comparing consecutively the resections of γ_1, γ_2 whose images are exactly the edge $\{v_i, v_{i+1}\}$ $(i = 0, 1, \ldots, n-1)$, we conclude the geometric uniqueness by applying the preceding property implied by (6.3).

Let $f : [a, c] \to X$ be the concatenation of two simple paths $\gamma : [a, b] \to X$ and $\gamma' : [b, c] \to X$ and assume that f is injective. Since simple paths are themselves concatenations of their restrictions, to prove that f is a simple path, it suffices to consider the case where both the images of γ and γ' are contained in an edge e and $a < b < c$. Given that $\gamma(b) = \gamma'(b)$ and f is injective, the geometric uniqueness implies that both γ and γ' are restrictions of the same simple path between the two ends of e. Hence, f is also a restriction of a simple path, which implies that f is a simple path.

We now prove the claimed existence of simple paths between two given points. Let e_1, e_2 be two edges containing them respectively. Note that, for any given path $v_0 v_1 \ldots v_n$ of the base graph, we can construct a simple path γ from v_0 to v_n that visits the vertices $v_0, v_1, \ldots, v_n$ in this order. Since the base graph is connected, we can thus construct simple paths from e_1 to e_2, which implies the existence of simple paths beginning at e_1 and ending at e_2. Hence, the claimed existence is proven and the map d_l is thus well-defined for any two given points.

We now prove that d_l is a valid metric on X. Fix three arbitrarily chosen points $x, y, z \in X$, we aim to show the following three properties,

$$1.\ d_l(x, y) = 0 \Longleftrightarrow x = y, \quad 2.\ d_l(x, y) = d_l(y, x), \quad 3.\ d_l(x, y) + d_l(y, z) \geq d_l(x, z).$$

Since the length function l is uniformly bounded from below by a positive constant, $d_l(x, y) = 0$ implies that x, y belong to the same edge, which further implies Property 1 by the injectivity of simple paths. For Property 2, note that if $\gamma : [a,b] \to X$ is a simple path from x to y, then $\gamma' : [a,b] \to X$ defined via $\gamma'(t) = \gamma(a+b-t)$ is a simple path from y to x with the same length. To prove Property 3, consider two arbitrarily chosen simple paths, $\gamma_1 : [a_1,b_1] \to X$ from x to y and $\gamma_2 : [a_2,b_2] \to X$ from y to z. As their concatenation is not necessarily injective, we define $c_1 := \inf\{t \in [a_1,b_1] \mid \gamma_1(t) \in \gamma_2([a_2,b_2])\}$ and $c_2 := \sup\{t \in [a_2,b_2] \mid \gamma_2(t) \in \gamma_1([a_1,b_1])\}$. Since $\gamma_1(b_1) = \gamma_2(a_2)$, both c_1 and c_2 are well-defined. Applying the geometric uniqueness to the two simple paths $t \in [0, b_1-a_1] \to \gamma_1(b_1-t)$ and $t \in [0, b_2-a_2] \to \gamma_1(b_2-t)$, we conclude that the two points $\gamma_1(c_1) = \gamma_2(c_2)$ coincide, and is either one of the three points x, y, z or a common vertex in the images of γ_1 and γ_2. Define $\gamma : [0, b_2-c_2+c_1-a_1] \to X$ by setting $\gamma|_{[0,c_1-a_1]} := \gamma_1|_{[a_1,c_1]}$ and $\gamma|_{[c_1-a_1, b_2-c_2+c_1-a_1]} := \gamma_2|_{[c_2,b_2]}$. By our choice of c_1 and c_2, γ is an injective concatenation of two simple paths. Hence, γ is a simple path from x to z, which implies $d_l(x,z) \le b_2 - a_2 + b_1 - a_1$. Since γ_1 and γ_2 are arbitrarily chosen, Property 3 is thus proven by our definition of d_l. □

Remark 6.5 (Explicit formulae for the length of a simple path). Via the simple path defined in Definition 6.1, (oriented) edges of metric graphs are realized as segments interpolating their ends, whose length is determined by the given length function. For a simple path between two different vertices, its length is equal to the sum of the lengths of all edges contained in its image set. For the typical case where both x and y are not vertices and not located at the same edge, we consider a simple path $\gamma : [a,b] \to X$ from x to y. Let $v_0, v_1, \dots, v_n$ be the sequence of vertices visited by γ in this order, and let w_0, w_1 be the two vertices such that $x = \overrightarrow{\{w_0, v_0\}}_{s_0}$ and $y = \overrightarrow{\{v_n, w_1\}}_{s_1}$, where $0 < s_0, s_1 < 1$. Then the length of γ can be calculated as follows,

$$\begin{aligned} \text{length of } \gamma := b - a &= t_0 - a + \sum_{k=0}^{n-1}(t_{k+1} - t_k) + b - t_n \\ &= (1-s_0)\, l(\overrightarrow{\{w_0, v_0\}}) + \sum_{k=0}^{n-1} l(\overrightarrow{\{v_k, v_{k+1}\}}) + s_1\, l(\overrightarrow{\{v_n, w_1\}}), \end{aligned}$$

where we regard γ as a restriction of the simple path corresponding to the path $w_0 v_0 v_1 \dots v_n w_1$, and the equalities $t_0 - a = (1-s_0)\, l(\overrightarrow{\{w_0, v_0\}})$, $b - t_n = s_1\, l(\overrightarrow{\{v_n, w_1\}})$ are implied by (6.3).

For metric trees, the distance between two given points can be reduced directly to the length of a simple path between them.

Lemma 6.6. *Let $\Gamma = (V, E, d_l)$ be a metric tree. For two given points $x, y \in \Gamma$, if γ is a simple path from x to y, then $d_l(x,y)$ is equal to the length of γ.*

Proof. It suffices to show the claim that, up to a translation of the domain, γ is the unique simple path from x to y. We prove this claim by contradiction and assume that there are two simple paths, $\gamma_1 : I_1 \to \Gamma$ and $\gamma_2 : I_2 \to \Gamma$, from x to y that are not a translation of each other. Since Γ is a tree, the geometric uniqueness in Lemma 6.4 excludes immediately the possibility that x and y are located at the same edge. Consider the two sequences of vertices visited by these two simple paths in order. Thanks to the local uniqueness of simple paths, these two sequences must be different, and we can thus find one vertex v_1 present in one sequence but not in the other one. Without loss of generality, we assume $v_1 \in \gamma_1(I_1)$ while $v_1 \notin \gamma_2(I_2)$. Chosen a vertex v_2 visited by γ_2. Since v_1 is the common end of two different edges, $\{v_1, w\}$ and $\{v_1, w'\}$, whose interiors intersect with γ_1, we can find two different paths in the base graph from v_1 to v_2, $v_1 w \dots v_2$ and $v_1 w' \dots v_2$. The union of these two paths contains a cycle in the base graph, which is a contradiction. □

Metric properties of metric graphs

Given an oriented edge $\vec{e}$, for two points $\vec{e}_s$ and $\vec{e}_t$ in it, it is not necessarily true that $d_l(\vec{e}_s, \vec{e}_t) = |t-s|\, l(\vec{e})$ since there could be simple paths between them that have smaller lengths and visit vertices other than $\vec{e}_0, \vec{e}_1$. Thanks to the requirement of a strictly positive global lower bound imposed on length functions, (oriented) edges are locally isometric to intervals of equal length.

Lemma 6.7 (Local isometries of oriented edges)**.** *Let $\Gamma = (V, E, d_l)$ be a metric graph. For an oriented edge $\vec{e}$ of Γ, the map $I^{\vec{e}} : \vec{e} \to [0, l(\vec{e})]$ defined by $I^{\vec{e}}(\vec{e}_s) := s\, l(\vec{e})$ is a local isometry of e,*

$$|I^{\vec{e}}(\vec{e}_s) - I^{\vec{e}}(\vec{e}_t)| = d_l(\vec{e}_s, \vec{e}_t) = |t-s|\, l(\vec{e}) \quad \textit{if } t, s \in [0,1] \textit{ and } |t-s| \le \frac{1}{l(\vec{e})} \inf_{e\in E} l(e).$$

Proof. Note that the map $I^{\vec{e}}$ is bijective. We denote its inverse map by $\gamma : [0, l(\vec{e})] \to \vec{e}$, which by definition is a simple path. If $t, s \in [0,1]$ satisfy

$$0 \le t - s \le \frac{1}{l(\vec{e})} \inf_{e\in E} l(e),$$

then the restriction $\gamma|_{[s\, l(\vec{e}), t\, l(\vec{e})]}$ attains the infimum length among all possible simple paths from $\vec{e}_s$ to $\vec{e}_t$. Indeed, the geometric uniqueness in Lemma 6.6 implies that any other path would necessarily include an edge other than $\vec{e}$ in its image, and thus have a length of at least $\inf_{e\in E} l(e)$. Therefore, $d_l(\vec{e}_s, \vec{e}_t) = |t-s|\, l(\vec{e})$, which concludes the proof. ☐

As a corollary, we prove that metric graphs are length spaces. For a metric graph $\Gamma = (V, E, d_l)$ and a curve $\gamma : [a,b] \to \Gamma$, recall that its length (Definition 1.1) is defined by

$$L_{d_l}(\gamma) := \sup_{a=t_0\le t_1 \le \cdots \le t_N = b} \sum_{i=0}^{N-1} d_l(\gamma(t_i), \gamma(t_{i+1})), \tag{6.4}$$

where the supremum is taken over all possible finite partitions of the compact interval $[a,b]$ using points $a = t_0 \le t_1 \le \cdots \le t_N = b$.

Corollary 6.8. *Let $\Gamma = (V, E, d_l)$ be a metric graph. Simple paths are 1-Lipschitz continuous, and a continuous map $\gamma : [a,b] \to \Gamma$ is a simple path if it is injective and locally isometric. For any two points $x, y \in \Gamma$, the distance between them satisfies $d_l(x,y) = \inf_\gamma L_{d_l}(\gamma)$, where the infimum is taken over all continuous curves γ from x to y. In particular, Γ is a length space.*

Proof. By definition, the length of a simple path between $x, y \in \Gamma$ is larger than the distance $d_l(x,y)$. Since restrictions of simple paths are still simple paths, it follows that simple paths are 1-Lipschitz continuous. Assuming that a continuous map $\gamma : [a,b] \to \Gamma$ is injective and locally isometric, we prove that it is a simple path. If the restriction γ to $(c,d) \subset [a,b]$ is isometric and its image $\gamma((c,d))$ is contained in an oriented edge $\vec{e}$, then by the local isometry of $I^{\vec{e}}(\vec{e}_s) := s\, l(\vec{e})$ (Lemma 6.7), the map $I^{\vec{e}} \circ \gamma|_{(c,d)}$ is simply a transition of intervals, which implies that γ is locally a simple path. By the compactness of $[a,b]$, γ is a concatenation of finitely many simple paths. Since γ is injective, Lemma 6.4 implies that γ is a simple path.

We now prove the last part of our proposition. In the infimum $\inf_\gamma L_{d_l}(\gamma)$ over all possible continuous curves γ from x to y, it suffices to consider only injective ones. Thanks to the existence of natural parameterization [23, Proposition 2.5.9], we can replace a continuous curve with a Lipschitz map that is locally isometric without changing the length. Therefore, according to the definition of d_l, it remains to show that for a simple path $\gamma : [a, b] \to X$, we have the equality $L_{d_l}(\gamma) = b - a$. By Lemma 6.7, there exists a partition $a = t_0 \le t_1 \le \cdots \le t_N = b$ of $[a, b]$ such that the restriction of γ on each interval $[t_i, t_{i+1}]$ is isometric. Since $\sum_{i=0}^{N-1} d_l(\gamma(t_i), \gamma(t_{i+1})) = b - a$, we have $L_{d_l}(\gamma) \ge b - a$. Note that the sum $\sum_{i=0}^{N-1} d_l(\gamma(t_i), \gamma(t_{i+1}))$ remains unchanged if we add more partition points. By adding the partition points $\{t_i\}_{0 \le i \le N}$ to any given partition, we obtain $L_{d_l}(\gamma) \le b - a$, which concludes the proof. □

We prove some basic metric properties of metric graphs in the following theorem. Recall that a length space is geodesic (Definition 1.2) if the distance of two given points is equal to the length of some rectifiable curve connecting them.

Theorem 6.9. *Metric graphs are proper, complete, separable, and geodesic metric spaces.*

Proof. We first show that metric graphs are proper metric spaces, i.e., closed and bounded subsets of metric graphs are compact. Fix a metric graph $\Gamma = (V, E, d_l)$ and let $\epsilon > 0$ be a uniform lower bound of the length function l of Γ. For any given vertex $v \in V$, the closed metric ball $B(v, \epsilon)$ is contained in finitely many edges since the base graph of Γ is required to be locally finite by Definition 6.1 and any edge intersecting with the ball $B(v, \epsilon)$ must have v as one of its ends. We now prove by mathematical induction the claim that for any $n = 1, 2, \ldots$, the closed metric ball $B(v, n\,\epsilon)$ is contained in finitely many edges. The case for $n = 1$ is already shown. Assume that the claim is true for $n = k$. Consider the set W of all ends of edges that intersect with $B(v, k\,\epsilon)$, which by assumption is finite. By the triangle inequality, we have $B(v, (k+1)\,\epsilon) \subset \cup_{w \in W} B(w, \epsilon)$, which implies the claim for $n = k + 1$ since each $B(w, \epsilon)$ is shown be contained in finitely many edges and W is a finite set. Therefore, the claim holds for any $n \in \mathbb{N}^*$. Since simple paths are continuous (Corollary 6.8), each edge of Γ and thus the union of finitely many edges is compact. Moreover, as Γ is connected, any bounded set must be contained in one of the metric balls $B(v, n\,\epsilon)$. It follows that a closed and bounded subset of Γ is compact. In particular, metric graphs are locally compact.

Recall that any proper metric space is complete and separable (Section 1.1). Moreover, a complete locally compact length space is always geodesic [23, Theorem 2.5.23]. □

Remark 6.10. The requirement of metric graphs being locally finite is crucial for Theorem 6.9 to hold. Consider the bouquet with countably many edges, as described in [23, Example 3.1.17]. This space can be viewed as the product space of countably many copies of the unit interval $[0, 1]$, all joined at the vertex 0. Since the set of all vertices (the endpoints 1 of each interval plus the shared endpoint 0) has no convergent subsequence, this space fails to be locally compact.

The canonical reference measure

The canonical measure $\mathcal{H}$ on Γ is defined using its edges and vertices, which is indeed the one-dimensional Hausdorff measure [23, §1.7] on Γ.

Definition 6.11 (Canonical measures on metric graphs)**.** Let $\Gamma = (V, E, d_l)$ be a metric graph. The canonical measure $\mathcal{H}$ is the measure on Γ that *gives no mass to the set of vertices* V, and for each oriented edge $\vec{e} \in E$, the image measure $I^{\vec{e}}_{\#}[\mathcal{H}|_{\vec{e}}]$ of its restriction $\mathcal{H}|_{\vec{e}}$ to $\vec{e}$ is the Lebesgue measure on $[0, l(\vec{e})]$, where $I^{\vec{e}} : \vec{e} \to [0, l(\vec{e})]$ is the local isometry sending $\vec{e}_s$ to $s\, l(\vec{e})$.

Since a metric graph Γ has at most countably many edges, the above definition uniquely determines a σ-finite canonical measure $\mathcal{H}$. We always assume that a metric graph is equipped with its canonical measure, and is thus a *metric measure graph*. For example, a Borel measure on a metric graph is said to be *absolutely continuous* if it is absolutely continuous with respect to the graph's canonical measure.

6.1.2 Curvature bounds on metric trees

Our focus in several upcoming sections will be on metric trees. In this subsection, we introduce the concept of metric spaces with curvature bounded above. This concept provides a basis for comparing metric trees with Riemannian manifolds, and thus offers a glimpse into how curvature bounds influence the properties of Wasserstein barycenters (to be illustrated later in Section 6.3). Our presentation, adapted from [22, Chapter II.1], is for illustration purpose only; these results will not be used in subsequent proofs. We begin with the definitions of geodesic triangles and their comparison triangles.

Definition 6.12 (Geodesic triangles, comparison triangles and comparison points)**.** Let (E, d) be a geodesic space. A geodesic segment connecting two points $p, q \in E$ is the image of a Lipschitz curve of length $d(p, q)$ from p to q. By convention, we denote by $[p, q]$ a definitely chosen geodesic segment connecting p and q. A geodesic triangle Δ in E consists of three points $p, q, r \in E$, its vertices, and a choice of three geodesic segments $[p, q], [q, r], [r, p]$ connecting them, its sides. Such a geodesic triangle will be denoted by $\Delta([p, q], [q, r], [r, p])$. If a point $x \in E$ lies in the union of $[p, q]$, $[q, r]$ and $[r, p]$, then we write $x \in \Delta$.

Given a real number $k \in \mathbb{R}$, denote by M_k^2 the model space of dimension 2 and constant sectional curvature k. If $k = 0$, then $M_0^2 := \mathbb{R}^2$ is the Euclidean plan. For $k \neq 0$, M_k^2 is obtained from the sphere $\mathbb{S}^2$ (if $k > 0$), or the hyperbolic plane $\mathbb{H}^2$ (if $k < 0$), by multiplying the distance function by the constant $1/\sqrt{|k|}$. Denote by d_k the Riemannian distance function of M_k^2. A geodesic triangle $\overline{\Delta}$ with vertices $\bar{p}, \bar{q}, \bar{r}$ in M_k^2 is called a comparison triangle for $\Delta = \Delta([p, q], [q, r], [r, p])$ if $d_k(\bar{p}, \bar{q}) = d(p, q)$, $d_k(\bar{q}, \bar{r}) = d(q, r)$ and $d_k(\bar{p}, \bar{r}) = d(p, r)$. A point $\overline{x} \in [\bar{q}, \bar{r}]$ is called a comparison point for $x \in [q, r]$ if $d_k(\bar{q}, \overline{x}) = d(q, x)$. Comparison points on the sides $[\bar{p}, \bar{q}]$ and $[\bar{p}, \bar{r}]$ are defined in the same way.

We now define metric spaces with curvature bounded above by k [22, Definition 1.2 of Chapter II.1]. For simplicity, we restrict our attention to the case where $k \leq 0$. The definition for $k > 0$ is similar, but requires slightly more care because the diameter of the model space M_k^2 is $\pi/\sqrt{k}$.

Definition 6.13 (Metric spaces with curvature bounded from above)**.** Let (E, d) be a geodesic metric space and let $k \leq 0$ be a real number. Fix a geodesic triangle Δ in E and a comparison triangle $\overline{\Delta} \subset M_k^2$ for Δ in the model space (M_k^2, d_k) of constant sectional curvature k. The triangle Δ is said to satisfy the CAT(k) inequality if for any two points $x, y \in \Delta$ with their comparison points $\overline{x}, \overline{y} \in \overline{\Delta}$,

$$d(x, y) \leq d_k(\overline{x}, \overline{y}).$$

The metric space E is said to be *of curvature* $\leq k$ if for every $x \in X$, there exists a metric ball centered at x such that any geodesic triangle contained in it satisfies the CAT(k) inequality.

We stress that metric spaces of curvature $\leq k$ are defined by *local* satisfaction of the CAT(k) inequality. By contrast, metric spaces satisfying the CAT(k) inequality globally are referred to as

CAT(k) spaces in the literature. The above definition via local CAT(k) inequality, often called curvature bounds from above in the sense of Alexandrov, generalizes the concept of Riemannian manifolds with sectional curvature bounded from above. This generalization is justified by the following theorem [22, 1A.6 of Chapter II.1].

Theorem 6.14. *Fix a real number $k \leq 0$. A complete Riemannian manifold (M, g) is a metric space of curvature $\leq k$ in the sense of Definition 6.13 if and only if the sectional curvature of M is less than or equal to k.*

Observe that for a geodesic triangle in a metric tree, any given side is contained in the union of the other two sides. Using this observation along with the triangle inequality for distance functions in model spaces, one can deduce the following curvature property of metric trees [22, (5) of Example 1.15 in Chapter II.1].

Proposition 6.15. *A metric tree is of curvature $\leq k$ for all real number $k \leq 0$.*

Proof. We fix a real number $k \leq 0$ and prove the CAT(k) inequality for geodesic triangles in metric trees. Fix a geodesic triangle Δ with vertices p, q, r in a metric tree and consider its comparison triangle $\overline{\Delta}$ with vertices $\bar{p}, \bar{q}, \bar{r}$ in the model space M_k^2. The CAT(k) inequality holds trivially, as an equality, for the case that the three vertices p, q, r of Δ located in the same geodesic segment.

We are left to consider the non-trivial case that the comparison triangle $\overline{\Delta}$ is not degenerate. Without loss of generality, it suffices to consider two points $x \in [p, q]$ and $y \in [q, r]$ and prove

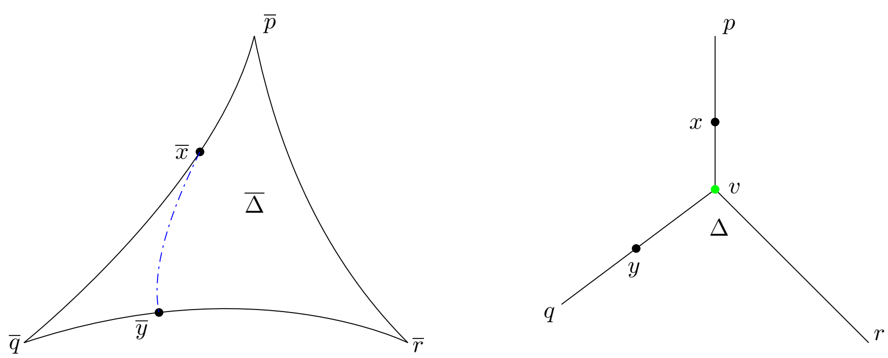


Figure 6.3: Comparison triangle $\overline{\Delta}$ in M_k^2 for Δ in a metric tree

$d(x, y) \leq d_k(\overline{x}, \overline{y})$, where $\overline{x} \in [\bar{p}, \bar{q}], \overline{y} \in [\bar{q}, \bar{r}]$ are the comparison points for x, y. As in Figure 6.3, let v be the unique vertex contained in the three sides of Δ. By the definition of comparison points, we have

$$d_k(\overline{x}, \bar{q}) = d(x, q) = d(x, v) + d(v, q), \quad d_k(\bar{q}, \overline{y}) = d(q, y) = d(q, v) - d(v, y)$$

It follows from the triangle inequality of d_k that

$$d_k(\overline{x}, \overline{y}) \geq d_k(\overline{x}, \bar{q}) - d_k(\bar{q}, \overline{y}) = d(x, v) + d(v, q) - d(q, v) + d(v, y) = d(x, v) + d(v, y) = d(x, y),$$

which is the equality to prove. $\square$

Remark 6.16. In the above proof of Proposition 6.15, the CAT(k) inequality is proven globally. In particular, each metric tree is a CAT(0) space, which is also alternatively referred to a *global NPC* space in [94, Proposition 3.4]. According to [94, Proposition 4.3], barycenters of probability measures on metric trees are unique. We shall see in Proposition 6.63 an example showing that Wasserstein barycenters on the Wasserstein space $\mathcal{W}_2(\Gamma)$ over a metric tree Γ are not unique, which in particular implies that $\mathcal{W}_2(\Gamma)$ is not a CAT(0) space (c.f. [14, Remark 2.10] [59, remark after Proposition 1.4] [8, Example 7.3.3]).

6.1.3 Wasserstein barycenters on the real line

Since the real line $\mathbb{R}$ can be represented as a metric tree (with integers being its vertices) and any edge of a metric tree is isometric to a compact interval, it is helpful to first investigate properties of Wasserstein barycenters on $\mathbb{R}$. As reviewed in (Section 1.4.1), any optimal transport problem on $\mathbb{R}$ admits an explicit solution. This solution underlies the formula of Wasserstein barycenter presented in Theorem 6.18.

Recall that $L^2([0,1])$ denotes the Hilbert space of squared integrable functions on $[0,1]$ with respect to the Lebesgue measure and f_μ^{-1} denotes the quantile function (Definition 1.28) of a probability measure μ on $\mathbb{R}$. The Wasserstein space $\mathcal{W}_2(\mathbb{R})$ inherits the linear structure of $L^2([0,1])$ via quantile functions, as shown by the following formula in Theorem 1.37,

$$d_W(\mu,\nu)^2 = \int_0^1 [f_\mu^{-1}(t) - f_\nu^{-1}(t)]^2 \,\mathrm{d}\, t. \tag{6.5}$$

Proposition 6.17. *The following subset Q of $L^2([0,1])$ is convex and closed,*

$$Q := \{g \in L^2([0,1]) \mid g \text{ coincides with a non-decreasing function on } (0,1) \text{ almost everywhere}\}.$$

The map $F : \mathcal{W}_2(\mathbb{R}) \to Q$ sending μ to f_μ^{-1} is a surjective isometry.

Proof. The convexity of Q follows from its definition. According to Theorem 1.37, $\mu \in \mathcal{W}_2(\mathbb{R})$ if and only if $f_\mu^{-1} \in L^2([0,1])$. Moreover, it follows from (6.5) that F is an isometry.

We now prove that F is surjective. Fix an element $g \in Q$. Since any monotone function has at most countably many points of discontinuity [17, Corollary 5.2.4], we can modify g on a negligible set such that g is right-continuous and non-decreasing on $(0,1)$ with $g(0) = \lim_{t\downarrow 0} g(t)$ and $g(1) = \lim_{t\uparrow 1} g(t)$. According to Lemma 1.29, the function $f(x) := \inf_{t\in(0,1)}\{t \mid g(t) > x\}$ defined for $x \in \mathbb{R}$ is right-continuous and non-decreasing. By definition of f, we also have that $\lim_{x\to-\infty} f(x) = 0$ and $\lim_{x\to+\infty} f(x) = 1$. If follows that there exists exactly one probability measure μ on $\mathbb{R}$ such that $f_\mu = f$ [29, Proposition 4.4.3]. By Lemma 1.34, g coincides with f_μ^{-1} on $(0,1)$, which shows that F is surjective.

Last, since $(\mathcal{W}_2(\mathbb{R}), d_W)$ is a complete metric space, Q is a closed set. ◻

Thanks to the above linear structure, explicit calculations of Wasserstein barycenters on $\mathbb{R}$ are possible.

Quantile function formula for Wasserstein barycenters on the real line

For a finitely supported probability measure $\mathbb{P} = \sum_{i=1}^{n} \lambda_i \delta_{\nu_i}$, the quantile function of its barycenter $\mu_{\mathbb{P}}$ is given by the formula [79, §3.1.4],

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

Via the isometric embedding $F : \mathcal{W}_2(\mathbb{R}) \to Q$ (see Proposition 6.17), this formula neatly translates to a linear combination in Q:

$$F(\mu_{\mathbb{P}}) = \sum_{i=1}^{n} \lambda_i\, F(\nu_i).$$

This linearity demonstrates that the Wasserstein barycenter problem on $\mathbb{R}$ simplifies significantly thanks to the linear structure of $L^2([0,1])$. Building upon this, the following theorem extends this result to general measures $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$. Our proof proceeds in two steps: first, we show that $g(t) := \int_{\mathcal{W}_2(\mathbb{R})} f_{\nu}^{-1}(t)\, \mathrm{d}\,\mathbb{P}(\nu)$ defines a valid quantile function; second, we prove that $g \in Q$ is indeed the barycenter of $F_{\#}\mathbb{P}$, crucially employing the linear structure of Q via Fubini's theorem.

Theorem 6.18 (Wasserstein barycenters on the real line)**.** *Let $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$ be a probability measure on the Wasserstein space $(\mathcal{W}_2(\mathbb{R}), d_W)$. Then $\mathbb{P}$ has a unique Wasserstein barycenter $\mu_{\mathbb{P}} \in \mathcal{W}_2(\mathbb{R})$, whose quantile function satisfies*

$$f_{\mu_{\mathbb{P}}}^{-1}(t) = \int_{\mathcal{W}_2(\mathbb{R})} f_{\nu}^{-1}(t)\, \mathrm{d}\,\mathbb{P}(\nu), \quad \forall\, t \in [0,1]. \tag{6.6}$$

In particular, the integral in (6.6) is finite for $t \in (0,1)$, and the inequality still holds when it takes possibly infinite values for the case $t = 0, 1$.

Proof. By Lemma 1.35, the map $\nu \mapsto f_{\nu}^{-1}(t)$ is upper semi-continuous for $t \in [0,1)$ and lower semi-continuous for $t = 1$, which implies that it is measurable for any $t \in [0,1]$ [29, p.176]. It follows from Theorem 1.37 and Fubini's theorem that

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

where we applied the property that the quantile function of δ_0 is the constant 0, i.e., $f_{\delta_0}^{-1} \equiv 0$. If follows from the Cauchy–Schwarz inequality that the function $g : [0,1] \to \overline{\mathbb{R}}$ defined by $g(t) := \int_{\mathcal{W}_2(\mathbb{R})} f_{\nu}^{-1}(t)\, \mathrm{d}\,\mathbb{P}(\nu)$ for $t \in [0,1]$ is an element in $L^2([0,1])$.

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

According to Proposition 6.17, since $g|_{(0,1)}$ is non-decreasing and right-continuous, there exists a unique measure $\mu_{\mathbb{P}} \in \mathcal{W}_2(\mathbb{R})$ such that $f_{\mu_{\mathbb{P}}}^{-1} = g$. It remains to show that $\mu_{\mathbb{P}}$ is the unique barycenter of $\mathbb{P}$. For any measure $\eta \in \mathcal{W}_2(\mathbb{R})$, by Theorem 1.37 and Fubini's theorem, we have

$$\begin{aligned}\int_{\mathcal{W}_2(\mathbb{R})} d_W(\eta,\nu)^2 \,\mathrm{d}\,\mathbb{P}(\nu) &= \int_0^1 \int_{\mathcal{W}_2(\mathbb{R})} [f_\eta^{-1}(t) - f_\nu^{-1}(t)]^2 \,\mathrm{d}\,\mathbb{P}(\nu)\,\mathrm{d}\,t \\ &= \int_0^1 \left[f_\eta^{-1}(t) - \int_{\mathcal{W}_2(\mathbb{R})} f_\nu^{-1}(t)\,\mathrm{d}\,\mathbb{P}(\nu)\right]^2 \mathrm{d}\,t + I(\mathbb{P}), \end{aligned} \tag{6.7}$$

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

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

It follows from (6.7) that the infimum $\inf_{\eta\in\mathcal{W}_2(\mathbb{R})} \int_{\mathcal{W}_2(\mathbb{R})} d_W(\eta,\nu)^2 \,\mathrm{d}\,\mathbb{P}(\nu)$ is reached by η if and only if $f_\eta^{-1}(t) = \int_{\mathcal{W}_2(\mathbb{R})} f_\nu^{-1}(t) = g(t) = f_{\mu_{\mathbb{P}}}^{-1}(t)$ for almost every $t \in [0,1]$. Therefore, $\mu_{\mathbb{P}}$ is a barycenter of $\mathbb{P}$ and its uniqueness follows from the injectivity of the embedding in Proposition 6.17. ☐

6.2 A reduction technique for metric trees

Metric graphs, being Polish spaces (Theorem 6.9), fit the general theory framework of optimal transport and Wasserstein spaces. In this section, we introduce a reduction technique to simplify optimal transport problems on metric trees. This technique allows us to recover the optimal transport plans for the case where one measure is supported in an edge, by reducing the problem to a corresponding optimal transport problem on the real line.

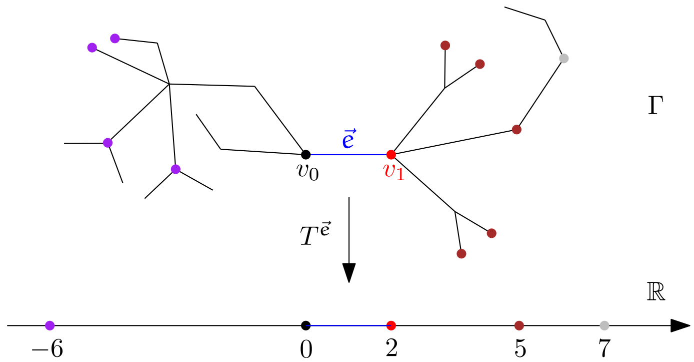
On a metric tree $\Gamma = (V, E, d_l)$, each edge $e \in E$ is isometric to an interval of length $l(e)$ (Lemma 6.6). Extending this isometry to the entire tree yields the following reduction map, which preserves distances for certain simple paths.

Proposition 6.19 (The reduction map associated to an oriented edge)**.** *Let $\Gamma = (V, E, d_l)$ be a metric tree and let $\vec{\mathfrak{e}} = \overrightarrow{\{v_0, v_1\}}$ be an oriented edge of Γ. There exists a unique map $T^{\vec{\mathfrak{e}}} : \Gamma \to \mathbb{R}$ of $\vec{\mathfrak{e}}$ defined by the following conditions,*

1. *$T^{\vec{\mathfrak{e}}}(v_0) = 0$, $T^{\vec{\mathfrak{e}}}(v_1) = l(\vec{\mathfrak{e}})$, and $T^{\vec{\mathfrak{e}}}$ restricted to $\overrightarrow{\{v_0, v_1\}}$ is an isometry onto $[0, l(\vec{\mathfrak{e}})]$;*
2. *given any simple path $\gamma : [a,b] \to \Gamma$ such that $\gamma(a)$ or $\gamma(b)$ is located at the edge $\overrightarrow{\{v_0, v_1\}}$, $T^{\vec{\mathfrak{e}}}$ composed with γ is an isometry,*

$$\forall\, t, s \in [a,b], \quad d_l(\gamma(t), \gamma(s)) = |T^{\vec{\mathfrak{e}}}(\gamma(t)) - T^{\vec{\mathfrak{e}}}(\gamma(s))|.$$

The map $T^{\vec{\mathfrak{e}}}$ is called the reduction map associated to $\vec{\mathfrak{e}} = \overrightarrow{\{v_0, v_1\}}$.



Proof. We first prove the existence. Fix a point $x \in \Gamma$. If $x \in \vec{e}$, then set $T^{\vec{e}}(x) := d_l(x, v_0)$. As for the case that $x \notin \vec{e}$, we first claim that $d_l(x, v_0) \neq d_l(x, v_1)$. Consider the two geodesics connecting x and v_0, v_1 respectively. If $d_l(x, v_0) = d_l(x, v_1)$, then by Lemma 6.6 none of them contains the edge $\vec{e}$, which contradicts the assumption that Γ is a metric tree since these two geodesics and the edge $\vec{e}$ produce a cycle. This contradiction proves our claim. With the help of this claim, in the case that $x \notin \vec{e}$, we can define $T^{\vec{e}}(x) := d_l(x, v_0)$ if $d(x, v_0) > d(x, v_1)$, and $T^{\vec{e}}(x) := -d_l(x, v_0)$ if $d_l(x, v_0) < d(x, v_1)$.

We now show that the previously defined function $T^{\vec{e}}$ satisfies the second property. Recall that the length of a simple path is equal to the distance between its two endpoints (Lemma 6.6). By our construction, $T^{\vec{e}}$ is a continuous function satisfying $|T^{\vec{e}}(y) - T^{\vec{e}}(z)| = d_l(y, z)$ for $y \in \vec{e}$ and $z \in \Gamma$. Therefore, if $\gamma : [a, b] \to \Gamma$ is a simple path such that $\gamma(a) \in \{v_0, v_1\}$, then

$$b - a = d_l(\gamma(b), \gamma(a)) = |T^{\vec{e}}(\gamma(b)) - T^{\vec{e}}(\gamma(a))|. \tag{6.8}$$

Consider the restrictions of γ to the intervals $[a, s]$ for $s \in (a, b)$. Since (6.8) also holds for these restrictions and the composited function $T^{\vec{e}} \circ \gamma : [a, b] \to \mathbb{R}$ is continuous, the function $T^{\vec{e}} \circ \gamma - T^{\vec{e}} \circ \gamma(a)$ must be always non-positive or always non-negative, i.e.,

$$T^{\vec{e}} \circ \gamma(s) - T^{\vec{e}} \circ \gamma(a) = s - a \text{ for } s \in (a, b] \quad \text{ or } \quad T^{\vec{e}} \circ \gamma(s) - T^{\vec{e}} \circ \gamma(a) = a - s \text{ for } s \in (a, b],$$

which further implies that $T^{\vec{e}} \circ \gamma$ is monotone and isometric. Hence, for $t, s \in [a, b]$, $d_l(\gamma(t), \gamma(s)) = |t - s| = |T^{\vec{e}}(\gamma(t)) - T^{\vec{e}}(\gamma(s))|$.

As for the uniqueness, note that a real number is uniquely determined by its distance to 0 and $l(\vec{e})$. Hence, for $x \in \Gamma$, $T^{\vec{e}}(x)$ is uniquely determined by the distances $d_l(x, v_0)$ and $d_l(x, v_1)$. $\square$

The reduction map associated to an edge induces a push-forward map from $\mathcal{W}_2(\Gamma)$ to $\mathcal{W}_2(\mathbb{R})$. To simplify notation in subsequent development, we use the symbol $\mathcal{T}$ to denote this map.

Definition 6.20 (The push-forward map associated to an oriented edge)**.** Let $\Gamma = (V, E, d_l)$ be a metric tree and let $\vec{e}$ be an oriented edge of Γ. We denote by $\mathcal{T}^{\vec{e}}$, or simply by $\mathcal{T}$ (when the oriented edge $\vec{e}$ is explicitly given in the context), the map

$$\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$$

defined for $\mu \in \mathcal{W}_2(\Gamma)$ by the formula $\mathcal{T}(\mu) := {T^{\vec{e}}}_{\#}\mu$, where $T^{\vec{e}}$ is the reduction map associated to $\vec{e}$ (Proposition 6.19). We call $\mathcal{T}$ the push-forward map associated to $\vec{e}$.

The push-forward map $\mathcal{T}$ preserves several properties of probability measures across the Wasserstein spaces $\mathcal{W}_2(\Gamma)$ and $\mathcal{W}_2(\mathbb{R})$. Let us start with the absolute continuity. To avoid confusion, we remark that there is no relation between the symbol $\vec{\mathfrak{e}}$ (an oriented edge) and e (an edge).

Lemma 6.21. *Let $\Gamma = (V, E, d_l)$ be a metric tree. Fix an oriented edge $\vec{\mathfrak{e}}$ of Γ and let $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ be the push-forward map associated to $\vec{\mathfrak{e}}$ (Definition 6.20). Then for any $\mu \in \mathcal{W}_2(\Gamma)$,*

$$\mu \text{ is absolutely continuous} \quad \Longleftrightarrow \quad \mathcal{T}(\mu) \text{ is absolutely continuous.}$$

Proof. Recall from Definition 6.11 that canonical measure $\mathcal{H}$ of Γ gives no mass to the vertices of Γ, and on each edge e, $\mathcal{H}|_e$ is the Lebesgue measure after identifying e with an interval of equal length.

We first prove the case where $\mu \in \mathcal{W}_2(\Gamma)$ is supported in some edge $e \in E$. Consider a simple path γ from $\vec{\mathfrak{e}}$ to e. The second property of $T^{\vec{\mathfrak{e}}}$ applied to γ implies that the map $T^{\vec{\mathfrak{e}}}|_e : e \to T^{\vec{\mathfrak{e}}}(e)$ is a metric isomorphism. Since μ is supported in e, μ is absolutely continuous with respect to $\mathcal{H}|_e$ if and only if $\mathcal{T}(\mu)$ is absolutely continuous with respect to $T^{\vec{\mathfrak{e}}}{}_{\#}[\mathcal{H}|_e]$. By definition of $\mathcal{H}$, $T^{\vec{\mathfrak{e}}}{}_{\#}[\mathcal{H}|_e]$ is the Lebesgue measure restricted to $T^{\vec{\mathfrak{e}}}(e)$, which proves the lemma for the particular case of μ.

Now consider the general case for $\mu \in \mathcal{W}_2(\Gamma)$. As Γ has at most countably many edges, we can re-write μ as $\mu := \sum_{j \in J} \lambda_j \, \mu_j$ such that for each index $j \in J \subset \mathbb{N}$, $0 < \lambda_j < 1$ and $\mu_j \in \mathcal{W}_2(\Gamma)$ is supported in some edge of Γ. Note that, with respect to a given measure, a sum of at most countably many non-negative measures is absolutely continuous if and only if each measure in the sum is so. Since $\mathcal{T}$ is a push-forward map, $\mathcal{T}(\mu) = \sum_{j \in J} \lambda_j \, \mathcal{T}(\mu_j)$. Hence, the general case follows from the previously proven case. □

The push-forward map $\mathcal{T}$ also helps to reduce an optimal transport problem on Γ to an optimal transport problem on $\mathbb{R}$, which relies on the following two properties of $T^{\vec{\mathfrak{e}}}$.

1. $T^{\vec{\mathfrak{e}}}$ preserves the distance of two given points if one of them is contained in the edge $\vec{\mathfrak{e}}$;

2. for any edge $e \in E$, $T^{\vec{\mathfrak{e}}}$ is injective on the set $\vec{\mathfrak{e}} \cup e$.

The first property ensures that the push-forward map induced by $T^{\vec{\mathfrak{e}}} \times T^{\vec{\mathfrak{e}}} : \Gamma \times \Gamma \to \mathbb{R} \times \mathbb{R}$ preserves the optimality of couplings thanks to the cyclical monotonicity characterization of optimal transport plans, as we shall see in (6.9). With the second property, we can show that the push-forward map is surjective as a map from the couplings of μ and ν to the couplings of $\mathcal{T}(\mu)$ and $\mathcal{T}(\nu)$. These two properties are used, in the following theorem, to demonstrate the following two inequalities respectively,

$$d_W(\mu, \nu) \leq d_W(\mathcal{T}(\mu), \mathcal{T}(\nu)) \quad \text{and} \quad d_W(\mu, \nu) \geq d_W(\mathcal{T}(\mu), \mathcal{T}(\nu)).$$

Note that the symbol d_W is employed to denote both the Wasserstein metrics of $\mathcal{W}_2(\Gamma)$ and $\mathcal{W}_2(\mathbb{R})$.

Theorem 6.22. *Let $\Gamma = (V, E, d_l)$ be a metric tree. Fix an oriented edge $\vec{\mathfrak{e}}$ of Γ and let $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ be the push-forward map $\mathcal{T}$ associated to $\vec{\mathfrak{e}}$ (Definition 6.20). For two given probability measures $\mu, \nu \in \mathcal{W}_2(\Gamma)$, if μ is supported in the edge $\vec{\mathfrak{e}}$, then*

$$d_W(\mu, \nu) = d_W(\mathcal{T}(\mu), \mathcal{T}(\nu)).$$

Proof. For any coupling γ of μ and ν, since μ is supported in $\vec{\mathfrak{e}}$, Proposition 6.19 implies that

$$\begin{aligned}\int_{\Gamma\times\Gamma} d_l(x,y)^2 \,\mathrm{d}\,\gamma(x,y) &= \int_{\Gamma\times\Gamma} |T^{\vec{\mathfrak{e}}}(x) - T^{\vec{\mathfrak{e}}}(y)|^2 \,\mathrm{d}\,\gamma(x,y) \\ &= \int_{\mathbb{R}\times\mathbb{R}} |x-y|^2 \,\mathrm{d}[T^{\vec{\mathfrak{e}}} \times T^{\vec{\mathfrak{e}}}]_{\#}\gamma(x,y),\end{aligned} \tag{6.9}$$

which shows $d_W(\mathcal{T}(\mu), \mathcal{T}(\nu)) \le d_W(\mu,\nu)$ since $[T^{\vec{\mathfrak{e}}} \times T^{\vec{\mathfrak{e}}}]_{\#}\gamma$ is a coupling of $\mathcal{T}(\mu)$ and $\mathcal{T}(\nu)$.

We now prove the inequality $d_W(\mu,\nu) \le d_W(\mathcal{T}(\mu), \mathcal{T}(\nu))$. Let η be an optimal transport plan between $\mathcal{T}(\mu)$ and $\mathcal{T}(\nu)$. According to (6.9), it suffices to find a coupling of μ and ν such that $T^{\vec{\mathfrak{e}}} \times T^{\vec{\mathfrak{e}}}$ pushes forward it to η. Rewrite $\nu = \sum_{e\in E} \lambda_e \,\nu_e + \sum_{v\in V} \lambda_v \,\delta_v$, where $\lambda_e, \lambda_v \in [0,1]$, δ_v denotes the Dirac measure supported at the vertex $v \in V$ and $\nu_e \in \mathcal{W}_2(\Gamma)$ is a probability measure supported in the edge e that gives no mass to the ends of e. Denote by f_e (respectively f_v) the density functions of $\mathcal{T}(\nu_e)$ (respectively $\mathcal{T}(\delta_v)$) with respect to $\mathcal{T}(\nu)$. It follows that $\sum_{e\in E} \lambda_e \,f_e + \sum_{v\in V} \lambda_v \,f_v = 1$ for $\mathcal{T}(\nu)$-almost everywhere. Introduce the probability measures $\eta_e(\mathrm{d}\,x, \mathrm{d}\,y) := f_e(y) \cdot \eta(\mathrm{d}\,x, \mathrm{d}\,y)$ and $\eta_v(\mathrm{d}\,x, \mathrm{d}\,y) := f_v(y) \cdot \eta(\mathrm{d}\,x, \mathrm{d}\,y)$. By our choices of f_e and f_v, η_e is a coupling of $\mathcal{T}(\mu)$ and $\mathcal{T}(\nu_e)$ if $\lambda_e \ne 0$ and $\eta_v = \mathcal{T}(\mu) \otimes \mathcal{T}(\delta_v)$ is the product measure of its two marginals if $\lambda_v \ne 0$. Moreover, $\eta = \sum_{e\in E} \lambda_e \,\eta_e + \sum_{v\in V} \lambda_v \,\eta_v$. We are now ready to construct a coupling between μ and ν as follows.

Fix an edge $e \in E$. We claim that $T^{\vec{\mathfrak{e}}}$ is injective on the set $\vec{\mathfrak{e}} \cup e$. Since the base graph of Γ is connected, there exists a simple path from $\vec{\mathfrak{e}}$ to e, and Proposition 6.19 shows that $T^{\vec{\mathfrak{e}}}$ is injective on its image, which proves our claim. Set $U_e := T^{\vec{\mathfrak{e}}}(\vec{\mathfrak{e}} \cup e) \subset \mathbb{R}$ and denote by $S_e : U_e \to \vec{\mathfrak{e}} \cup e$ the inverse map of $T^{\vec{\mathfrak{e}}}|_{\vec{\mathfrak{e}}\cup e}$. Since $\vec{\mathfrak{e}} \cup e$ contains the support of μ and ν_e, U_e contains the support of $\mathcal{T}(\nu_e)$ and $\mathcal{T}(\mu)$. Hence, $U_e \times U_e$ contains the support of η_e, which allows us to define the probability measure $\gamma_e := [S_e \times S_e]_{\#}\eta_e$ on $\Gamma \times \Gamma$. For $e \in E$ such that $\lambda_e \ne 0$, since S_e is the inverse map of $T^{\vec{\mathfrak{e}}}|_{\vec{\mathfrak{e}}\cup e}$ and η_e is a coupling of $\mathcal{T}(\mu)$ and $\mathcal{T}(\nu_e)$, γ_e is a coupling of μ and ν_e, and $\eta_e = [T^{\vec{\mathfrak{e}}} \times T^{\vec{\mathfrak{e}}}]_{\#}\gamma_e$.

For $v \in V$, note that the measure $\gamma_v := \mu \otimes \delta_v$ satisfies $\eta_v = [T^{\vec{\mathfrak{e}}} \times T^{\vec{\mathfrak{e}}}]_{\#}\gamma_v$ if $\lambda_v \ne 0$. As a sum of at most countably many probability measures on $\Gamma \times \Gamma$, $\gamma := \sum_{e\in E} \lambda_e \,\gamma_e + \sum_{v\in V} \lambda_v \,\gamma_v$ is a well-defined probability measure satisfying $[T^{\vec{\mathfrak{e}}} \times T^{\vec{\mathfrak{e}}}]_{\#}\gamma = \eta$. Moreover, γ is coupling of μ and ν since γ_e (respectively γ_v) is a coupling of μ and ν_e (respectively δ_v). Therefore, we have $d_W(\mu,\nu) \le d_W(\mathcal{T}(\mu), \mathcal{T}(\nu))$, which implies the equality $d_W(\mu,\nu) = d_W(\mathcal{T}(\mu), \mathcal{T}(\nu))$. ☐

Remark 6.23. Note that Theorem 6.22 does not assert the uniqueness of optimal transport plans between μ and ν even if there is a unique optimal transport plan between $\mathcal{T}(\mu)$ and $\mathcal{T}(\nu)$. In Proposition 6.63, we shall see how the reduction technique is applied and also an example illustrating the non-uniqueness of Wasserstein barycenters due to the branching structure of metric trees.

Remark 6.24. Our proof of Theorem 6.22 relies on the injectivity of $T^{\vec{\mathfrak{e}}}$ on $\vec{\mathfrak{e}} \cup e$ to define γ_e such that $\eta_e = [T^{\vec{\mathfrak{e}}} \times T^{\vec{\mathfrak{e}}}]_{\#}\gamma_e$. However, the assumption that ν_e assigns no mass to the endpoints of e allows us to weaken this requirement: injectivity of $T^{\vec{\mathfrak{e}}}$ on $\vec{\mathfrak{e}} \cup \mathring{e}$ (where $\mathring{e}$ is the interior of e) is sufficient. This assumption, introduced in the decomposition $\nu = \sum_{e\in E} \lambda_e \,\nu_e + \sum_{v\in V} \lambda_v \,\delta_v$ to prevent multiple choices for ν_e when endpoint mass is allowed, unexpectedly also contributes to this weakening of the injectivity condition on $T^{\vec{\mathfrak{e}}}$.

According to Theorem 6.9, metric graphs are proper metric spaces, which implies that Wasserstein barycenters always exist on metric trees. By combining Theorem 6.22 with the formula of Wasserstein barycenters on the real line (Theorem 6.18), we shall prove some interesting properties of Wasserstein barycenters on metric trees.

6.3 Almost absolute continuity of barycenters

In Chapter 4, it is shown that lower Ricci curvature bounds ensure the absolute continuity of Wasserstein barycenters. However, metric trees are metric spaces with curvature $\leq k$ for all $k \leq 0$ (Proposition 6.15), and are thus usually considered to have curvature $-\infty$. The following example, inspired by [50, Example 1], shows that on metric trees, the absolute continuity of Wasserstein barycenters is no longer guaranteed.

Example 6.25. Consider the metric tree Γ with a tripod shape, which is constructed by attaching three unit intervals $[0,1]$ at the endpoint 0. Denote by ν_i for $i = 1,2,3$ the three probability measures supported in the three different edges of Γ, such that each of them is the uniform probability measure on $[\frac{1}{2}, 1]$. Then the Dirac measure δ_0 at vertex 0 is the unique barycenter of the measure $\mathbb{P} := \sum_{i=1}^{3} \frac{1}{3}\delta_{\nu_i}$. To streamline our presentation, we postpone the proof of this property to Proposition 6.59. Therefore, with respect to the canonical measure on Γ, we see that while $\mathbb{P}$ gives mass to absolute continuous measures, its barycenter is not absolutely continuous.

This section is devoted to proving that Wasserstein barycenters on metric trees are *almost absolutely continuous*, meaning that the above singularity can only occur at vertices. We start with the following lemma, which characterizes Wasserstein barycenters when they are supported in an edge. Its proof relies on the reduction technique (Theorem 6.22) introduced in the previous section.

Lemma 6.26. *Let $\Gamma = (V, E, d_l)$ be a metric graph. Fix an oriented edge $\vec{e}$ of Γ and a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$. Suppose that $\mathbb{P}$ has a barycenter $\mu_{\mathbb{P}} \in \mathcal{W}_2(\Gamma)$ that is supported in the edge $\vec{e}$ of Γ. Denote by $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ the push-forward map associated to $\vec{e}$ (Definition 6.20) and define $\mathbb{Q} := \mathcal{T}_{\#}\mathbb{P}$. Then the quantile function of $\mathcal{T}(\mu_{\mathbb{P}})$ is determined by the quantile function of $\mu_{\mathbb{Q}}$ as follows: for $t \in [0,1]$,*

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

Proof. Since $\mu_{\mathbb{P}}$ is a barycenter of $\mathbb{P}$ that is supported in the edge $\vec{e}$, Theorem 6.22 implies that

$$\begin{aligned} \inf_{\mu \in \mathcal{W}_2(\vec{e})} \int_{\mathcal{W}_2(\Gamma)} d_W(\mathcal{T}(\mu), \mathcal{T}(\nu))^2 \, \mathrm{d}\,\mathbb{P}(\nu) &= \inf_{\mu \in \mathcal{W}_2(\vec{e})} \int_{\mathcal{W}_2(\Gamma)} d_W(\mu, \nu)^2 \, \mathrm{d}\,\mathbb{P}(\nu) \\ &= \int_{\mathcal{W}_2(\Gamma)} d_W(\mu_{\mathbb{P}}, \nu)^2 \, \mathrm{d}\,\mathbb{P}(\nu) = \int_{\mathcal{W}_2(\Gamma)} d_W(\mathcal{T}(\mu_{\mathbb{P}}), \mathcal{T}(\nu))^2 \, \mathrm{d}\,\mathbb{P}(\nu). \end{aligned} \tag{6.10}$$

Since the restriction of the reduction map $T^{\vec{e}}$ to $\vec{e}$ is an isometry onto $[0, l(\vec{e})]$, $\mathcal{T}$ maps $\mathcal{W}_2(\vec{e})$ bijectively to $\mathcal{W}_2([0, l(\vec{e})])$. Applying the definition $\mathbb{Q} := \mathcal{T}_{\#}\mathbb{P}$ to (6.10), we obtain

$$\inf_{\mu \in \mathcal{W}_2([0, l(\vec{e})])} \int_{\mathcal{W}_2(\mathbb{R})} d_W(\mu, \nu)^2 \, \mathrm{d}\,\mathbb{Q}(\nu) = \int_{\mathcal{W}_2(\mathbb{R})} d_W(\mathcal{T}(\mu_{\mathbb{P}}), \nu)^2 \, \mathrm{d}\,\mathbb{Q}(\nu). \tag{6.11}$$

Denote by $\mu_{\mathbb{Q}}$ the unique Wasserstein barycenter of $\mathbb{Q}$, which satisfies $f^{-1}_{\mu_{\mathbb{Q}}} = \int_{\mathcal{W}_2(\mathbb{R})} f^{-1}_{\nu} \, \mathrm{d}\,\mathbb{Q}(\nu)$ (Theorem 6.18). To further simplify (6.11), we apply Theorem 1.37 with Fubini's theorem (c.f.

(6.7) in the proof of Theorem 6.18) and obtain

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

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

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

Hence, (6.11) is equivalent to

$$\inf_{\mu \in \mathcal{W}_2([0,l(\vec{\mathrm{e}})])} \int_0^1 [f_\mu^{-1} - f_{\mu_\mathbb{Q}}^{-1}]^2 \,\mathrm{d}\,\lambda = \int_0^1 [f_{\mathcal{T}(\mu_\mathbb{P})}^{-1} - f_{\mu_\mathbb{Q}}^{-1}]^2 \,\mathrm{d}\,\lambda. \tag{6.12}$$

By Lemma 1.33, a probability measure μ is in the space $\mathcal{W}_2([0,l(\vec{\mathrm{e}})])$ if and only the image of f_μ^{-1} is contained in the interval $[0,l(\vec{\mathrm{e}})]$. Hence, a solution $\mu \in \mathcal{W}_2([0,l(\vec{\mathrm{e}})])$ of the minimization problem in the left-hand side of (6.12) must satisfy the requirements

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

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

Lemma 6.26 implies that Wasserstein barycenters on metric trees can be fully reduced to the real line, provided that they are supported in the interior of an egde.

Corollary 6.27. *Let $\Gamma = (V,E,d_l)$ be a metric graph. Fix an oriented edge $\vec{\mathrm{e}}$ of Γ and a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$. Suppose that $\mathbb{P}$ has a barycenter $\mu_\mathbb{P} \in \mathcal{W}_2(\Gamma)$ that is supported in the edge $\vec{\mathrm{e}}$ of Γ and assigns no mass to the ends of $\vec{\mathrm{e}}$. Denote by $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ the push-forward map associated to $\vec{\mathrm{e}}$ (Definition 6.20). Then $\mathcal{T}(\mu_\mathbb{P})$ is the unique barycenter of $\mathbb{Q} := \mathcal{T}_\#\mathbb{P}$.*

Proof. Since $\mu_\mathbb{P}$ is assumed to give no mass to the ends of $\vec{\mathrm{e}}$, $\mathcal{T}(\mu_\mathbb{P})$ gives no mass to the endpoints of $[0,l(\vec{\mathrm{e}})]$. Hence, Lemma 1.34 implies

$$0 = f_{\mathcal{T}(\mu_\mathbb{P})}(0) = \inf_{t\in(0,1)} \{t \mid f_{\mathcal{T}(\mu_\mathbb{P})}^{-1}(t) > 0\}, \quad 1 = \lim_{s\uparrow l(\vec{\mathrm{e}})} f_{\mathcal{T}(\mu_\mathbb{P})}(s) = \lim_{s\uparrow l(\vec{\mathrm{e}})} \inf_{t\in(0,1)} \{t \mid f_{\mathcal{T}(\mu_\mathbb{P})}^{-1}(t) > s\}.$$

It follows from these two equalities that for $0 < t < 1$, $f_{\mathcal{T}(\mu_\mathbb{P})}^{-1}(t) \neq 0$ and $f_{\mathcal{T}(\mu_\mathbb{P})}^{-1}(t) \neq l(\vec{\mathrm{e}})$. According to the requirements satisfied by $\mathcal{T}(\mu_\mathbb{P})$ in Lemma 6.26, we must have $0 \le f_{\mu_\mathbb{Q}}^{-1}(t) \le l(\vec{\mathrm{e}})$ and $f_{\mathcal{T}(\mu_\mathbb{P})}^{-1}(t) = f_{\mu_\mathbb{Q}}^{-1}(t)$ for $t \in (0,1)$. Hence, $\mathcal{T}(\mu_\mathbb{P}) = \mu_\mathbb{Q}$ is the unique barycenter of $\mathbb{Q}$. □

At first glance, the assumption in Lemma 6.26 (and thus in Corollary 6.27), that $\mathbb{P}$ has a barycenter $\mu_{\mathbb{P}} \in \mathcal{W}_2(\Gamma)$ supported in the edge $\vec{e}$, might appear limited. However, this assumption becomes useful when combined with the restriction property of Wasserstein barycenters (Proposition 5.2). The proof below, demonstrating the almost absolute continuity of Wasserstein barycenters, provides a concrete example of this idea.

Theorem 6.28 (Almost absolute continuity of Wasserstein barycenters on metric trees)**.** *Let $\Gamma = (V, E, d_l)$ be a metric graph. Fix a measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$ that gives mass to absolutely continuous measures on Γ. If $\mu_{\mathbb{P}}$ is a barycenter of $\mathbb{P}$, then the restriction of $\mu_{\mathbb{P}}$ to the interior of any given edge is absolutely continuous. Therefore, if $\mu_{\mathbb{P}}$ is not absolutely continuous, then its singular part is a sum of Dirac measures at the vertices of Γ.*

Proof. Fix an oriented edge $\vec{e}$ of Γ. Denote by $\mathring{e}$ the interior of $\vec{e}$. If μ gives no mass to the set $\mathring{e}$, then its restriction on $\mathring{e}$ is null, and thus absolutely continuous.

Consider now the case that $\mu_{\mathbb{P}}(\mathring{e}) > 0$ and denote by $\mu_{\mathring{e}} \in \mathcal{W}_2(\Gamma)$ the normalized probability measures of the restriction of $\mu_{\mathbb{P}}$ to $\mathring{e}$. Let $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ be the push-forward map associated to $\vec{e}$ (Definition 6.20). According to Lemma 6.21, for $\eta \in \mathcal{W}_2(\Gamma)$, $\mathcal{T}(\eta)$ is absolutely continuous if and only if η is absolutely continuous. We prove the absolute continuity of $\mu_{\mathring{e}}$ by discussing two different cases.

If $\mu_{\mathring{e}} = \mu_{\mathbb{P}}$, then $\mathcal{T}(\mu_{\mathring{e}})$ is the unique barycenter of $\mathbb{Q} := \mathcal{T}_{\#}\mathbb{P}$ according to Corollary 6.27, which is absolutely continuous since $\mathbb{Q}$ gives mass to absolutely continuous measures on $\mathbb{R}$ (Theorem 4.5). It follows that $\mu_{\mathbb{P}}$ is absolutely continuous when $\mu_{\mathring{e}} = \mu_{\mathbb{P}}$. We now prove that $\mu_{\mathring{e}}$ is absolutely continuous when $\mu_{\mathring{e}} \neq \mu_{\mathbb{P}}$. For the division $\mu_{\mathbb{P}} = \lambda\, \mu_{\mathring{e}} + (1-\lambda)\nu$ with $\lambda := \mu_{\mathbb{P}}(\mathring{e}) \in (0,1)$ and $\nu \in \mathcal{W}_2(\Gamma)$, Corollary 5.4 provides two measures $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$ such that $\mu_{\mathring{e}}$ is a barycenter of $\mathbb{P}_1$ and ν is a barycenter of $\mathbb{P}_2$. Moreover, Corollary 5.4 implies that both $\mathbb{P}_1$ and $\mathbb{P}_2$ give mass to absolutely continuous measures since $\mathbb{P}$ does so. We then have $\mu_{\mathring{e}} = \mu_{\mathbb{P}_1}$ is absolutely continuous as proven in the previous case.

Since the directed edge $\vec{e}$ is arbitrarily chosen, our theorem is proven. □

6.4 New results of Wasserstein barycenters on $\mathbb{R}$

Theorem 6.28 shows how properties of barycenters supported in an edge can be extended, via the restriction property of Wasserstein barycenters, to general barycenters (with appropriate modifications) on metric trees. This motivates our study of Wasserstein barycenters on the real line with compact support, since their properties can then be translated to barycenters on metric trees that assign full mass to the interior of some edge (Corollary 6.27). Our investigation proceeds as follows: given a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$ with its unique barycenter possessing certain properties, we aim to identify necessary properties satisfied by $\mathbb{P}$-almost every measure. The following proposition concerning compact support illustrates this paradigm.

Proposition 6.29. *If the unique barycenter $\mu_{\mathbb{P}}$ of $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$ has compact support, then for $\mathbb{P}$-almost every measure $\nu \in \mathcal{W}_2(\mathbb{R})$, ν has compact support. Moreover, if $\mu_{\mathbb{P}}$ is a Dirac measure, then for $\mathbb{P}$-almost every measure $\nu \in \mathcal{W}_2(\mathbb{R})$, ν is a Dirac measure.*

Proof. Lemma 1.33 shows that a probability measure μ on $\mathbb{R}$ has compact support if and only if its quantile function is finite on the unit interval $[0,1]$. By the formula of Wasserstein barycenters on

$\mathbb{R}$ (Theorem 6.18), we have

$$f_{\mu_{\mathbb{P}}}^{-1}(0) = \int_{\mathcal{W}_2(\mathbb{R})} f_\nu^{-1}(0)\,\mathrm{d}\,\mathbb{P}(\nu), \quad f_{\mu_{\mathbb{P}}}^{-1}(1) = \int_{\mathcal{W}_2(\mathbb{R})} f_\nu^{-1}(1)\,\mathrm{d}\,\mathbb{P}(\nu).$$

We remind the reader that the existence of the above integrals is part of the conclusions of Theorem 6.18. Since both $f_{\mu_{\mathbb{P}}}^{-1}(0)$ and $f_{\mu_{\mathbb{P}}}^{-1}(1)$ are finite by our assumption, the above equalities imply that for $\mathbb{P}$-almost every measure $\nu \in \mathcal{W}_2(\mathbb{R})$), both $f_\nu^{-1}(0)$ and $f_\nu^{-1}(1)$ are finite, which further implies that ν has compact support.

Now consider the special case that $\mu_{\mathbb{P}} = \delta_x$ is a Dirac measure. Since $f_{\mu_{\mathbb{P}}}^{-1}$ is the constant function on $[0,1]$ with value x, Theorem 6.18 implies

$$0 \leq \int_{\mathcal{W}_2(\mathbb{R})} f_\nu^{-1}(1) - f_\nu^{-1}(0)\,\mathrm{d}\,\mathbb{P}(\nu) = f_{\mu_{\mathbb{P}}}^{-1}(1) - f_{\mu_{\mathbb{P}}}^{-1}(0) = x - x = 0.$$

Hence, for $\mathbb{P}$-almost every ν, $f_\nu^{-1}(1) = f_\nu^{-1}(0)$, and the last part of our proposition follows from Lemma 1.33. $\square$

The preceding proof relies on the key idea of expressing the properties of a probability measure through its quantile function. To further develop this idea, we introduce dual measures in the following subsection.

6.4.1 Dual measures

Lemma 1.30 states that a quantile function is uniquely determined by its values on the open interval $(0,1)$, where it is also right-continuous. This property is shared by the distribution function of a probability measure on $[0,1]$. Based on this, we define dual measures as follows.

Definition 6.30 (Dual measures)**.** Let $\mu, \widetilde{\mu}$ be two probability measures on the real line that are supported in the unit interval $[0,1]$. Denote by f_μ^{-1} the quantile function of μ and by $f_{\widetilde{\mu}}$ the distribution function of $\widetilde{\mu}$. The measure $\widetilde{\mu}$ is the dual measure of μ if

$$f_{\widetilde{\mu}}(t) = f_\mu^{-1}(t), \quad \forall\, 0 < t < 1.$$

Our definition of dual measures is justified by the following lemma.

Lemma 6.31. *Let μ be a probability measure supported in the unit interval $[0,1]$. Its dual measure $\widetilde{\mu}$ always exists and is unique. Moreover, μ is the dual measure of $\widetilde{\mu}$, i.e., $\widetilde{\widetilde{\mu}} = \mu$.*

Proof. Consider the following function $f : \mathbb{R} \to [0,1]$ defined by setting

$$f(x) := f_\mu^{-1}(x) \text{ for } 0 \leq x < 1, \quad f(x) := 0 \text{ for } x < 0, \quad f(x) := 1 \text{ for } x \geq 1.$$

Since μ is supported in the unit interval $[0,1]$, Lemma 1.33 implies $f_\mu^{-1}(x) \in [0,1]$ for $x \in [0,1)$. It follows that the function f defined above is non-decreasing and right-continuous. Hence, there exists a unique probability measure $\widetilde{\mu}$ on $\mathbb{R}$ such that $f = f_{\widetilde{\mu}}$ is the distribution function of $\widetilde{\mu}$ [29, Proposition 4.4.3]. Given that the distribution function of a probability measure supported in $[0,1]$ is uniquely determined by its values on the open interval $(0,1)$, we conclude that f is the sole distribution function of this kind that matches f_μ^{-1} on $(0,1)$. Consequently, according to Definition 6.30, $\widetilde{\mu}$ is the unique dual measure of μ.

It remains to that μ is the dual measure of $\widetilde{\mu}$. Since $f_{\widetilde{\mu}}(t) = f_\mu^{-1}(t)$ for $0 < t < 1$, Lemma 1.34 implies that $f_\mu(x) = \inf_{t\in(0,1)}\{t \in (0,1) \mid f_{\widetilde{\mu}}(t) > x\}$ for $x \in \mathbb{R}$. We now re-write the infimum part of this formula of $f_\mu(x)$. Since $f_{\widetilde{\mu}}(t) = 1$ for $t \geq 1$, if the set $\{t \in (0,1) \mid f_{\widetilde{\mu}}(t) > x\}$ is empty for some $0 < x < 1$, then by convention (1.15),

$$\inf_{t\in(0,1)}\{t \in (0,1) \mid f_{\widetilde{\mu}}(t) > x\} = 1 = \inf_t\{t \in \mathbb{R} \mid f_{\widetilde{\mu}}(t) > x\}. \tag{6.13}$$

As $f_{\widetilde{\mu}}(t) = 0$ for $t < 0$, by considering two different cases according to whether the set $\{t \in (0,1) \mid f_{\widetilde{\mu}}(t) > x\}$ is empty, the established eqaulity (6.13) implies that, for $0 < x < 1$,

$$f_\mu(x) = \inf_{t\in(0,1)}\{t \in (0,1) \mid f_{\widetilde{\mu}}(t) > x\} = \inf_t\{t \in \mathbb{R} \mid f_{\widetilde{\mu}}(t) > x\} =: f_{\widetilde{\mu}}^{-1}(x),$$

which shows that μ is the dual measure of $\widetilde{\mu}$ according to Definition 6.30. □

Remark 6.32. In the proof of Lemma 6.31, we see that $f_{\widetilde{\mu}}$ and f_μ^{-1} coincide on $[0,1)$, with the possibility of being distinct at the point 1. For example, if the point 1 is not in the support of μ, then $f_\mu^{-1}(1) < 1$ by Lemma 1.33, while $f_{\widetilde{\mu}}(1) = \widetilde{\mu}([0,1]) = 1$ by the assumption $\widetilde{\mu} \in \mathcal{W}_2([0,1])$.

Now we are ready to investigate some basic properties of dual measures. In the following proposition, we characterize atoms of a probability measure using its dual measure. By *discrete measure*, we mean a σ-finite measure that is a weighted sum of (at most countably many) Dirac measures. For example, we can assign non-zero mass properly to the rational numbers in $[0,1]$ to construct a discrete probability measure. Note that the support of this example is the closure of all rational numbers in $[0,1]$, which is exactly the interval $[0,1]$.

Proposition 6.33 (Atoms and dual measures)**.** *Fix two probability measures $\mu, \widetilde{\mu}$ supported in $[0,1]$ such that $\widetilde{\mu}$ is the dual measure of μ. We have the following characterizations of μ,*

1. *μ is a discrete measure if and only if the support of $\widetilde{\mu}$ is negligible with respect to the Lebesgue measure on $\mathbb{R}$;*
2. *μ is atomless, i.e., $\mu(\{x\}) = 0$ for all $x \in \mathbb{R}$ if and only if $[0,1]$ is the support of $\widetilde{\mu}$;*
3. *the support of μ consists of finitely many points if and only if the support of $\widetilde{\mu}$ consists of finitely many points;*
4. *the support of μ consists of countably many points if and only if the support of $\widetilde{\mu}$ consists of countably many points.*

Proof. Recall from Lemma 1.33 that if a measure ν has compact support, then $[f_\nu^{-1}(0), f_\nu^{-1}(1)]$ is the convex hall of $\mathrm{supp}(\nu)$. Hence, the intervals $(-\infty, f_{\widetilde{\mu}}^{-1}(0))$ and $(f_{\widetilde{\mu}}^{-1}(1), +\infty)$ are the two unbounded connected components of $\mathbb{R} \setminus \mathrm{supp}(\widetilde{\mu})$. Moreover, atoms of μ are characterized by connected components of $\mathbb{R} \setminus \mathrm{supp}(\widetilde{\mu})$ as follows.

1. Since $\mu(\{0\}) = f_\mu(0) - \lim_{s\uparrow 0} f_\mu(s) = f_{\widetilde{\mu}}^{-1}(0)$, Lemma 1.33 implies that $\mu(\{0\}) = x > 0$ if and only if the interval $(-\infty, x)$ is a connected component of $\mathbb{R} \setminus \mathrm{supp}(\widetilde{\mu})$.
2. For $0 < t < 1$, since $\mu(\{t\}) = f_\mu(t) - \lim_{s\uparrow t} f_\mu(s) = f_{\widetilde{\mu}}^{-1}(t) - \lim_{s\uparrow t} f_{\widetilde{\mu}}^{-1}(s)$, Lemma 1.32 implies that $\mu(\{t\}) = x > 0$ if and only if the interval $(f_{\widetilde{\mu}}^{-1}(t) - x, f_{\widetilde{\mu}}^{-1}(t)) \subset (0,1)$ is a connected component of $\mathbb{R} \setminus \mathrm{supp}(\widetilde{\mu})$.

3. Since $\mu(\{1\}) = 1 - \lim_{s\uparrow 1} f_\mu(s) = 1 - f_{\widetilde{\mu}}^{-1}(1)$, Lemma 1.33 implies that $\mu(\{1\}) = x > 0$ if and only if the interval $(1-x, +\infty)$ is a connected component of $\mathbb{R} \setminus \mathrm{supp}(\widetilde{\mu})$.

Note that the open set $\mathbb{R}\setminus\mathrm{supp}(\widetilde{\mu})$ is a disjoint union of at most countably many intervals, which are connected components of $\mathbb{R} \setminus \mathrm{supp}(\widetilde{\mu})$. The above characterizations associate each of the intervals with an atom of μ. It is also shown that, for each interval, the length of its intersection with $(0,1)$ is equal to the jump of f_μ at the associated atom. Consequently, the sum of all jumps of f_μ, given by $\sum_{t\in[0,1]}\{f_\mu(t) - \lim_{s\uparrow t} f_\mu(s)\}$, is equal to the Lebesgue measure of the open set $(0,1)\setminus\mathrm{supp}(\widetilde{\mu})$. Recall that a probability measure μ on $\mathbb{R}$ is discrete if and only if the sum of all jumps of f_μ is 1, and it is atomless if and only if f_μ is continuous. Hence, Property 1 and Property 2 follow from the previously established equality $\sum_{t\in[0,1]}\{f_\mu(t) - \lim_{s\uparrow t} f_\mu(s)\} = 1 - \mathcal{L}^1(\mathrm{supp}(\widetilde{\mu}))$.

For Property 3, we first prove the claim that a probability measure ν on $\mathbb{R}$ is a weighted sum of finitely many Dirac measures if and only if both f_ν and f_ν^{-1} admit only finitely many values. If $\nu = \sum_{j=1}^N \lambda_j\, \delta_{x_j}$, then the image set $f_\mu([0,1])$ is contained in the set $\{0, \lambda_1, \lambda_2, \ldots, \sum_{j=1}^N \lambda_j = 1\}$, and the image set $f_\mu^{-1}([0,1])$ is contained in the set $\{x_1, x_2, \ldots, x_N\}$. Conversely, as f_μ is non-decreasing, if f_μ admits only finitely many values, then $[0,1]$ is a union of finitely intervals such that f_μ is constant on each of them, which implies that ν is a weighted sum of finitely many Dirac measures. Therefore, the claim is proven. Since f_μ^{-1} and $f_{\widetilde{\mu}}$ can only differ at 1 (Definition 6.30 and Remark 6.32), Property 3 follows from the preceding claim.

For Property 4, we assume that the support of μ consists of countably many points and prove that so does the support of $\widetilde{\mu}$. We claim that the closure of the image of f_μ, i.e., the set $\overline{f_\mu(\mathbb{R})}$, is countable. Denote by $A \subset [0,1]$ the set of discontinuity points of f_μ, which is a countable set since $f_\mu : \mathbb{R} \to [0,1]$ is a monotone function. For our claim, it suffices to prove

$$\overline{f_\mu(\mathbb{R})} = f_\mu(\mathrm{supp}(\mu)) \bigcup \{0,1\} \bigcup f_\mu(A_-), \qquad \text{where } f_\mu(A_-) := \bigcup_{x\in A} \lim_{y\uparrow x} f_\mu(y). \tag{6.14}$$

For $x \in \mathbb{R} \setminus \mathrm{supp}(\mu)$ such that $f_\mu(x) \notin \{0,1\}$, there exist two points $a, b \in \mathrm{supp}(\mu)$ such that $x \in (a,b) \subset \mathbb{R} \setminus \mathrm{supp}(\mu)$ since $\mathrm{supp}(\mu)$ is a closed set. It follows from the right-continuity of f_μ that $f(x) = f(a)$, which implies $f_\mu(\mathbb{R}) \subset f_\mu(\mathrm{supp}(\mu)) \cup \{0,1\}$. Therefore, for a fixed point $t \in \overline{f_\mu(\mathbb{R})} \setminus f_\mu(\mathrm{supp}(\mu))$ in the open interval $(0,1)$, there exists a sequence $\{x_i\}_{i\in\mathbb{N}} \subset \mathrm{supp}(\mu)$ such that $t = \lim_{i\to+\infty} f(x_i)$. By passing to a subsequence, we can assume without loss of generality that $\{x_i\}_{i\in\mathbb{N}}$ is monotone and $\lim_{i\to+\infty} x_i =: x \in \mathrm{supp}(\mu)$. Since $t \neq f(x) \in f(\mathrm{supp}(\mu))$ and f_μ is right-continuous, the sequence $\{x_i\}$ must be non-decreasing. Hence, $t = \lim_{i\to+\infty} f(x_i) = \lim_{y\uparrow x} f(x) \in f_\mu(A_-)$, which proves (6.14) by our previous choice of t. Our claim is thus proven, which implies that the closure set $\overline{f_{\widetilde{\mu}}^{-1}([0,1])}$ is countable. By Lemma 1.32 and Lemma 1.33, $\widetilde{\mu}$ gives no mass to the complement of $\overline{f_{\widetilde{\mu}}^{-1}([0,1])}$. It follows that $\mathrm{supp}(\widetilde{\mu})$ consists of at most countably many points. By Property 3, $\mathrm{supp}(\widetilde{\mu})$ is necessarily a countable set since $\mathrm{supp}(\mu)$ is not a finite set. Property 4 follows from the duality $\mu = \widetilde{\widetilde{\mu}}$. □

Recall that the Cantor measure $\mathfrak{c}$ on $[0,1]$ is an atomless probability measure whose support is the Cantor set [29, Example 2.1.10, Exercies 7 of §2.1]. Hence, its dual measure $\widetilde{\mathfrak{c}}$ is a discrete measure with support $[0,1]$ according to Proposition 6.33. This result might initially appear counter-intuitive, as one might expect that for a discrete probability measure μ, both its distribution function f_μ and its quantile function f_μ^{-1} would admit at most countably many values. This expectation seems to contradict the fact that $f_{\widetilde{\mathfrak{c}}}^{-1}$ coincides the distribution function $f_{\mathfrak{c}}$ of $\mathfrak{c}$ on $(0,1)$. However, a closer look at the proof of Proposition 6.33 reveals that f_μ admits countably many values if and only if the support of μ consists of countably many points. To further clarify the subtle distinction between discrete measures and measures with countable support, we present the following example.

Example 6.34 (A discrete measure whose distribution function admits uncountably many values)**.** Denote by $A \subset (0,1)$ the set of all rational numbers between 0 and 1 of the form $(2k-1)/2^m$, where $k, m \in \mathbb{N}^*$ are strictly positive integers such that $k \leq 2^{m-1}$. We define a discrete probability measure μ by setting $\mu(\{a\}) = 1/3^m$ for $a = (2k-1)/2^m \in A$. Note that μ assigns full mass to the set A since

$$\mu(A) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3^2} + 2^2 \cdot \frac{1}{3^3} + \cdots = \sum_{m=1}^{\infty} \frac{1}{3} (\frac{2}{3})^{m-1} = 1,$$

which implies that μ is a well-defined probability measure. We now calculate the distribution function f_μ of μ. For $x \in (0,1)$, since $y \in A \cap (0, \frac{x}{2}] \iff 2y \in A \cap (0, x]$, we have $\mu((0, \frac{x}{2}]) = \frac{1}{3}\mu((0,x])$. Hence,

$$\forall\, x \in (0,1), \quad f_\mu(\frac{1}{2}x) = \frac{1}{3} f_\mu(x), \tag{6.15}$$

which implies that $f_\mu(\frac{1}{2}) = \mu(\{\frac{1}{2}\}) + \lim_{y\uparrow 1} f_\mu(\frac{y}{2}) = \frac{1}{3} + \frac{1}{3}\lim_{y\uparrow 1} f_\mu(y) = \frac{2}{3}$. For $x \in (0,1)$ and $y \in A \cap (0, \frac{x}{2}]$, μ assigns the same mass to the point y as to the point $y + \frac{1}{2} \in A \cap (\frac{1}{2}, \frac{1+x}{2}]$. Hence, $\mu((0, \frac{x}{2}]) = \mu((\frac{1}{2}, \frac{1+x}{2}])$. It follows that, by (6.15),

$$\forall\, x \in (0,1), \quad f_\mu(\frac{1+x}{2}) = f_\mu(\frac{1}{2}) + \mu((\frac{1}{2}, \frac{1+x}{2}]) = \frac{2}{3} + f_\mu(\frac{x}{2}) = \frac{1}{3}[2 + f_\mu(x)]. \tag{6.16}$$

We now prove by mathematical induction the claim that, for $m \in \mathbb{N}^*$ and $t = \sum_{i=1}^m \frac{a_i}{2^i}$ with $a_i \in \{0,1\}$, $f_\mu(t) = \sum_{i=1}^m \frac{2\,a_i}{3^i}$. For the case $m = 1$, the equality $f_\mu(0) = 0$ is trivial and the equality $f_\mu(\frac{1}{2}) = \frac{2}{3}$ is already shown. Assume that the claim is true for $m = k$ ($k \in \mathbb{N}^*$). For $t = \sum_{i=1}^{k+1} \frac{a_i}{2^i}$, we write $t = \frac{a_1}{2} + \frac{1}{2}\sum_{i=1}^k \frac{a_{i+1}}{2^i}$. If $a_1 = 0$, then (6.15) implies that $f(t) = \frac{1}{3}\sum_{i=1}^k \frac{2a_{i+1}}{3}$. If $a_1 = 1$, then (6.16) implies that $f(t) = \frac{2}{3} + \frac{1}{3}\sum_{i=1}^k \frac{2a_{i+1}}{3}$. Therefore, the claim for $m = k+1$ is shown, which proves the claim for any $m \in \mathbb{N}^*$. Note that any real number $s \in [0,1]$ can be represented as a sum $s = \sum_{i=1}^\infty \frac{a_i}{2^i}$ with $a_i \in \{0,1\}$ such that the sequence $\{a_i\}_{i\geq 1}$ is not asymptotically identical to 1. Denote by $N(j)$ the j-th index such that $a_{N(j)} = 0$. Then, one can approximate s from above by $t_j = \sum_{i=1}^{N(j)-1} \frac{a_i}{2^i} + \frac{1}{2^{N(j)}}$ as $j \to \infty$. It follows from the right-continuity of f_μ that $f_\mu(s) = \sum_{i=1}^\infty \frac{2\,a_i}{3^i}$. Hence, $f_\mu([0,1])$ is the Cantor set, which is uncountable.

Remark 6.35. In the literature, the properties of the distribution functions of discrete measures are investigated under the name of *saltus function* or *jump function*. See references such as [40, 226B], [99, Definition 1.6.30], [11, §13.2 of Chapter 1], [54, Definition 1.1.5], [84, §7 of Chapter I], and [75, §1 of Chapter VIII].

Proposition 6.33 shows that the dual measures of various types of singular measures (with respect to the Lebesgue measure) retain singularity. To generalize this observation, we begin with some technical preparations.

For a specific class of complex measures, which will suffice for our subsequent development, we introduce the total variation norm as defined in [10, Definition 5.1.11, Definition 5.1.13].

Definition 6.36 (Total variation norm of complex measures)**.** Let μ and ν be two finite measures on $[0,1]$. The total variation norm of the *complex measure* $\mu + i\,\nu$ is

$$\|\mu + i\,\nu\|_{TV} := \sup_{\pi} \sum_{A\in\pi} |\mu(A) + i\,\nu(A)|,$$

where the supremum is taken over all possible partitions $\pi = \{A_1, A_2, \ldots, A_k\}$ of the unit interval $[0,1] = \cup_{j=1}^{k} A_j$ into finitely many pairwise disjoint measurable sets $\{A_j\}_{1\le j\le k}$, and $|x + i\,y| := \sqrt{x^2+y^2}$ denotes the modulus of the complex number $x + i\,y$ with $x, y \in \mathbb{R}$.

As shown in the following lemma [10, Theorem 5.3.5], the total variation norm can be calculated with the aid of Radon-Nikodym derivatives.

Lemma 6.37. *Let μ be a finite measure on $[0,1]$. For two given squared integrable functions $f, g \in L^2(\mu)$ with respect to μ, we have*

$$\|f\cdot\mu + i\,g\cdot\mu\|_{TV} = \int_{[0,1]} |f + i\,g| \,\mathrm{d}\,\mu = \int_{[0,1]} \sqrt{f^2+g^2}\,\mathrm{d}\,\mu.$$

To prove that a measure is singular, we shall use the following proposition showing that certain sets are negligible with respect to the Lebesgue measure, whose proof and generalizations can be found in [111, Theorem 7.29], [82, Theorem 4.1.4], [67, Corollary 3.37], [49, Corollary 6.2.2], and [89, Theorem (4.5) of Chapter IX]. For completeness, we provide a proof using Vitali covering theorem.

Proposition 6.38. *Let $f : [0,1] \to \mathbb{R}$ be a real function on $[0,1]$. Define*

$$A_f := \{x \in (0,1) \mid \textit{the derivative of } f \textit{ exists at } x \textit{ and } f'(x) = 0\}.$$

The image set $f(A_f)$ is contained in a Borel set with Lebesgue measure 0.

Proof. We first recall the following Vitali covering theorem stated for $\mathbb{R}$ [17, Theorem 5.5.1]. Let $E \subset \mathbb{R}$ be an arbitrary set. A fine covering of E is a collection $\mathcal{F}$ of compact intervals such that for every $x \in E$ and $\varepsilon > 0$, the exists an interval $I \in \mathcal{F}$ in the collection that contains x and has length less than ε. The Vitali covering theorem asserts that we can extract from any fine covering of E a sub-collection of at most countably many intervals $\mathcal{F}' = \{I_j, j \in J\} \subset \mathcal{F}$ $(J \subset \mathbb{N})$ such that $I_m \cap I_n = \emptyset$ if $m \neq n \in J$ and $\mathcal{L}^1(E \setminus \cup_{j\in J} I_j) = 0$.

Fix an $\varepsilon > 0$. For $x \in A_f$, there exists a positive number $\delta_x > 0$ such that $0 < x - \delta_x < x + \delta_x < 1$ and $|f(x+h) - f(x)| \le \varepsilon|h|$ if $|h| \le \delta_x$. Consider the following fine covering $\mathcal{F}$ of $f(A_f(E))$. To define $\mathcal{F}$, we associate x with the intervals $[f(x) - \varepsilon\,h, f(x) + \varepsilon\,h]$ for all $0 < h < \delta_x$, i.e.,

$$\mathcal{F} := \bigcup_{x\in A_f} \{[f(x) - \varepsilon\,h, f(x) + \varepsilon\,h] \subset \mathbb{R} \mid 0 < h < \delta_x\}.$$

By the Vitali covering theorem, there exists an at most countable subfamily of pairwise disjoint closed intervals, $\mathcal{F}' := \{I_j := [f(x_j) - \varepsilon\,h_j, f(x_j) + \varepsilon\,h_j],\ j \in J\} \subset \mathcal{F}$, that covers the set $f(A_f)$ up to a negligible set. For $j \in J$, define $\Delta_j := [x_j - h_j, x_j + h_j]$. We claim that for $k \neq l \in J$, the two intervals Δ_k and Δ_l are disjoint. Indeed, if $y \in \Delta_k \cap \Delta_l$, then $|f(y) - f(x_k)| \le \varepsilon\,h_k$ and $|f(y) - f(x_l)| \le \varepsilon\,h_l$, and thus

$$f(y) \in [f(x_k) - \varepsilon\,h_k, f(x_k) + \varepsilon\,h_k] \cap [f(x_l) - \varepsilon\,h_l, f(x_l) + \varepsilon\,h_l],$$

which contradicts the property that $\mathcal{F}'$ is a family of pairwise disjoint closed intervals. Therefore, our claim is proven. Since $\Delta_j \subset [0,1]$ for $j \in J$, we have $\sum_{j\in J} \lambda(\Delta_j) \leq \lambda([0,1]) = 1$, which implies

$$\mathcal{L}^*(f(A_f)) \leq \lambda(\bigcup_{j\in J} I_j) = \varepsilon \sum_{j\in J} 2\, h_j = \varepsilon \sum_{j\in J} \mathcal{L}^1(\Delta_j) \leq \varepsilon,$$

where $\mathcal{L}^*$ denotes the outer measure of the Lebesgue measure $\mathcal{L}^1$ on $\mathbb{R}$. Since $\varepsilon > 0$ is arbitrarily chosen, we have $\mathcal{L}^1(f(A_f)) = 0$, which concludes the proof. $\square$

The symmetry between a probability measure and its dual is not yet fully exploited, partially due to the discontinuity of distribution functions. To overcome this, we employ the Minty parameterization, introduced in [74, §3] and further explained in [86, B of Chapter 12], which transforms monotone (possibly multivalued) mappings into 1-Lipschitz functions. This allows us to express the symmetry between dual measures in terms of Wasserstein barycenters, which will be illustrated by figures later (Remark 6.41).

Proposition 6.39 (Symmetry between dual measures)**.** *Let μ be a probability measure supported in the unit interval $[0,1]$. Denote by $\mathfrak{u} := \mathcal{L}^1|_{[0,1]}$ the uniform probability measure on $[0,1]$ and by $\mathfrak{b}_{\widetilde{\mu}}$ the unique barycenter of $\frac{1}{2}\delta_{\mathfrak{u}} + \frac{1}{2}\delta_{\widetilde{\mu}}$. The probability measure $\mathfrak{b}_\mu := 2\mathfrak{u} - \mathfrak{b}_{\widetilde{\mu}}$ is the unique barycenter of $\frac{1}{2}\delta_{\mathfrak{u}} + \frac{1}{2}\delta_\mu$, and the distribution function $f_{\mathfrak{b}_{\widetilde{\mu}}}$ of $\mathfrak{b}_{\widetilde{\mu}}$ is the optimal transport map pushing forward $\mathfrak{b}_\mu$ to μ,*

Proof. Note that the barycenter measure $\mathfrak{b}_{\widetilde{\mu}}$ is absolutely continuous since so is the measure $\mathfrak{u}$. Since the distribution function $f_{\mathfrak{u}}$ of $\mathfrak{u}$ coincides with the identity function on $[0,1]$, we have $\widetilde{\mathfrak{u}} = \mathfrak{u}$. It follows from the formula of Wasserstein barycenters on $\mathbb{R}$ (Theorem 6.18) that $f_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}} = \frac{1}{2} f_{\mathfrak{u}} + \frac{1}{2} f_\mu$. We first show that $\mathfrak{b}_\mu := 2\mathfrak{u} - \mathfrak{b}_{\widetilde{\mu}}$ is a probability measure, which is equivalent, after passing to the distribution functions, to prove that, for any $0 \leq x < y \leq 1$,

$$f_{\mathfrak{b}_{\widetilde{\mu}}}(y) - f_{\mathfrak{b}_{\widetilde{\mu}}}(x) \leq 2y - 2x. \tag{6.17}$$

Fix $x, y \in [0,1]$ such that $x < y$. Define $B := \{t \in [0,1] \mid x < f_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}}(t) \leq y\}$. By Lemma 1.38, $\mathfrak{b}_{\widetilde{\mu}} = [f^{-1}_{\mathfrak{b}_{\widetilde{\mu}}}]_\# \mathfrak{u}$, which further implies $\mathfrak{b}_{\widetilde{\mu}} = [f_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}}]_\# \mathfrak{u}$ and thus $f_{\mathfrak{b}_{\widetilde{\mu}}}(y) - f_{\mathfrak{b}_{\widetilde{\mu}}}(x) = \mathfrak{u}(B)$. If B is empty, then (6.17) holds trivially. It remains to consider the case that B is non-empty, which implies

$$f_{\mathfrak{b}_{\widetilde{\mu}}}(y) - f_{\mathfrak{b}_{\widetilde{\mu}}}(x) = \sup B - \inf B. \tag{6.18}$$

By definition of B and the right-continuity of $f_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}} = \frac{1}{2} f_{\mathfrak{u}} + \frac{1}{2} f_\mu$,

$$x \leq f_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}}(\inf B) = \frac{1}{2} \inf B + \frac{1}{2} f_\mu(\inf B).$$

Moreover, for any $t \in B$, since f_μ is non-decreasing,

$$\frac{1}{2} t + \frac{1}{2} f_\mu(\inf B) \leq \frac{1}{2} t + \frac{1}{2} f_\mu(t) = f_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}}(t) \leq y,$$

which implies $\frac{1}{2} \sup B + \frac{1}{2} f_\mu(\inf B) \leq y$. Therefore, (6.18) implies

$$f_{\mathfrak{b}_{\widetilde{\mu}}}(y) - f_{\mathfrak{b}_{\widetilde{\mu}}}(x) = \sup B - \inf B \leq 2y - f_\mu(\inf B) - 2x + f_\mu(\inf B) \leq 2y - 2x.$$

Since x, y is arbitrarily chosen and (6.17) is proven, $\mathfrak{b}_\mu = 2\mathfrak{u} - \mathfrak{b}_{\widetilde{\mu}}$ is a probability measure. Note that $\mathfrak{b}_\mu$ is absolutely continuous since $\mathfrak{b}_{\widetilde{\mu}}$ is so.

We now show that $d_W(\mathfrak{u},\, \mathfrak{b}_\mu) = d_W(\mathfrak{b}_\mu,\, \mu)$. Since $\mathfrak{b}_{\widetilde{\mu}}$ is atomless, $[f_{\mathfrak{b}_{\widetilde{\mu}}}]_\#\mathfrak{b}_{\widetilde{\mu}} = \mathfrak{u}$ according to Lemma 1.38. As $f_{\mathfrak{b}_{\widetilde{\mu}}} = f^{-1}_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}}$ for $\mathfrak{u}$-almost everywhere, Lemma 1.38 also implies that $[f_{\mathfrak{b}_{\widetilde{\mu}}}]_\#\mathfrak{u} = [f^{-1}_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}}]_\#\mathfrak{u} = \widetilde{\mathfrak{b}_{\widetilde{\mu}}}$. Hence, $[f_{\mathfrak{b}_{\widetilde{\mu}}}]_\#(\mathfrak{b}_\mu) = [f_{\mathfrak{b}_{\widetilde{\mu}}}]_\#(2\mathfrak{u} - \mathfrak{b}_{\widetilde{\mu}}) = 2\widetilde{\mathfrak{b}_{\widetilde{\mu}}} - \mathfrak{u} = \mu$ thanks to the equality $f_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}} = \frac{1}{2}f_{\mathfrak{u}} + \frac{1}{2}f_\mu$. As $\mathfrak{b}_\mu$ is also an atomless measure and its distribution function is $2f_{\mathfrak{u}} - f_{\mathfrak{b}_{\widetilde{\mu}}}$, $[2f_{\mathfrak{u}} - f_{\mathfrak{b}_{\widetilde{\mu}}}]_\#(\mathfrak{b}_\mu) = \mathfrak{u}$ by Lemma 1.38. Since both $f_{\mathfrak{b}_{\widetilde{\mu}}}$ and $2f_{\mathfrak{u}} - f_{\mathfrak{b}_{\widetilde{\mu}}}$ are non-decreasing functions, they are optimal transport maps pushing forward the measure $\mathfrak{b}_\mu$ to measures μ and $\mathfrak{u}$ respectively. Therefore,

$$\begin{aligned} d_W(\mathfrak{u},\, \mathfrak{b}_\mu)^2 &= \int_0^1 |2f_{\mathfrak{u}}(x) - f_{\mathfrak{b}_{\widetilde{\mu}}}(x) - x|^2 \,\mathrm{d}\,\mathfrak{b}_\mu(x) = \int_0^1 |f_{\mathfrak{u}} - f_{\mathfrak{b}_{\widetilde{\mu}}}|^2 \,\mathrm{d}\,\mathfrak{b}_\mu \text{ and} \\ d_W(\mathfrak{b}_\mu,\, \mu)^2 &= \int_0^1 |x - f_{\mathfrak{b}_{\widetilde{\mu}}}(x)|^2 \,\mathrm{d}\,\mathfrak{b}_\mu(x) = \int_0^1 |f_{\mathfrak{u}} - f_{\mathfrak{b}_{\widetilde{\mu}}}|^2 \,\mathrm{d}\,\mathfrak{b}_\mu, \end{aligned}$$

which implies $d_W(\mathfrak{u},\, \mathfrak{b}_\mu) = d_W(\mathfrak{b}_\mu,\, \mu)$.

Consider the map $g := f_{\mathfrak{b}_{\widetilde{\mu}}} \circ f^{-1}_{\mathfrak{b}_\mu}$. Since g is non-decreasing and $g_\#\mathfrak{u} = [f_{\mathfrak{b}_{\widetilde{\mu}}}]_\#(\mathfrak{b}_\mu) = \mu$, g is the optimal transport map pushing forward $\mathfrak{u}$ to μ. Since $\mathfrak{b}_\mu$ is atomless, $f_{\mathfrak{b}_\mu} \circ f^{-1}_{\mathfrak{b}_\mu}$ is the identity function on $(0, 1)$ by Corollary 1.39. Hence,

$$(\mathrm{Id}, g)_\#\mathfrak{u} = (f_{\mathfrak{b}_\mu} \circ f^{-1}_{\mathfrak{b}_\mu}, f_{\mathfrak{b}_{\widetilde{\mu}}} \circ f^{-1}_{\mathfrak{b}_\mu})_\#\mathfrak{u} = (f_{\mathfrak{b}_\mu}, f_{\mathfrak{b}_{\widetilde{\mu}}})_\#(\mathfrak{b}_\mu)$$

is the optimal transport plan between $\mathfrak{u}$ and μ. Since $\mathfrak{b}_\mu = 2\mathfrak{u} - \mathfrak{b}_{\widetilde{\mu}}$, we have $f_{\mathfrak{b}_\mu} = 2f_{\mathfrak{u}} - f_{\mathfrak{b}_{\widetilde{\mu}}}$ and thus

$$d_W(\mathfrak{u}, \mu)^2 = \int_0^1 |2f_{\mathfrak{u}}(x) - f_{\mathfrak{b}_{\widetilde{\mu}}}(x) - f_{\mathfrak{b}_{\widetilde{\mu}}}(x)|^2 \,\mathrm{d}\,\mathfrak{b}_\mu(x) = 4\int_0^1 |f_{\mathfrak{u}} - f_{\mathfrak{b}_{\widetilde{\mu}}}|^2 \,\mathrm{d}\,\mathfrak{b}_\mu = 4\, d_W(\mathfrak{u}, \mathfrak{b}_\mu)^2.$$

Note that for any $\eta \in \mathcal{W}_2(\mathbb{R})$,

$$\frac{1}{2}d_W(\mathfrak{u}, \eta)^2 + \frac{1}{2}d_W(\eta, \mu)^2 \geq [\frac{1}{2}d_W(\mathfrak{u}, \eta) + \frac{1}{2}d_W(\eta, \mu)]^2 \geq d_W(\mathfrak{u}, \mu)^2,$$

which becomes an equality if $\eta = \mathfrak{b}_\mu$ since $d_W(\mathfrak{u},\, \mathfrak{b}_\mu) = d_W(\mathfrak{b}_\mu,\, \mu) = \frac{1}{2}d_W(\mathfrak{u}, \mu)$. It follows that $\mathfrak{b}_\mu$ is the unique barycenter of $\frac{1}{2}\delta_{\mathfrak{u}} + \frac{1}{2}\delta_\mu$. □

Having established the necessary background, we are now in a position to prove the following characterizations of singular measures.

Theorem 6.40. *Fix a probability measure μ supported in the unit interval $[0, 1]$. Let $\mathfrak{u} := \mathcal{L}^1|_{[0,1]}$ be the uniform probability measure on $[0, 1]$. Denote by $\widetilde{\mu}$ the dual measure of μ and by $\mathfrak{b}_{\widetilde{\mu}}$ the unique barycenter of $\frac{1}{2}\delta_{\mathfrak{u}} + \frac{1}{2}\delta_{\widetilde{\mu}}$. The following statements are equivalent.*

1. *The measure μ is singular, i.e., μ and $\mathfrak{u}$ are mutually singular.*
2. *The total variation norm of the complex measure $\mathfrak{u} + i\,\mu$ is 2.*
3. *The function $\mathcal{V}_{\mathfrak{b}_{\widetilde{\mu}}} : [0, 1] \to \mathbb{R}$ that sends $x \in [0, 1]$ to $\mathcal{V}_{\mathfrak{b}_{\widetilde{\mu}}}(x) := x - f_{\mathfrak{b}_{\widetilde{\mu}}}(x)$ is a 1-Lipschitz function and the length of its graph is $\sqrt{2}$.*

4. $\mathfrak{b}_{\widetilde{\mu}}$ *is the restriction of* $2\,\mathfrak{u}$ *on some measurable subset of* $[0,1]$ *of Lebesgue measure* $1/2$.

5. *The dual measure* $\widetilde{\mu}$ *is singular.*

Proof. The equivalence between Statement 1 and Statement 2 is proven in [10, Proposition 5.4.4 (a)]. For completeness, we repeat the part of their proof demonstrating that Statement 1 implies Statement 2. If μ and $\mathfrak{u}$ are mutually singular, then by definition [17, Definition 3.2.1] there exists a measurable set $\Omega \subset [0,1]$ such that $\mathfrak{u}(\Omega) = 1$ and $\mu([0,1] \setminus \Omega) = 1$. By refining any finite partition π of $[0,1]$ such that for $A \in \pi$, either $A \subset \Omega$ or $A \subset [0,1] \setminus \Omega$, we obtain that $\|\mathfrak{u} + i\,\mu\|_{TV} = |\mathfrak{u}(\Omega)| + |i\,\mu([0,1] \setminus \Omega)| = 1 + 1 = 2$.

For Statement 3, before proving its relations with other statements, we exhibit the following properties of $\mathfrak{b}_{\widetilde{\mu}}$ and $\mathcal{V}_{\mathfrak{b}_{\widetilde{\mu}}}$. Since $\mathfrak{b}_{\mu} := 2\mathfrak{u} - \mathfrak{b}_{\widetilde{\mu}}$ is a probability measure according to Proposition 6.39, $f_{\mathfrak{b}_{\widetilde{\mu}}}$ is a 2-Lipschitz function (c.f. (6.17)). It follows by direct calculation that $\mathcal{V}_{\mathfrak{b}_{\widetilde{\mu}}} = f_{\mathfrak{u}} - f_{\mathfrak{b}_{\widetilde{\mu}}}$ is a 1-Lipschitz function. Note that the barycenter measure $\mathfrak{b}_{\widetilde{\mu}}$ is absolutely continuous since $\mathfrak{u}$ is so. Denote by h the density function of $\mathfrak{b}_{\widetilde{\mu}}$, i.e., $\mathfrak{b}_{\widetilde{\mu}} = h \cdot \mathfrak{u}$. The length of the graph of $\mathcal{V}_{\mathfrak{b}_{\widetilde{\mu}}}$ is

$$\begin{aligned}\int_0^1 \sqrt{1 + (\mathcal{V}'_{\mathfrak{b}_{\widetilde{\mu}}})^2}\,\mathrm{d}\,\mathfrak{u} &= \int_0^1 \sqrt{1 + (1-h)^2}\,\mathrm{d}\,\mathfrak{u} = \frac{1}{\sqrt{2}} \int_0^1 \sqrt{h^2 + (2-h)^2}\,\mathrm{d}\,\mathfrak{u} \\ &= \frac{1}{\sqrt{2}} \int_0^1 |h + i(2-h)|\,\mathrm{d}\,\mathfrak{u} = \frac{1}{\sqrt{2}} \|\mathfrak{b}_{\widetilde{\mu}} + i\,\mathfrak{b}_{\mu}\|_{TV},\end{aligned}$$

where for the last equality, we applied Lemma 6.37 to the equality $\mathfrak{b}_{\widetilde{\mu}} + i\,\mathfrak{b}_{\mu} = (h + i\,(2-h)) \cdot \mathfrak{u}$. Since $\mathfrak{b}_{\widetilde{\mu}}$ is atomless, $[f_{\mathfrak{b}_{\widetilde{\mu}}}]_{\#}\mathfrak{b}_{\widetilde{\mu}} = \mathfrak{u}$ by Lemma 1.38. Moreover, Proposition 6.39 implies $[f_{\mathfrak{b}_{\widetilde{\mu}}}]_{\#}(\mathfrak{b}_{\mu}) = \mu$.

We prove that Statement 2 implies Statement 3. To avoid confusion, for a subset $A \subset [0,1]$, we denote by $[f_{\mathfrak{b}_{\widetilde{\mu}}}]^{-1}(A)$ the pre-image of A under the map $f_{\mathfrak{b}_{\widetilde{\mu}}}$, which is not necessarily the image set $f^{-1}_{\mathfrak{b}_{\widetilde{\mu}}}(A)$ of A under the map $f^{-1}_{\mathfrak{b}_{\widetilde{\mu}}}$. Recall that for any two subsets $A, B \subset [0,1]$, $[f_{\mathfrak{b}_{\widetilde{\mu}}}]^{-1}(A) \cap [f_{\mathfrak{b}_{\widetilde{\mu}}}]^{-1}(B) = [f_{\mathfrak{b}_{\widetilde{\mu}}}]^{-1}(A \cap B)$ and $[f_{\mathfrak{b}_{\widetilde{\mu}}}]^{-1}(A) \cup [f_{\mathfrak{b}_{\widetilde{\mu}}}]^{-1}(B) = [f_{\mathfrak{b}_{\widetilde{\mu}}}]^{-1}(A \cup B)$. Hence, for a given disjoint partition $\pi = \{A_1, A_2, \ldots, A_k\}$ of $[0,1]$ with finitely many measurable sets $\{A_j\}_{1 \le j \le k}$, $\pi' := \{[f_{\mathfrak{b}_{\widetilde{\mu}}}]^{-1}(A_1), [f_{\mathfrak{b}_{\widetilde{\mu}}}]^{-1}(A_2), \ldots, [f_{\mathfrak{b}_{\widetilde{\mu}}}]^{-1}(A_k)\}$ is also a disjoint partition of $[0,1]$. It follows from $[f_{\mathfrak{b}_{\widetilde{\mu}}}]_{\#}\mathfrak{b}_{\widetilde{\mu}} = \mathfrak{u}$ and $[f_{\mathfrak{b}_{\widetilde{\mu}}}]_{\#}(\mathfrak{b}_{\mu}) = \mu$ that

$$\sum_{A \in \pi} |\mathfrak{u}(A) + i\,\mu(A)| = \sum_{A' \in \pi'} |\mathfrak{b}_{\widetilde{\mu}}(A') + i\,\mathfrak{b}_{\mu}(A')|. \tag{6.19}$$

Since (6.19) holds for arbitrarily chosen partition π and $|\mathcal{V}'_{\mathfrak{b}_{\widetilde{\mu}}}| \le 1$, if Statement 2 is true, then

$$\sqrt{2} = \frac{1}{\sqrt{2}} \|\mathfrak{u} + i\,\mu\|_{TV} \le \frac{1}{\sqrt{2}} \|\mathfrak{b}_{\widetilde{\mu}} + i\,\mathfrak{b}_{\mu}\|_{TV} = \int_0^1 \sqrt{1 + (\mathcal{V}'_{\mathfrak{b}_{\widetilde{\mu}}})^2}\,\mathrm{d}\,\mathfrak{u} \le \sqrt{2},$$

which implies Statement 3.

Assuming that Statement 3 is true, we prove Statement 4 as follows. Since $\mathcal{V}_{\mathfrak{b}_{\widetilde{\mu}}}$ is a 1-Lipschitz function, Statement 3 implies that $|\mathcal{V}'_{\mathfrak{b}_{\widetilde{\mu}}}| = |1 - h| = 1$ for $\mathfrak{u}$-almost everywhere. Hence, the density function h of $\mathfrak{b}_{\widetilde{\mu}}$ satisfies that for $\mathfrak{u}$-almost every $x \in [0,1]$, $h(x)$ is either 0 or 2, which implies Statement 4.

Assume that Statement 4 is true. Consider the distribution function $f_{\mathfrak{b}_{\widetilde{\mu}}}$ and define $A_0 := \{x \in (0,1) \mid f'_{\mathfrak{b}_{\widetilde{\mu}}}(x) = 0\}$, $A_2 := \{x \in (0,1) \mid f'_{\mathfrak{b}_{\widetilde{\mu}}}(x) = 2\}$ and $\Omega := [0,1] \setminus (A_0 \cup A_2)$. Statement 4 implies that $\mathfrak{u}(A_0) = \mathfrak{u}(A_2) = \frac{1}{2}$, $\mathfrak{u}(\Omega) = 0$ and $\mathfrak{b}_{\widetilde{\mu}} = 2\,\mathfrak{u}|_{A_2}$. Consider the measure $\mathfrak{b}_{\mu} := 2\mathfrak{u} - \mathfrak{b}_{\widetilde{\mu}}$.

According to Proposition 6.39, $\mathfrak{b}_\mu$ is the barycenter of $\frac{1}{2}\delta_{\mathfrak{u}} + \frac{1}{2}\delta_\mu$. Moreover, since μ is the dual measure of $\widetilde{\mu}$ and $\mathfrak{b}_{\widetilde{\mu}} = 2\mathfrak{u} - \mathfrak{b}_\mu$, applying Proposition 6.39 again, we obtain $[f_{\mathfrak{b}_\mu}]_\# \mathfrak{b}_{\widetilde{\mu}} = \widetilde{\mu}$. Since the derivative of $f_{\mathfrak{b}_\mu} = 2 f_{\mathfrak{u}} - f_{\mathfrak{b}_{\widetilde{\mu}}}$ exists at points in A_2 with value 0, Proposition 6.38 implies that there is a (Borel) measurable set $X \subset [0,1]$ such that $f_{\mathfrak{b}_\mu}(A_2) \subset X$ and $\mathfrak{u}(X) = 0$. As $\mathfrak{b}_{\widetilde{\mu}} = 2\,\mathfrak{u}|_{A_2}$ and $A_2 \subset [f_{\mathfrak{b}_\mu}]^{-1}(X)$,

$$\widetilde{\mu}(X) = \mathfrak{b}_{\widetilde{\mu}}([f_{\mathfrak{b}_\mu}]^{-1}(X)) \geq [2\,\mathfrak{u}|_{A_2}](A_2) = 1,$$

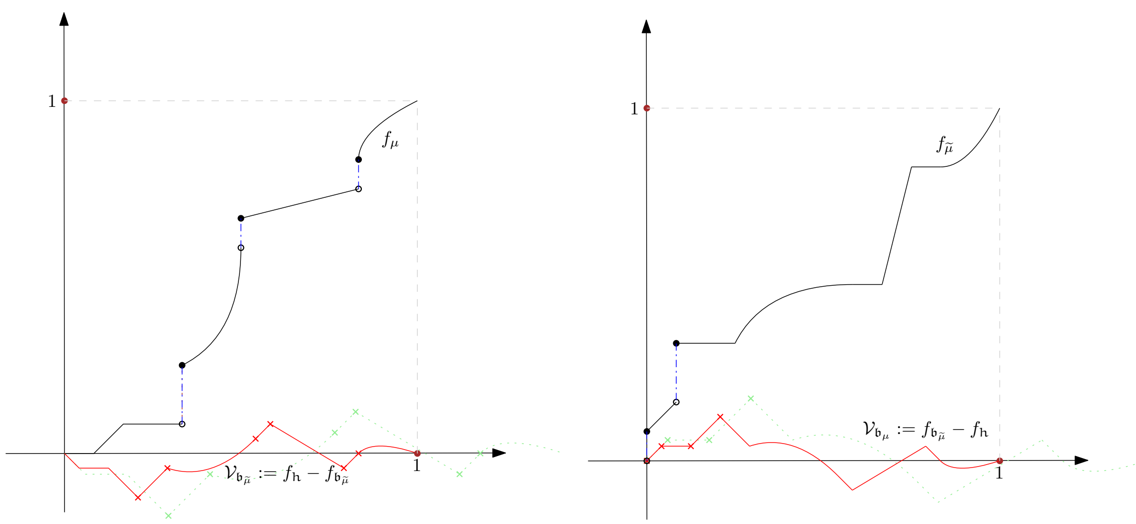
which implies $\widetilde{\mu}$ and $\mathfrak{u}$ are mutually singular, i.e., Statement 5.

We have now shown that

$$\text{Statement 1} \implies \text{Statement 2} \implies \text{Statement 3} \implies \text{Statement 4} \implies \text{Statement 5}.$$

Since μ is the dual measure of $\widetilde{\mu}$, by applying the above statements to the measure $\widetilde{\mu}$ in place of μ, it follows that Statement 5 implies Statement 1. $\square$

Remark 6.41. To explain the relation between measures μ and $\mathfrak{b}_{\widetilde{\mu}}$ in Proposition 6.39 and Theorem 6.40, we illustrate some geometric operations in Figure 6.4.



(a) Rotate and scale f_μ to obtain $\mathcal{V}_{\mathfrak{b}_{\widetilde{\mu}}}$

(b) Rotate and scale $f_{\widetilde{\mu}}$ to obtain $\mathcal{V}_{\mathfrak{b}_\mu}$

Figure 6.4: Geometric interpretations for μ and $\mathfrak{b}_{\widetilde{\mu}}$

For measure μ, we represent the graph of its distribution f_μ using the image of the complex curve $\Gamma_\mu = f_{\mathfrak{u}} + i\,f_\mu : t \in [0,1] \to t + i\,f_\mu(t)$. By rotating Γ_μ with degree $-\frac{\pi}{4}$ and then scaling it with factor $\frac{\sqrt{2}}{2}$, we obtain

$$\frac{\sqrt{2}}{2} \cdot e^{-\frac{i\,\pi}{4}} \cdot \Gamma_\mu : t \mapsto \frac{1}{2}t + \frac{1}{2}f_\mu(t) + i\,[(\frac{1}{2}t + \frac{1}{2}f_\mu(t)) - t]. \tag{6.20}$$

The measure $\mathfrak{b}_{\widetilde{\mu}}$, denoting the barycenter of $\frac{1}{2}\delta_{\mathfrak{u}} + \frac{1}{2}\delta_{\widetilde{\mu}}$, is introduced to simplify the term $\frac{1}{2}t + \frac{1}{2}f_\mu(t)$. Since $f_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}} = \frac{1}{2}t + \frac{1}{2}f_\mu(t)$, (6.20) can be equivalently written as $s \mapsto f_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}}(s) + i\,(f_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}}(s) - s)$. With the intuition that $f_{\mathfrak{b}_{\widetilde{\mu}}}$ is approximately the inverse of $f_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}}$ in mind (c.f. Corollary 1.39), we apply the "re-parameterization" (not bijective in general, and thus not rigorous) that replaces the pair $(s, f_{\widetilde{\mathfrak{b}_{\widetilde{\mu}}}}(s))$ with $(f_{\mathfrak{b}_{\widetilde{\mu}}}(t), t)$. With the new parameterization, (6.20) becomes $t \in [0,1] \mapsto t + i\,(t - f_{\mathfrak{b}_{\widetilde{\mu}}}(t))$, which also represents the graph of $\mathcal{V}_{\mathfrak{b}_{\widetilde{\mu}}} : t \to t - f_{\mathfrak{b}_{\widetilde{\mu}}}(t)$. Therefore, approximately speaking, $\mathcal{V}_{\mathfrak{b}_{\widetilde{\mu}}}$ can be obtained by rotating and scaling f_μ.

To explain the geometric meaning of Proposition 6.39, we consider the function $\mathcal{V}_{\mathfrak{b}_\mu} : t \mapsto t - f_{\mathfrak{b}_\mu}(t) = f_{\mathfrak{b}_{\widetilde{\mu}}}(t) - t$, where $\mathfrak{b}_\mu := 2\mathfrak{u} - \mathfrak{b}_{\widetilde{\mu}}$. With the preceding re-parameterization, its graph becomes

$$s \mapsto \frac{1}{2}s + \frac{1}{2}f_\mu(s) + i\,[s - (\frac{1}{2}s + \frac{1}{2}f_\mu(s))].$$

Applying the "re-parameterization" that replaces $(s, f_\mu(s))$ with $(f_{\widetilde{\mu}}(t), t)$, we obtain

$$t \mapsto \frac{1}{2}t + \frac{1}{2}f_{\widetilde{\mu}}(t) + i\,[(\frac{1}{2}t + \frac{1}{2}f_{\widetilde{\mu}}(t)) - t],$$

which is $\frac{\sqrt{2}}{2} \cdot e^{-\frac{i\pi}{4}} \cdot \Gamma_{\widetilde{\mu}}$. Therefore, $\mathfrak{b}_\mu$ is the barycenter of $\frac{1}{2}\delta_{\mathfrak{u}} + \frac{1}{2}\delta_\mu$ according to the relation between f_μ and $\mathcal{V}_{\mathfrak{b}_{\widetilde{\mu}}}$ that we deduced in the preceding paragraph. Using the rotation of angle $-\frac{\pi}{4}$, Proposition 6.39 translates the symmetry between f_μ and $f_{\widetilde{\mu}}$ with respect to the line $y = x$ into the symmetry between $\mathcal{V}_{\mathfrak{b}_{\widetilde{\mu}}}$ and $\mathcal{V}_{\mathfrak{b}_\mu}$ with respect to the x-axis.

As for Theorem 6.40, the equivalence between Statement 1 and Statement 2 relies on the following idea: the length of the plane curve, obtained from the map $t \mapsto (t, f_\mu(t))$ by connecting discontinuity points of $f_\mu(t)$ with segments, is equal to the total variation norm $\|\mathfrak{u} + i\,\mu\|_{TV}$, which can reach the maximum value 2 if and only if $\mathfrak{u}, \mu$ are mutually singular. The equivalence between Statement 2 and Statement 3 follows directly from the geometric relation between the graphs of f_μ and $\mathcal{V}_{\mathfrak{b}_{\widetilde{\mu}}}$.

6.4.2 Rigid properties

Let $\mu_{\mathbb{P}}$ be the unique barycenter of some probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$. A measure property $\mathcal{Q}$ of $\mu_{\mathbb{P}}$ is a *rigid property* of Wasserstein barycenters on $\mathbb{R}$ if $\mu_{\mathbb{P}}$ possessing the property $\mathcal{Q}$ implies that for $\mathbb{P}$-almost every ν, ν also has property $\mathcal{Q}$. For example, Proposition 6.29 shows that having compact support and being a Dirac measure are two of the rigid properties. In this subsection, we prove some rigid properties of barycenter measures related to singularity, applying the theory of dual measures.

Some results proven in the preceding subsection 6.4, though stated for dual measures, can be applied in a wider context via the Lebesgue decomposition theorem. Let us first clarify the definition of singular functions.

Definition 6.42 (Singular functions and jump functions)**.** Let $F : I \to \mathbb{R}$ be a real function defined on an interval $I \subset \mathbb{R}$. F is a singular function if its derivative exists and is equal to 0 almost everywhere (with respect to the Lebesgue measure $\mathcal{L}^1|_I$). F is a jump function if it coincides with the distribution function of a discrete measure μ on $\mathbb{R}$ up to a constant, i.e., $F - f_\mu|_I$ is a constant function on I, where μ is not necessarily a probability measure while $f_\mu(t) = \mu((-\infty, t])$.

We now state a particular case of the Lebesgue decomposition theorem for monotone functions [17, Theorem 5.4.5], which is particularly applicable to quantile functions restricted to $(0, 1)$.

Lemma 6.43. *Let $g:(0,1)\to\mathbb{R}$ be a right-continuous and non-decreasing function. It can be uniquely written as follows,*

$$g = g^{ac} + g^{sc} + g^{j}, \tag{6.21}$$

such that g^{ac}, g^{sc}, g^{sa} are three non-decreasing functions defined on $(0,1)$ satisfying

1. *g^{ac} is absolutely continuous with $g^{ac}(t) = \int_0^t g'(s)\,\mathrm{d}\,s$;*
2. *g^{sc} is singular and continuous;*
3. *g^{j} is a right-continuous jump function with $\lim_{t\downarrow 0} g^{j}(t) = 0$.*

Alternatively, we can also uniquely decompose g as the sum of two non-decreasing functions,

$$g = g^{c} + g^{j}, \tag{6.22}$$

such that g^{c} is a continuous function and g^{j} satisfies the previous requirement.

Proof. The existence of the decomposition (6.21) is explicitly constructed in [63, Corollary to Theorem 5.7.1] or [10, Theorem 5.4.1, Theorem 5.4.3]. In particular, g^{j} is defined as follows,

$$g^{j}(t) := \sum_{0<s\le t}\left[g(s) - \lim_{q\uparrow s} g(q)\right], \quad \forall\, t\in(0,1). \tag{6.23}$$

For the existence of (6.22), it suffices to set $g^{c} := g^{ac} + g^{sc}$.

As for the uniqueness of (6.21), we assume that $g = h^{ac} + h^{sc} + h^{j}$ is another decomposition satisfying the same requirements. The equality $h^{ac} = g^{ac}$ is trivial. By Definition 6.42, $h^{j} - g^{j}$ is the difference of two distribution functions of discrete measures up to a constant. Hence, being continuous, $h^{j} - g^{j} = g^{sc} - h^{sc}$ is forced to be a constant function. It follows from $\lim_{t\downarrow 0} h^{j}(t) = \lim_{t\downarrow 0} g^{j}(t) = 0$ that $h^{j} = g^{j}$, which furthers implies $h^{sc} = g^{sc}$. The uniqueness of the decomposition (6.22) can be proven similarly. □

With the help of Lemma 6.43, we characterize singular measures on $\mathbb{R}$ as follows.

Proposition 6.44 (Characterization of singular measures on $\mathbb{R}$)**.** *Let μ be a probability measure on $\mathbb{R}$. Denote by $f_\mu^{-1}|_{(0,1)} = g^{ac} + g^{sc} + g^{j} = g^{c} + g^{j}$ the decomposition of its quantile function as in Lemma 6.43. The measure μ is singular if and only if $f_\mu^{-1}|_{(0,1)}$ is a singular function, i.e., g^{ac} is a zero function. The support of μ is negligible if and only if $f_\mu^{-1}|_{(0,1)}$ is a jump function, i.e., g^{c} is a constant function.*

Proof. By Lemma 1.30, $f_\mu^{-1}|_{(0,1)}$ is a real-valued right-continuous and non-decreasing function, which allows us to apply Lemma 6.43.

To prove the proposition, we first consider the case when μ is supported in $[0,1]$. Theorem 6.40 states that μ is singular if and only if $\widetilde{\mu}$ is singular, which is also equivalent to that $f_\mu^{-1}|_{(0,1)}$ is singular by Definition 6.42. By the same arguments, Proposition 6.33 implies that μ has negligible support if and only if $f_\mu^{-1}|_{(0,1)}$ is a jump function.

We now consider the general case that μ is a probability measure on $\mathbb{R}$. Fix an interval $I := (a,b]$ such that $\mu(I) > 0$. Consider the map $Q^{I} : (a,b] \to [0,1]$ defined by $Q^{I}(x) := \frac{x-a}{b-a}$. Define the measure

$$\nu_I := \frac{1}{\mu(I)} {Q^{I}}_{\#}[\mu|_{(a,b]}],$$

which is obtained by first transforming $\mu|_{(a,b]}$ into the measure ${Q^I}_{\#}[\mu|_{(a,b]}]$ on $[0,1]$ and then normalizing it as a probability measure. Note that ν_I is singular if and only if $\mu|_{(a,b]}$ is so; and ν_I has negligible support if and only if $\mu|_{(a,b]}$ does so. By definition of ν_I, for $x \in [0,1]$,

$$f_{\nu_I}(x) = \frac{1}{\mu(I)}\left[f_\mu(x(b-a)+a) - f_\mu(a)\right].$$

According to the definition $f_{\nu_I}^{-1}(t) := \inf_x\{x \in \mathbb{R} \mid f_{\nu_I}(x) > t\}$ for $t \in (0,1)$, we have

$$f_{\nu_I}^{-1}(t) = \frac{1}{b-a}\left[f_\mu^{-1}\left(t\mu(I) + f_\mu(a)\right) - a\right], \quad \forall\, t \in (0,1).$$

In particular, $f_{\nu_I}^{-1}$ is singular if and only if $f_\mu^{-1}|_{(f_\mu(a),f_\mu(b)]}$ is singular; $f_{\nu_I}^{-1}$ is a jump function if and only if $f_\mu^{-1}|_{(f_\mu(a),f_\mu(b)]}$ is a jump function. Since ν_I is a probability measure supported in $[0,1]$, it follows from the previously proven case that $\mu|_{(a,b]}$ is singular if and only if $f_\mu^{-1}|_{(f_\mu(a),f_\mu(b)]}$ is singular; the support of $\mu|_{(a,b]}$ is negligible if and only if $f_\mu^{-1}|_{(f_\mu(a),f_\mu(b)]}$ is a jump function.

Since the interval $I = (a,b]$ satisfying $\mu(I) > 0$ is arbitrarily chosen, our proposition is proven after choosing a collection of such intervals covering the support of μ. □

Measurability related to quantile functions

Giving $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$, Proposition 6.44 inspires us to analyze the singularity of barycenters $\mu_{\mathbb{P}}$ and measures $\nu \in \operatorname{supp}(\mathbb{P})$ via the decomposition of quantile functions. For example, with the decomposition $f_\nu^{-1}|_{(0,1)} = g_\nu^c + g_\nu^j$ given by (6.22), it is natural to deduce for $t \in (0,1)$ that

$$f_{\mu_{\mathbb{P}}}^{-1}(t) = \int_{\mathcal{W}_2(\mathbb{R})} f_\nu^{-1}(t)\,\mathrm{d}\,\mathbb{P}(\nu) = \int_{\mathcal{W}_2(\mathbb{R})} g_\nu^c(t)\,\mathrm{d}\,\mathbb{P}(\nu) + \int_{\mathcal{W}_2(\mathbb{R})} g_\nu^j(t)\,\mathrm{d}\,\mathbb{P}(\nu). \tag{6.24}$$

However, to rigorously justify (6.24), we must show that the function $\nu \mapsto g_\nu^c(t)$ is measurable so that its integral against $\mathbb{P}$ is well-defined. In this subsection, we shall prove some measurability properties related to quantile functions.

While the measurability of $\nu \mapsto f_\nu^{-1}(t)$ is already proven in Lemma 1.35, the measurability of $\nu \mapsto g_\nu^c(t)$ is still non-trivial. In the following proposition, we use notation from the domain of stochastic processes, since its proof is extracted from the related literature. We refer to[28, Theorem 8.1.23], [33, Theorem 3], [48, Theorem 3.42 of Chapter III] or [53, Theorem 2.1.37] for the standard statement of this proposition, which is proved for adapted stochastic processes with finite variation. For simplicity, some technical details are left out to the classic reference [51].

Proposition 6.45. *Let $(\Omega, \mathcal{F})$ be a measurable space. For each $\omega \in \Omega$, we associate it with a non-decreasing and right-continuous function $g_\omega : (0,1) \to \mathbb{R}$. Denote by $g_\omega = g_\omega^c + g_\omega^j$ the decomposition of g_ω as in Lemma 6.43. For $t \in (0,1)$, we define two functions $X_t : (\Omega,\mathcal{F}) \to \mathbb{R}$ and $Y_t : (\Omega,\mathcal{F}) \to \mathbb{R}$ by setting*

$$X_t(\omega) := g_w(t) \quad \textit{and} \quad Y_t(\omega) := g_w^j(t).$$

If X_t is $\mathcal{F}$-measurable for all $t \in (0,1)$, then Y_t is $\mathcal{F}$-measurable for all $t \in (0,1)$.

Proof. We say a function defined on $(\Omega,\mathcal{F})$ is measurable if it is $\mathcal{F}$-measurable. To fit our proposition correctly into the settings of stochastic processes, we choose an arbitrary interval $[a,b)$ satisfying $0 < a < b < 1$ and re-define

$$X_t := X_a,\ Y_t := Y_a \text{ for } t \in [0,a] \quad \text{and} \quad X_s := X_b,\ Y_s := Y_b \text{ for } s \in [b,+\infty),$$

which allows us to regard X_t and Y_t as functions on $(\Omega, \mathcal{F})$ indexed by $\mathbb{R}_+$. Moreover, we consider the constant filtration $\{\mathcal{F}_t\}_{t\geq 0}$ with $\mathcal{F}_t := \mathcal{F}$. In the following arguments, by deducing the measurability of $\{Y_t\}_{t>0}$ from the measurability of $\{X_t\}_{t\geq 0}$, our proposition is also proven since the preceding interval $[a, b)$ is arbitrarily chosen.

For $t > 0$, define functions $X_{t-} : (\Omega, \mathcal{F}) \to \mathbb{R}$ and $\Delta X_t : (\Omega, \mathcal{F}) \to \mathbb{R}$ by setting

$$X_{t-}(\omega) := \lim_{s\uparrow t} X_s(\omega) \quad \text{ and } \quad \Delta X_t(w) := X_t(\omega) - X_{t-}(\omega).$$

Moreover, we further define $X_{0-} := X_0$ and $\Delta X_0 := X_0 - X_{0-} = 0$. According to the definition (6.23) of g_w^j, $Y_t = \sum_{0<s\leq t} \Delta X_s$ for any $t > 0$.

Fix $t > 0$. The measurability of Y_t is proven by constructing a sequence of measurable functions $S_n : (\Omega, \mathcal{F}) \to (0, 1)$, $n \in \mathbb{N}^*$, such that their graphs $\mathrm{graph}(S_n) := \{(\omega, s) \in \Omega \times (0, 1) \mid s = S_n(\omega)\}$ are pairwise disjoint and cover the set $\{\Delta X \neq 0\}$, i.e.,

$$\{(\omega, s) \in \Omega \times (0, 1) \mid \Delta X_s(w) \neq 0\} \subset \bigcup_{n\geq 1} \mathrm{graph}(S_n). \tag{6.25}$$

The explicit construction of $\{S_n\}_{n\in\mathbb{N}^*}$ can be found in [51, Proposition 1.32 of Chapter I], and we skip it for simplicity. Since a non-zero term in the sum $\sum_{0<s\leq t} \Delta X_s(\omega)$ must be one of $\Delta X_{S_n(\omega)}(\omega)$ according to (6.25), it follows from the relation $S_i(\omega) \neq S_j(\omega)$ for $i \neq j$ that

$$\sum_{0<s\leq t} \Delta X_s(\omega) = Y_t(\omega) = \sum_{n=1}^{\infty} \Delta X_{S_n(\omega)}(\omega)\, \mathbb{1}_{\{S_n\leq t\}}(w).$$

Hence, we are left to show that $\omega \mapsto \Delta X_{S_n(\omega)}(\omega)\, \mathbb{1}_{\{S_n\leq t\}}(w)$ is measurable. In the context of stochastic process, it is equivalent to show that the stopped process ΔX^{S_n} is adapted, which holds [51, Definition 1.20, Proposition 1.21 and Corollary 1.25 of Chapter I] thanks to our assumption that functions f_ω for $w \in \Omega$, are non-decreasing and right-continuous. ☐

We also need to deal with the measurability involving total variation.

Definition 6.46 (Total variations of functions)**.** Let $[a, b] \subset \mathbb{R}$ be a compact interval and let $f : [a, b] \to \mathbb{R}$ be a function defined on it. We define the total variation of f on $[a, b]$ as

$$V_a^b(f) := \sup_{a=t_0<t_1<\cdots<t_N=b} \sum_{i=0}^{N-1} |f(t_{i+1}) - f(t_i)|, \tag{6.26}$$

where the supremum is taken over all partitions $a = t_0 < t_1 < \cdots < t_N = b$ of the interval $[a, b]$. We say that f is of bounded variation on $[a, b]$ if $V_a^b(f) < +\infty$ is finite.

Note that uncountably many partitions are compared in the supremum (6.26), and thus pose the problem of measurability when we consider the total variations of a family of functions. However, for right-continuous functions, it suffices to consider only countably many partitions, a widely used conclusion when we consider the variation of stochastic processes [51, Proposition 3.3 of Chapter I] [33, Proof of Theorem 4] [48, Proof of Theorem 3.44] [28, Remark 8.1.10]. We prove it for the case $[a, b] = [0, 1]$ in the following lemma to clarify the details.

Lemma 6.47. *Let $f : [0,1] \to \mathbb{R}$ be a right-continuous function with bounded variation. For $n \in \mathbb{N}^*$, define*

$$Q_n(f) := \sum_{k=0}^{2^n-1} |f(\frac{k+1}{2^n}) - f(\frac{k}{2^n})|.$$

Then $\lim_{n\to\infty} Q_n(f) = V_0^1(f)$.

Proof. By triangle inequality, $Q_n(f)$ is increasing in n. Since $Q_n(f) \le V_0^1(f)$ by definition (6.26), the limit $\lim_{n\to\infty} Q_n(f)$ exists and $\lim_{n\to\infty} Q_n(f) \le V_0^1(f)$. Therefore, it suffices to show that given any partitions $0 = t_0 < t_1 < \cdots < t_N = 1$ and $\epsilon > 0$, there exists $m \in \mathbb{N}^*$, such that

$$\sum_{i=0}^{N-1} |f(t_{i+1}) - f(t_i)| < Q_m(f) + \epsilon,$$

as the right-hand side is always dominated by $\lim_{n\to\infty} Q_n(f) + \epsilon$. Since f is right-continuous, we may choose m sufficiently large such that for any $i = 1, 2, \ldots, N-1$, there exists $k_i \in \mathbb{N}^*$ such that $t_i < k_i/2^m < t_{i+1}$ and $|f(k_i/2^m) - f(t_i)| < \frac{\epsilon}{2N+1}$. By further setting $k_N := 1$, we obtain from the triangle inequality that

$$\begin{aligned}
|f(t_{i+1}) - f(t_i)| &\le |f(\frac{k_{i+1}}{2^m}) - f(t_{i+1})| + |f(\frac{k_{i+1}}{2^m}) - f(\frac{k_i}{2^m})| + |f(\frac{k_i}{2^m}) - f(t_i)| \\
&= |f(\frac{k_{i+1}}{2^m}) - f(\frac{k_i}{2^m})| + \frac{2\epsilon}{2N+1}.
\end{aligned}$$

It follows that

$$\sum_{i=0}^{N-1} |f(t_{i+1}) - f(t_i)| < \sum_{i=0}^{N-1} |f(\frac{k_{i+1}}{2^m}) - f(\frac{k_i}{2^m})| + \frac{2N\epsilon}{2N+1} < Q_m(f) + \epsilon,$$

which concludes the proof. ∎

We are now ready to prove the following two rigid properties.

Barycenter measures with negligible support

In the following proposition, we prove that having negligible support is a rigid property, and Proposition 6.45 is employed to ensure the measurability for the equality (6.24).

Theorem 6.48. *Let $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$ be a probability measure on $\mathcal{W}_2(\mathbb{R})$. If the support of its barycenter $\mu_{\mathbb{P}}$ is negligible, then for $\mathbb{P}$-almost every ν, the support of ν is negligible.*

Proof. For each $\nu \in \mathcal{W}_2(\mathbb{R})$, we apply the decomposition (6.22) to its quantile function f_ν, and obtain $f_\nu^{-1}|_{(0,1)} = g_\nu^c + g_\nu^j$. For each $t \in (0,1)$, Since $\nu \mapsto f_\nu^{-1}(t)$ is measurable for each $t \in (0,1)$, Proposition 6.45 guarantees that the functions $\nu \mapsto g_\nu^j(t)$ and thus $\nu \mapsto g_\nu^c(t)$ are measurable. It follows from the barycenter formula (6.6) that

$$f_{\mu_{\mathbb{P}}}^{-1}(t) = \int_{\mathcal{W}_2(\mathbb{R})} g_\nu^c(t) \,\mathrm{d}\,\mathbb{P}(\nu) + \int_{\mathcal{W}_2(\mathbb{R})} g_\nu^j(t) \,\mathrm{d}\,\mathbb{P}(\nu), \quad \forall\, t \in (0,1). \tag{6.27}$$

We remark that in the above equality, both sides are finite thanks to Lemma 1.30. We claim that the function $F^c : (0,1) \to \mathbb{R}$ defined by $F^c(t) := \int_{\mathcal{W}_2(\mathbb{R})} g^c_\nu(t) \,\mathrm{d}\,\mathbb{P}(\nu)$ is a constant function. Since g^c_ν is continuous and non-decreasing, F^c is also continuous and non-decreasing according to the monotone convergence theorem (c.f. proof of Theorem 6.18). Consider the function $F^j := \int_{\mathcal{W}_2(\mathbb{R})} g^j_\nu(t) \,\mathrm{d}\,\mathbb{P}(\nu)$ defined on $(0,1)$. Since F^j is non-decreasing and right-continuous, (6.22) implies the decomposition $F^j = h^c + h^j$. Hence, $f^{-1}_{\mu_\mathbb{P}}|_{(0,1)} = (F^c + h^c) + h^j$, which is a valid decomposition of the form (6.22). According to Proposition 6.44, $F^c + h^c$ must be a constant function, which further implies that both F^c and h^c are constant functions since they are non-decreasing. Therefore, our claim is proven.

Since the integral of non-decreasing functions, $F^c(t) = \int_{\mathcal{W}_2(\mathbb{R})} g^c_\nu(t) \,\mathrm{d}\,\mathbb{P}(\nu)$, is constant, g^c_ν is a constant function for $\mathbb{P}$-almost every ν. Hence, our proposition follows from Proposition 6.44. ☐

Singular barycenter measure

As stated in Proposition 6.44, a probability measure on $\mathbb{R}$ is singular if and only if its quantile function is singular on $(0,1)$. Hence, we begin with a criterion for singular functions via total variation. In this subsection, the map Id refers to the identity function on $(0,1)$.

Lemma 6.49. *Let $f : (0,1) \to \mathbb{R}$ be a right-continuous and non-decreasing function. The function f is singular if and only if for any compact intervals $[a,b] \subset (0,1)$,*

$$V_a^b(f - \mathrm{Id}) = V_a^b(f + \mathrm{Id}) = f(b) - f(a) + b - a.$$

Proof. The equality $V_a^b(f + \mathrm{Id}) = f(b) - f(a) + b - a$ follows directly from Definition 6.46 since f is non-decreasing. Denote by $f = g^{ac} + g^{sc} + g^j$ the decomposition (6.21) of f. Hence, we obtain the following re-writings,

$$f + \mathrm{Id} = (g^{ac} + \mathrm{Id}) + g^{sc} + g^j \quad \text{and} \quad f - \mathrm{Id} = (g^{ac} - \mathrm{Id}) + g^{sc} + g^j,$$

whose restrictions to $[a,b]$ correspond to the decomposition of a function of bounded variations as a sum of an absolutely continuous function, a singular and continuous function and a jump function. A classic result on total variation [67, Corollary 3.90], which can be deduced from the corresponding decomposition of signed measures [10, Theorem 5.3.6, Theorem 7.5.10], implies that

$$V_a^b(f + \mathrm{Id}) = \int_a^b |f'(x) + 1| \,\mathrm{d}\,x + V_a^b(g^{sc}) + V_a^b(g^j),$$
$$V_a^b(f - \mathrm{Id}) = \int_a^b |f'(x) - 1| \,\mathrm{d}\,x + V_a^b(g^{sc}) + V_a^b(g^j).$$

Therefore, $V_a^b(f + \mathrm{Id}) = V_a^b(f - \mathrm{Id})$ is equivalent to $|f'(x) - 1| = |f'(x) + 1|$ for $\mathcal{L}^1$-almost every $x \in [a,b]$, which is further equivalent to $f'(x) = 0$ almost everywhere. ☐

One advantage of Lemma 6.49 is its compatibility with integrals as illustrated by the following lemma, allowing us to apply it with the formula of Wasserstein barycenters on $\mathbb{R}$,

Lemma 6.50. *Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. Let $f : \Omega \times (0,1) \to \mathbb{R}$ be a function such that for each $\omega \in \Omega$, the function $f_\omega : t \mapsto f(\omega, t)$ is right-continuous on $(0,1)$, and for each $t \in (0,1)$, the function $\omega \mapsto f(\omega, t)$ is $\mathcal{F}$-measurable. Then, for any sub-interval $[a,b] \subset (0,1)$,*

$$V_a^b\left(\int_\Omega f_\omega \,\mathrm{d}\,\mu(\omega)\right) \le \int_\Omega V_a^b(f_\omega) \,\mathrm{d}\,\mu(\omega). \tag{6.28}$$

Proof. The measurability of the function $\omega \mapsto V_a^b(f_\omega)$ is guaranteed by Lemma 6.47. We are left to show the inequality (6.28) for the case that the right-hand is finite. Given a partition $a = t_0 < t_1 < \cdots < t_N = b$, since

$$\begin{aligned}\sum_{i=0}^{N-1} \left| \int_\Omega f_\omega(t_{i+1}) \,\mathrm{d}\,\mu(\omega) - \int_\Omega f_\omega(t_i) \,\mathrm{d}\,\mu(\omega) \right| &\le \sum_{i=0}^{N-1} \int_\Omega |f_\omega(t_{i+1}) - f_\omega(t_i)| \,\mathrm{d}\,\mu(\omega) \\ &\le \int_\Omega V_a^b(f_\omega) \,\mathrm{d}\,\mu(\omega),\end{aligned}$$

(6.28) follows directly from Definition 6.46. □

With the above technical preparations, we are ready to show that being singular is a rigid property of Wasserstein barycenters.

Theorem 6.51. *Let $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$ be a probability measure on $\mathcal{W}_2(\mathbb{R})$. If its barycenter $\mu_\mathbb{P}$ is singular (with respect to $\mathcal{L}^1$), then for $\mathbb{P}$-almost every ν, ν is also singular.*

Proof. Proposition 6.44 reduces our task to showing that $f_\nu^{-1}|_{(0,1)}$ is singular for $\mathbb{P}$-almost every ν. Thanks to Lemma 6.49, it suffices to fix an arbitrarily chosen compact interval $[a,b] \subset (0,1)$, and then prove that

$$V_a^b(f_\nu^{-1} - \mathrm{Id}) = V_a^b(f_\nu^{-1} + \mathrm{Id}) \quad \text{for } \mathbb{P}\text{-almost every } \nu. \tag{6.29}$$

Applying Lemma 6.50 with the formula of Wasserstein barycenter on $\mathbb{R}$ (Theorem 6.18), we obtain

$$\begin{aligned} V_a^b(f_{\mu_\mathbb{P}}^{-1} - \mathrm{Id}) &= V_a^b\left(\int_{\mathcal{W}_2(\mathbb{R})} f_\nu^{-1} \,\mathrm{d}\,\mathbb{P}(\nu) - \mathrm{Id}\right) \\ &\le \int_{\mathcal{W}_2(\mathbb{R})} V_a^b(f_\nu^{-1} - \mathrm{Id}) \,\mathrm{d}\,\mathbb{P}(\nu) \le \int_{\mathcal{W}_2(\mathbb{R})} V_a^b(f_\nu^{-1}) + V_a^b(\mathrm{Id}) \,\mathrm{d}\,\mathbb{P}(\nu) \\ &= \int_{\mathcal{W}_2(\mathbb{R})} \left[f_\nu^{-1}(b) - f_\nu^{-1}(a) + b - a\right] \mathrm{d}\,\mathbb{P}(\nu) = f_{\mu_\mathbb{P}}^{-1}(b) - f_{\mu_\mathbb{P}}^{-1}(a) + b - a \\ &= V_a^b(f_{\mu_\mathbb{P}}^{-1} + \mathrm{Id}), \end{aligned} \tag{6.30}$$

where we used that quantile functions are right-continuous (for the right-continuity of $f_\nu - \mathrm{Id}$) and non-decreasing (for the calculations of total variation). Since $\mu_\mathbb{P}$ is singular, Lemma 6.49 implies $V_a^b(f_{\mu_\mathbb{P}}^{-1} - \mathrm{Id}) = V_a^b(f_{\mu_\mathbb{P}}^{-1} + \mathrm{Id})$, i.e., the inequalities (6.30) must be equalities, which proves the statement (6.29) and thus the theorem. □

6.5 Singularity at vertices

For Wasserstein barycenters on metric trees, Theorem 6.28 proves their almost absolute continuity, drawing our attention to their singularity at vertices, a feature that marks a fundamental difference from the real line $\mathbb{R}$. The aim of this subsection is to deepen our comprehension of how the distinct branching structure of metric trees shapes barycenter properties, and to illuminate the potential for extending established results from $\mathbb{R}$ to this setting. Recall that $\mathbb{R}_+ = [0, +\infty)$ and $\mathbb{R}_- = (-\infty, 0]$ are two half axes containing the origin point.

Necessary conditions for singularity at vertices

We begin with a necessary condition for barycenters to be Dirac measures on vertices.

Lemma 6.52. *Let $\Gamma = (V, E, d_l)$ be a metric tree. Fix a vertex $v \in V$ and an oriented edge $\vec{\mathfrak{e}} = \overrightarrow{\{v, w\}}$. Let $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$ be a probability measure such that $\mu_{\mathbb{P}} = \delta_v$ is a barycenter of $\mathbb{P}$. Denote by $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ the push-forward map associated to $\vec{\mathfrak{e}}$ (Definition 6.20). Then the unique barycenter $\mu_{\mathbb{Q}}$ of $\mathbb{Q} := \mathcal{T}_{\#}\mathbb{P}$ is supported in the half axis $\mathbb{R}_-$.*

Proof. Since $\mu_{\mathbb{P}}$ is supported in $\vec{\mathfrak{e}}$, Lemma 6.26 is applicable. As $f^{-1}_{\mathcal{T}(\mu_{\mathbb{P}})}$ is the constant function with value 0, Lemma 6.26 implies $f_{\mu_{\mathbb{Q}}} \le 0$, which concludes the proof by Lemma 1.33. $\square$

We can generalize Lemma 6.52 via the restriction property of Wasserstein barycenters (Proposition 5.2). Let us first prove the following property.

Proposition 6.53. *Let $\Gamma = (V, E, d_l)$ be a metric tree. Fix an oriented edge $\vec{\mathfrak{e}}$ of Γ and a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$. Denote by $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ the push-forward map associated to $\vec{\mathfrak{e}}$ (Definition 6.20). Let $F : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\Gamma)$ be a measurable map such that $F(\nu)$ is absolutely continuous with respect to ν. Then the barycenters of $\mathbb{Q}_1 := \mathcal{T}_{\#}\mathbb{P}$ and $\mathbb{Q}_2 := [\mathcal{T} \circ F]_{\#}\mathbb{P}$ satisfy*

$$\mathrm{Conv}\left(\mathrm{supp}(\mu_{\mathbb{Q}_2})\right) \subset \mathrm{Conv}\left(\mathrm{supp}(\mu_{\mathbb{Q}_1})\right),$$

where $\mathrm{Conv}(A)$ for $A \subset \mathbb{R}$ denotes the convex hull of A.

Proof. We first prove the claim that for $\nu \in \mathcal{W}_2(\Gamma)$, $\mathcal{T} \circ F(\nu)$ is absolutely continuous with respect to $\mathcal{T}(\nu)$. Recall that $\mathcal{T}$ is indeed a push-forward map. Denote by $T^{\vec{\mathfrak{e}}} : \Gamma \to \mathbb{R}$ the reduction map associated to $\vec{\mathfrak{e}}$ (Proposition 6.19). For $A \in \mathcal{B}(\mathbb{R})$, if $\mathcal{T}(\nu)(A) := \nu([T^{\vec{\mathfrak{e}}}]^{-1}(A)) = 0$, then $\mathcal{T} \circ F(\nu)(A) := F(\nu)([T^{\vec{\mathfrak{e}}}]^{-1}(A)) = 0$ since $F(\nu)$ is absolutely continuous with respect to ν. Hence, the claim is proven, which implies $\mathrm{supp}(\mathcal{T} \circ F(\nu)) \subset \mathrm{supp}(\mathcal{T}(\nu))$.

We now prove our proposition for the case that $\mu_{\mathbb{Q}_1}$ has compact support, which is reduced to the following inequalities according to Lemma 1.33,

$$f^{-1}_{\mu_{\mathbb{Q}_2}}(0) \ge f^{-1}_{\mu_{\mathbb{Q}_1}}(0) > -\infty \quad \text{and} \quad f^{-1}_{\mu_{\mathbb{Q}_2}}(1) \le f^{-1}_{\mu_{\mathbb{Q}_1}}(1) < +\infty \tag{6.31}$$

By Proposition 6.29, for $\mathbb{P}$-almost every ν, $F(\nu)$ has compact support since the barycenter of $\mathbb{Q}_1 = F_{\#}\mathbb{P}$ does so, which further implies

$$f^{-1}_{\mathcal{T}\circ F(\nu)}(0) \ge f^{-1}_{\mathcal{T}(\nu)}(0) > -\infty \quad \text{and} \quad f^{-1}_{\mathcal{T}\circ F(\nu)}(1) \le f^{-1}_{\mathcal{T}(\nu)}(1) < +\infty, \tag{6.32}$$

thanks to the inclusion $\mathrm{supp}(\mathcal{T} \circ F(\nu)) \subset \mathrm{supp}(\mathcal{T}(\nu))$ and Lemma 1.33. By Theorem 6.18, for $t \in [0, 1]$,

$$f^{-1}_{\mu_{\mathbb{Q}_2}}(t) = \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}\circ F(\nu)}(t)\, \mathrm{d}\,\mathbb{P}(\nu) \quad \text{and} \quad f^{-1}_{\mu_{\mathbb{Q}_1}}(t) = \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}(\nu)}(t)\, \mathrm{d}\,\mathbb{P}(\nu).$$

Hence, (6.31) follows from the inequalities (6.32). As for the case that $\mathrm{supp}(\mu_{\mathbb{Q}_1})$ is not compact, either we have $\mathrm{Conv}\left(\mathrm{supp}(\mu_{\mathbb{Q}_1})\right) = \mathbb{R}$ and the proposition is trivial, or it suffices to prove one inequality in (6.31) according to Lemma 1.33, which can be done via similar arguments as above. $\square$

The map $F : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\Gamma)$ in Proposition 6.53 is provided by Proposition 5.2. In the following proposition, we also demonstrate how to leverage the explicit construction of F.

Theorem 6.54. *Let $\Gamma = (V, E, d_l)$ be a metric tree. Fix a vertex $v \in V$ and an oriented edge $\vec{e} = \overrightarrow{\{v, w\}}$. Let $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$ be a probability measure such that it has a barycenter $\mu_{\mathbb{P}}$ satisfying $\mu_{\mathbb{P}}(\{v\}) > 0$. Denote by $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ the push-forward map associated to $\vec{e}$ (Definition 6.20) and define $\mathbb{Q} := \mathcal{T}_{\#}\mathbb{P}$. Then the unique barycenter $\mu_{\mathbb{Q}}$ of $\mathbb{Q}$ satisfies*

$$\mu_{\mathbb{Q}}(\mathbb{R}_-) > 0.$$

Proof. The case that $\mu_{\mathbb{P}}(\{v\}) = 1$ follows directly from Lemma 6.52. Define $\lambda := \mu_{\mathbb{P}}(\{v\})$. We are left to prove our proposition for the case $0 < \lambda < 1$. Consider the decomposition $\mu_{\mathbb{P}} = \lambda\, \mu^1 + (1-\lambda)\mu^2$ with $\mu^1 := \delta_v$ and $\mu^2 \in \mathcal{W}_2(\Gamma)$. Proposition 5.2 provides a measurable map $F : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\Gamma)$ such that $\delta_v = \mu^1$ is a barycenter of $F_{\#}\mathbb{P}$ and

$$\forall\, \nu \in \mathcal{W}_2(\Gamma), \quad \nu = \lambda\, F(\nu) + (1-\lambda)\nu^2 \quad \text{with} \quad \nu^2 \in \mathcal{W}_2(\Gamma). \tag{6.33}$$

According to Proposition 6.53, the barycenters of $\mathbb{Q} = \mathcal{T}_{\#}\mathbb{P}$ and $\mathbb{Q}' := [\mathcal{T} \circ F]_{\#}\mathbb{P}$ satisfy

$$\operatorname{Conv}\left(\operatorname{supp}(\mu_{\mathbb{Q}'})\right) \subset \operatorname{Conv}\left(\operatorname{supp}(\mu_{\mathbb{Q}})\right). \tag{6.34}$$

Since δ_v is a barycenter of $F_{\#}\mathbb{P}$, Lemma 6.52 implies that

$$\operatorname{supp}(\mu_{\mathbb{Q}'}) \subset \mathbb{R}_-.$$

We prove by contradiction that $\mu_{\mathbb{Q}}(\mathbb{R}_-) > 0$ and assume now that $\mu_{\mathbb{Q}}(\mathbb{R}_-) = 0$. Denote by $T^{\vec{e}} : \Gamma \to \mathbb{R}$ the reduction map associated to $\vec{e}$ (Proposition 6.19). Since $\operatorname{supp}(\mu_{\mathbb{Q}}) \subset \mathbb{R}_+$, (6.34) implies that $\operatorname{supp}(\mu_{\mathbb{Q}'}) \subset \mathbb{R}_- \cap \mathbb{R}_+ = \{0\}$ and thus $\mu_{\mathbb{Q}'} = \delta_0$. Moreover, according to Lemma 1.33, $\mu_{\mathbb{Q}}(\mathbb{R}_-) = 0$ and the inclusion (6.34), i.e., $\{0\} \subset \operatorname{Conv}(\operatorname{supp}(\mu_{\mathbb{Q}}))$, imply that $f^{-1}_{\mu_{\mathbb{Q}}}(0) = 0$. Since $\mu_{\mathbb{Q}'} = \delta_0$ is a barycenter of $\mathbb{Q}' = [\mathcal{T} \circ F]_{\#}\mathbb{P}$, Proposition 6.29 implies that $\mathbb{Q}'$ is supported in Dirac measures. Hence, for $\mathbb{P}$-almost every ν, $q_\nu := f^{-1}_{\mathcal{T}\circ F(\nu)}(0) \in \mathbb{R}$ is finite and $\mathcal{T} \circ F(\nu) = \delta_{q_\nu}$ is a Dirac measure, which further implies

$$\mathcal{T}(\nu)(\{q_\nu\}) = \nu\left([T^{\vec{e}}]^{-1}(q_\nu)\right) \geq \lambda\, F(\nu)\left([T^{\vec{e}}]^{-1}(q_\nu)\right) = \lambda\, \mathcal{T} \circ F(\nu)(\{q_\nu\}) = \lambda, \tag{6.35}$$

where we applied (6.33) for the above inequality. Since $f^{-1}_{\mu_{\mathbb{Q}'}}(0) = 0 = f^{-1}_{\mu_{\mathbb{Q}}}(0)$, Theorem 6.18 implies

$$\int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}\circ F(\nu)}(0)\, \mathrm{d}\,\mathbb{P}(\nu) = \int_{\mathcal{W}_2(\Gamma)} q_\nu\, \mathrm{d}\,\mathbb{P}(\nu) = 0 = \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}(\nu)}(0)\, \mathrm{d}\,\mathbb{P}(\nu). \tag{6.36}$$

According to (6.35) and Lemma 1.33, for $\mathbb{P}$-almost every ν, $f^{-1}_{\mathcal{T}(\nu)}(0) \leq q_\nu$, which further implies $f^{-1}_{\mathcal{T}(\nu)}(0) = q_\nu$ thanks to (6.36). Therefore, for $\mathbb{P}$-almost every ν, $\mathcal{T}(\nu)$ is supported in $[q_\nu, +\infty)$ with $\mathcal{T}(\nu)(\{q_\nu\}) \geq \lambda > 0$, which implies $f^{-1}_{\mathcal{T}(\nu)}(\frac{\lambda}{2}) = q_\nu$ by definition of quantile function. Hence, by Theorem 6.18,

$$0 = f^{-1}_{\mu_{\mathbb{Q}}}(0) = \int_{\mathcal{W}_2(\Gamma)} q_\nu\, \mathrm{d}\,\mathbb{P}(\nu) = \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}(\nu)}(\frac{\lambda}{2})\, \mathrm{d}\,\mathbb{P}(\nu) = f^{-1}_{\mu_{\mathbb{Q}}}(\frac{\lambda}{2}),$$

which implies $\mu_{\mathbb{Q}}(\{0\}) \geq \frac{\lambda}{2} > 0$, a contradiction to the assumption that $\mu_{\mathbb{Q}}(\mathbb{R}_-) = 0$. $\square$

Dirac measures at vertices as barycenters

The reduction technique for Wasserstein barycenter problems on metric trees might seem to offer only an edge-dependent perspective. However, by comparing the barycenter problems reduced to $\mathbb{R}$ by different push-forward maps (Definition 6.20), we gain significant insight into the original problem. The following technical but crucial proposition provides such an example, where we consider two edges with opposite orientations. To impose opposite orientations on two edges (see Figure 6.5), we label vertices via simple paths. Recall that the simple path from one given vertex to another one is unique up to transitions of its domain (c.f. proof of Lemma 6.6).

Proposition 6.55. *Let $\Gamma = (V, E, d_l)$ be a metric tree. Let $e_1 = \{v_1, w_1\}, e_2 = \{v_2, w_2\}$ be two different edges of Γ. We label the vertices of e_1, e_2 such that by restricting a simple path from w_1 to w_2, we can obtain a simple path from v_1 to v_2. Denote by $\mathcal{T}_1, \mathcal{T}_2 : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ respectively the push-forward maps associated to $\overrightarrow{\{v_1, w_1\}}$ and $\overrightarrow{\{v_2, w_2\}}$ (Definition 6.20). Fix a measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$, denote by $\mu_{\mathbb{Q}_1}, \mu_{\mathbb{Q}_2}$ respectively the unique barycenters of $\mathbb{Q}_1 := \mathcal{T}_{1\#}\mathbb{P}$ and $\mathbb{Q}_2 := \mathcal{T}_{2\#}\mathbb{P}$.*

Assume $\operatorname{supp}(\mu_{\mathbb{Q}_1}) \subset \mathbb{R}_+$ and $\operatorname{supp}(\mu_{\mathbb{Q}_2}) \subset \mathbb{R}_+$. Then $\mu_{\mathbb{Q}_1} = \mu_{\mathbb{Q}_2} = \delta_0$, and the edges e_1 and e_2 share a common vertex $v_1 = v_2$.

Proof. Thanks to the way how we label the vertices v_1, v_2, w_1, w_2, we can divide measures in the support of $\mathbb{P}$ into three groups. The first group corresponds to measures, excluding δ_{v_1}, whose images under $\mathcal{T}_1$ are supported in $\mathbb{R}_+$. The second group corresponds to measures, excluding δ_{v_2}, whose images under $\mathcal{T}_2$ are supported in $\mathbb{R}_+$. The third group collects all measures not included in the preceding two groups. In other words, we write

$$\mathbb{P} = \lambda_1 \mathbb{P}_1 + \lambda_2 \mathbb{P}_2 + \lambda_3 \mathbb{P}_3, \tag{6.37}$$

where $\mathbb{P}_1$ is supported in $A_1 := \mathcal{T}_1^{-1}[\mathcal{W}_2(\mathbb{R}^+) \setminus \delta_0]$, $\mathbb{P}_2$ is supported in $A_2 := \mathcal{T}_2^{-1}[\mathcal{W}_2(\mathbb{R}^+) \setminus \delta_0]$, $\lambda_1 := \mathbb{P}(A_1)$, $\lambda_2 := \mathbb{P}(A_2)$, and $\lambda_3 := 1 - \lambda_1 - \lambda_2$. To uniquely determine (6.37), we further require that $\mathbb{P}_1 = \delta_{w_1}$ if $\lambda_1 = 0$, $\mathbb{P}_2 = \delta_{w_2}$ if $\lambda_2 = 0$, and $\mathbb{P}_3 = \frac{1}{2}\delta_{v_1} + \frac{1}{2}\delta_{v_2}$ if $\lambda_3 = 0$. To show that A_1 and A_2 are disjoint, we note that if a measure $\mu \in \mathcal{W}_2(\Gamma)$ satisfies $\operatorname{supp}(\mathcal{T}_1(\mu)) \subset \mathbb{R}^+$ and $\operatorname{supp}(\mathcal{T}_2(\mu)) \subset \mathbb{R}^+$, then $\mu = \delta_{\nu_1} = \delta_{\nu_2}$ by our labelling of the vertices, which implies $A_1 \cap A_2 = \emptyset$. According to Lemma 1.33,

$$\begin{aligned} A_1 &:= \{\mu \in \mathcal{W}_2(\Gamma) \mid f^{-1}_{\mathcal{T}_1(\mu)}(0) \geq 0 \text{ and } f^{-1}_{\mathcal{T}_1(\mu)}(1) \neq 0\}, \\ A_2 &:= \{\mu \in \mathcal{W}_2(\Gamma) \mid f^{-1}_{\mathcal{T}_2(\mu)}(0) \geq 0 \text{ and } f^{-1}_{\mathcal{T}_2(\mu)}(1) \neq 0\}. \end{aligned}$$

In particular,

$$\mu \in \operatorname{supp}(\mathbb{P}_3) \implies f^{-1}_{\mathcal{T}_1(\mu)}(0) \leq 0 \quad \text{and} \quad f^{-1}_{\mathcal{T}_2(\mu)}(0) \leq 0. \tag{6.38}$$

Define $\mathbb{T}_1 := \mathcal{T}_{1\#}\mathbb{P}_1$ and $\mathbb{T}_2 := \mathcal{T}_{2\#}\mathbb{P}_2$. Denote by $C := d_l(v_1, v_2) \geq 0$ the distance between v_1 and v_2. We now deduce the relation between $f^{-1}_{\mathcal{T}_1(\nu)}$ and $f^{-1}_{\mathcal{T}_2(\nu)}$ for $\nu \in A_1 \cup A_2$, using the fact that $\overrightarrow{\{v_1, w_1\}}$ and $\overrightarrow{\{v_2, w_2\}}$ are pointing toward different directions of the simple path from w_1 to w_2. By definition of reduction maps, if $\nu \in A_1$ or $\nu \in A_2$, then $\mathcal{T}_1(\nu)(\{x\}) = \mathcal{T}_2(\nu)(\{-x - C\})$ for $x \in \mathbb{R}$. Hence, by Lemma 1.33, for $\nu \in A_1$,

$$\begin{aligned} f^{-1}_{\mathcal{T}_2(\nu)}(0) &= -f^{-1}_{\mathcal{T}_1(\nu)}(1) - C \quad \text{(both sides can be } -\infty) \\ f^{-1}_{\mathcal{T}_2(\nu)}(1) &= -f^{-1}_{\mathcal{T}_1(\nu)}(0) - C \quad \text{(both sides are finite as } \nu \in A_1). \end{aligned}$$

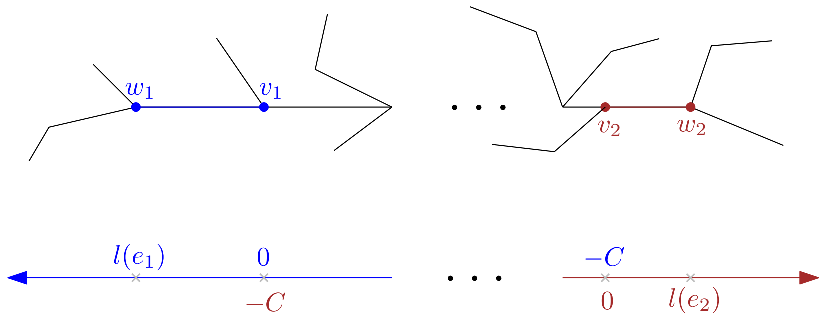


Figure 6.5: Reduction maps associated to $\overrightarrow{\{v_1, w_1\}}$ and $\overrightarrow{\{v_2, w_2\}}$

Similarly, if $\nu \in A_2$, then $f^{-1}_{\mathcal{T}_1(\nu)}(0) = -f^{-1}_{\mathcal{T}_2(\nu)}(1) - C$ (both sides can be $-\infty$) and $f^{-1}_{\mathcal{T}_1(\nu)}(1) = -f^{-1}_{\mathcal{T}_2(\nu)}(0) - C$ (both sides are finite as $\nu \in A_2$).

Applying the formula of Wasserstein barycenter on $\mathbb{R}$ (Theorem 6.18) with (6.37), we obtain

$$\begin{aligned} f^{-1}_{\mu_{\mathbb{Q}_1}}(0) &= \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}_1(\nu)}(0) \,\mathrm{d}\,\mathbb{P}(\nu) \\ &= \lambda_1 \int_{A_1} f^{-1}_{\mathcal{T}_1(\nu)}(0) \,\mathrm{d}\,\mathbb{P}_1(\nu) + \lambda_2 \int_{A_2} f^{-1}_{\mathcal{T}_1(\nu)}(0) \,\mathrm{d}\,\mathbb{P}_2(\nu) + \lambda_3 \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}_1(\nu)}(0) \,\mathrm{d}\,\mathbb{P}_3(\nu) \\ &= \lambda_1 f^{-1}_{\mu_{\mathbb{T}_1}}(0) + \lambda_2 \left[-f^{-1}_{\mu_{\mathbb{T}_2}}(1) - C \right] + \lambda_3 \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}_1(\nu)}(0) \,\mathrm{d}\,\mathbb{P}_3(\nu), \end{aligned}$$

where we used $f^{-1}_{\mathcal{T}_1(\nu)}(0) = -f^{-1}_{\mathcal{T}_2(\nu)}(1) - C$ for $\nu \in A_2$. We have $f^{-1}_{\mu_{\mathbb{Q}_1}}(0) \geq 0$ by assumption, and $\lambda_3 \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}_1(\nu)}(0) \,\mathrm{d}\,\mathbb{P}_3(\nu) \leq 0$ by (6.38). Hence,

$$\lambda_1 f^{-1}_{\mu_{\mathbb{T}_1}}(0) \geq \lambda_2 \left[f^{-1}_{\mu_{\mathbb{T}_2}}(1) + C \right]. \tag{6.39}$$

By the same arguments,

$$f^{-1}_{\mu_{\mathbb{Q}_2}}(0) = \lambda_1 \left[-f^{-1}_{\mu_{\mathbb{T}_1}}(1) - C \right] + \lambda_2 f^{-1}_{\mu_{\mathbb{T}_2}}(0) + \lambda_3 \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}_2(\nu)}(0) \,\mathrm{d}\,\mathbb{P}_3(\nu),$$

where we used $f^{-1}_{\mathcal{T}_2(\nu)}(0) = -f^{-1}_{\mathcal{T}_1(\nu)}(1) - C$ for $\nu \in A_1$. And we also have

$$\lambda_2 f^{-1}_{\mu_{\mathbb{T}_2}}(0) \geq \lambda_1 \left[f^{-1}_{\mu_{\mathbb{T}_1}}(1) + C \right]. \tag{6.40}$$

According to (6.39), the inequality $f^{-1}_{\mu_{\mathbb{T}_2}}(1) \geq f^{-1}_{\mu_{\mathbb{T}_2}}(0)$, and (6.40),

$$\lambda_1 f^{-1}_{\mu_{\mathbb{T}_1}}(0) \geq \lambda_2 \left[f^{-1}_{\mu_{\mathbb{T}_2}}(1) + C \right] \geq \lambda_2 \left[f^{-1}_{\mu_{\mathbb{T}_2}}(0) + C \right] \geq (\lambda_1 + \lambda_2) C + \lambda_1 f^{-1}_{\mu_{\mathbb{T}_1}}(1). \tag{6.41}$$

It follows from $f^{-1}_{\mu_{\mathbb{T}_1}}(1) \geq f^{-1}_{\mu_{\mathbb{T}_1}}(0)$ that $(\lambda_1 + \lambda_2)C = 0$, $\lambda_1 f^{-1}_{\mu_{\mathbb{T}_1}}(0) = \lambda_1 f^{-1}_{\mu_{\mathbb{T}_1}}(1)$, $\lambda_2 f^{-1}_{\mu_{\mathbb{T}_2}}(0) = \lambda_2 f^{-1}_{\mu_{\mathbb{T}_2}}(1)$, which further implies, by the previous expressions of $f^{-1}_{\mu_{\mathbb{Q}_1}}(0)$ and $f^{-1}_{\mu_{\mathbb{Q}_2}}(0)$,

$$f^{-1}_{\mu_{\mathbb{Q}_1}}(0) = \lambda_3 \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}_1(\nu)}(0) \,\mathrm{d}\,\mathbb{P}_3(\nu) = \lambda_3 \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}_2(\nu)}(0) \,\mathrm{d}\,\mathbb{P}_3(\nu) = f^{-1}_{\mu_{\mathbb{Q}_2}}(0) = 0. \tag{6.42}$$

We now prove that $\mu_{\mathbb{Q}_1} = \mu_{\mathbb{Q}_2} = \delta_0$. Since $f^{-1}_{\mu_{\mathbb{T}_1}} = \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}_1(\nu)} \,\mathrm{d}\,\mathbb{P}_1(\nu)$, it follows from $\lambda_1 f^{-1}_{\mu_{\mathbb{T}_1}}(0) = \lambda_1 f^{-1}_{\mu_{\mathbb{T}_1}}(1)$ that $\lambda_1 f_{\mathcal{T}_1(\nu)}$ is a constant function for $\mathbb{P}_1$-almost every ν (c.f. proof of Proposition 6.29). Thanks to the relation between $f_{\mathcal{T}_1(\nu)}$ and $f_{\mathcal{T}_2(\nu)}$ for $\nu \in A_1$, $\lambda_1 f_{\mathcal{T}_2(\nu)}$ is also a constant function for $\mathbb{P}_1$-almost every ν. Similarly, $\lambda_2 f^{-1}_{\mu_{\mathbb{T}_2}}(0) = \lambda_2 f^{-1}_{\mu_{\mathbb{T}_2}}(1)$ implies that $\lambda_2 f_{\mathcal{T}_2(\nu)}$ and $\lambda_2 f_{\mathcal{T}_1(\nu)}$ are both constant functions for $\mathbb{P}_2$-almost every ν. Let us prove the claim that $\lambda_3 f_{\mathcal{T}_1(\nu)}$ and $\lambda_3 f_{\mathcal{T}_2(\nu)}$ are both constant functions for $\mathbb{P}_3$-almost every ν. Since the case $\lambda_3 = 0$ is trivial, we are left to prove the claim for the case $\lambda_3 \neq 0$. In this case, (6.38) and (6.42) imply that for $\nu \in \operatorname{supp}(\mathbb{P}_3)$, $f^{-1}_{\mathcal{T}_1(\nu)}(0) = f^{-1}_{\mathcal{T}_2(\nu)}(0) = 0$, which further imply $f^{-1}_{\mathcal{T}_1(\nu)}(1) = f^{-1}_{\mathcal{T}_2(\nu)}(1) = 0$ as $\nu \notin A_1 \cup A_2$. Hence, the claim is proven. Therefore, according to the following equalities for $t \in [0,1]$

$$f^{-1}_{\mu_{\mathbb{Q}_1}}(t) = \lambda_1 \int_{A_1} f^{-1}_{\mathcal{T}_1(\nu)}(t) \,\mathrm{d}\,\mathbb{P}_1(\nu) + \lambda_2 \int_{A_2} f^{-1}_{\mathcal{T}_1(\nu)}(t) \,\mathrm{d}\,\mathbb{P}_2(\nu) + \lambda_3 \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}_1(\nu)}(t) \,\mathrm{d}\,\mathbb{P}_3(\nu),$$

$$f^{-1}_{\mu_{\mathbb{Q}_2}}(t) = \lambda_1 \int_{A_1} f^{-1}_{\mathcal{T}_2(\nu)}(t) \,\mathrm{d}\,\mathbb{P}_1(\nu) + \lambda_2 \int_{A_2} f^{-1}_{\mathcal{T}_2(\nu)}(t) \,\mathrm{d}\,\mathbb{P}_2(\nu) + \lambda_3 \int_{\mathcal{W}_2(\Gamma)} f^{-1}_{\mathcal{T}_2(\nu)}(t) \,\mathrm{d}\,\mathbb{P}_3(\nu),$$

both functions $f^{-1}_{\mu_{\mathbb{Q}_1}}$ and $f^{-1}_{\mu_{\mathbb{Q}_2}}$ are constant. Since $f^{-1}_{\mu_{\mathbb{Q}_1}}(0) = f^{-1}_{\mu_{\mathbb{Q}_2}}(0) = 0$ by (6.42), we have $\mu_{\mathbb{Q}_1} = \mu_{\mathbb{Q}_2} = \delta_0$.

We are left to prove that $v_1 = v_2$, i.e., $C = 0$. Since $(\lambda_1+\lambda_2)C = 0$, the case $\lambda_1+\lambda_2 > 0$ is trivial. If $\lambda_1 + \lambda_2 = 0$, then $\lambda_3 > 0$ and we have shown in the preceding paragraph that $\mathcal{T}_1(\nu) = \mathcal{T}_2(\nu) = \delta_0$ for $\mathbb{P}_3$-almost every ν. In particular, $\mathbb{P}_3 = \delta_{\delta_{v_1}} = \delta_{\delta_{v_2}}$, which concludes the proof. □

By properly choosing two oriented edges to compare their reduction maps, we prove the following powerful proposition that helps us to determine the support of a Wasserstein barycenter on metric trees.

Proposition 6.56. *Let $\Gamma = (V, E, d_l)$ be a metric tree. Fix an oriented edge $\vec{e}$ of Γ. Denote by $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ the push-forward map associated to $\vec{e}$ (Definition 6.20). Let $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$ be a probability measure. Assume that the unique barycenter $\mu_{\mathbb{Q}}$ of $\mathbb{Q} := \mathcal{T}_\#\mathbb{P}$ is supported in $\mathbb{R}_-$. Then for any barycenter $\mu_{\mathbb{P}}$ of $\mathbb{P}$, its image $\mathcal{T}(\mu_{\mathbb{P}})$ under $\mathcal{T}$ is supported in $\mathbb{R}_-$.*

Proof. We prove the proposition by contradiction. Denote by $T^{\vec{e}} : \Gamma \to \mathbb{R}$ the reduction map associated to $\vec{e}$ (Proposition 6.19). Assume that there exists a barycenter $\mu_{\mathbb{P}}$ such that $\mu_{\mathbb{P}}((0, +\infty)) > 0$. Since $(0, +\infty) = \cup_{x>0}(x, +\infty)$, there exists a real number $0 < x \leq l(\vec{e})$ such that $\mu_{\mathbb{P}}((x, +\infty)) > 0$. Denote by $v_1 := \vec{e}_{x/l(\vec{e})} \in \Gamma$ the pre-image of x under $T^{\vec{e}}$. If $x < l(\vec{e})$, then we consider the following modification of the vertices and edges of Γ. We add v_1 to V as a vertex, and replace the edge $\{\vec{e}_0, \vec{e}_1\} \in E$ with two edges $\{\vec{e}_0, v_1\}$ and $\{v_1, \vec{e}_1\}$ of lengths x and $l(\vec{e}) - x$ respectively. Such a modification does not change the metric structure of Γ. Moreover, the reduction map associated to $\overrightarrow{\{\vec{e}_0, v_1\}}$ coincides with $T^{\vec{e}}$. Therefore, by possibly replacing $\vec{e}$ with $\overrightarrow{\{\vec{e}_0, v_1\}}$, we can assume without loss of generality that $x = l(\vec{e})$ and $v_1 = \vec{e}_1$.

Since $\mu_{\mathbb{P}}((l(\vec{e}), +\infty)) > 0$, there exists an edge $e_2 = \{v_2, w_2\} \in E$ such that $l(\vec{e}) \leq T^{\vec{e}}(v_2) < T^{\vec{e}}(w_2)$ and we can write

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

where $0 < \lambda < 1$, $\mu^1 \in \mathcal{W}_2(\Gamma)$, $\mu^2 \in \mathcal{W}_2(\Gamma)$, and $\mu^1(e_2 \setminus \{v_2\}) = 1$. According to Proposition 5.2, there exists a measurable map $F : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\Gamma)$ such that μ^1 is a barycenter of $F_\#\mathbb{P}$. By Proposition 6.53, the barycenter $\mu_{\mathbb{Q}'}$ of $\mathbb{Q}' := [\mathcal{T} \circ F]_\#\mathbb{P}$ is supported in $\mathbb{R}_-$ since by assumption, the barycenter $\mu_{\mathbb{Q}}$ of $\mathbb{Q} = \mathcal{T}_\#\mathbb{P}$ is supported in $\mathbb{R}_-$.

We shall derive a contradiction using Proposition 6.55. Define $w_1 := \vec{e}_0$. Denote by $\mathcal{T}_1$ and $\mathcal{T}_2$ respectively the push-forward maps associated to $\overrightarrow{\{v_1, w_1\}}$ and $\overrightarrow{\{v_2, w_2\}}$. Since $\vec{e} = \overrightarrow{\{w_1, v_1\}}$ and the barycenter $\mu_{\mathbb{Q}'}$ of $\mathbb{Q}' = [\mathcal{T} \circ F]_{\#}\mathbb{P}$ is supported in $\mathbb{R}_-$, the barycenter $\mu_{\mathbb{Q}_1}$ of $\mathbb{Q}_1 := [\mathcal{T}_1 \circ F]_{\#}\mathbb{P}$ is supported in $[l(\vec{e}), +\infty)$. Consider the barycenter $\mu_{\mathbb{Q}_2}$ of $\mathbb{Q}_2 := [\mathcal{T}_2 \circ F]_{\#}\mathbb{P}$. Since $\mu^1(e_2 \setminus \{v_2\}) = 1$, $\mathcal{T}_2(\mu^1)((0, l(e_2)]) = 1$, which implies $f^{-1}_{\mathcal{T}_2(\mu^1)}(t) > 0$ for $t \in (0,1)$. As μ^1 is a barycenter of $F_{\#}\mathbb{P}$, Lemma 6.26 implies that $f^{-1}_{\mu_{\mathbb{Q}_2}} \geq 0$. Hence, $\mathrm{supp}(\mu_{\mathbb{Q}_2}) \subset \mathbb{R}_+$. Applying Proposition 6.55 to $\mu_{\mathbb{Q}_1}$ and $\mu_{\mathbb{Q}_2}$, we obtain that $\mu_{\mathbb{Q}_1} = \mu_{\mathbb{Q}_2} = \delta_0$, which is contradiction since $\mathrm{supp}(\mu_{\mathbb{Q}_1}) \subset [l(\vec{e}), +\infty)$. ☐

As a corollary to Lemma 6.52 and Proposition 6.56, we get a sufficient and necessary condition for a Dirac measure at some vertex to be a barycenter.

Theorem 6.57. *Let $\Gamma = (V, E, d_l)$ be a metric tree. Fix a vertex $v \in V$ and a probability measure $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$. We enumerate all the edges $e_k := \{v, w_k\} \in E$, $k = 1, 2, \ldots, n$, such that v is an end of each e_k. For each edge e_k, denote by $\mathcal{T}_k : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ the push-forward map associated to $\overrightarrow{\{v, w_k\}}$ (Definition 6.20), define $\mathbb{Q}_k := \mathcal{T}_{k\#}\mathbb{P}$, and denote by $\mu_{\mathbb{Q}_k}$ the unique barycenter of $\mathbb{Q}_k$. Then δ_v is a barycenter of $\mathbb{P}$ if and only if*

$$\mathrm{supp}(\mu_{\mathbb{Q}_k}) \subset (-\infty, 0], \quad \textit{for } k = 1, 2, \ldots, n. \tag{6.43}$$

Moreover, if (6.43) holds, then δ_v is the unique barycenter of $\mathbb{P}$.

Proof. If δ_v is a barycenter of $\mathbb{P}$, then (6.43) follows directly from Lemma 6.52.

Assuming that (6.43) holds, we prove that δ_v is the unique barycenter of $\mathbb{P}$. Let $\mu_{\mathbb{P}}$ be a barycenter of $\mathbb{P}$. Denote by T_k the reduction map associated to $\overrightarrow{\{v, w_k\}}$ (Proposition 6.19). By Proposition 6.56, (6.43) implies that

$$\mathrm{supp}(\mathcal{T}_k(\mu_{\mathbb{P}})) \subset \mathbb{R}_- \quad \text{for } k = 1, 2, \ldots, n.$$

Since $\{e_k\}_{1 \leq k \leq n}$ is the set of all edges at v,

$$\mu_{\mathbb{P}}(\Gamma \setminus \{v\}) = \mu_{\mathbb{P}}(\cup_{k=1}^{n} T_k^{-1}((0, +\infty)) \leq \sum_{k=1}^{n} \mu_{\mathbb{P}}(T_k^{-1}((0, +\infty))) = 0,$$

which implies $\mu_{\mathbb{P}} = \delta_v$. ☐

Remark 6.58. Inspired by Theorem 6.54 and Theorem 6.57, one may wonder if the condition

$$\mu_{\mathbb{Q}_k}(\mathbb{R}_-) > 0, \quad \text{for } k = 1, 2, \ldots, n, \tag{6.44}$$

implies that $\mu_{\mathbb{P}}(\{v\}) > 0$. We shall see in Proposition 6.64 a counter-example for this implication.

6.6 Summary and examples of barycenters

The preceding subsections established several key results and, in doing so, identified a systematic approach for studying Wasserstein barycenters on metric trees. This approach is built upon the reduction technique for metric trees and the restriction property of Wasserstein barycenters.

Our systematic approach for Wasserstein barycenters on metric trees

To fix the notation, let $\Gamma = (E, V, d_l)$ be a metric tree and let $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$ be a probability measure. Fix an oriented edge $\vec{e} := \overrightarrow{\{v, w\}}$ of Γ. Denote by $T^{\vec{e}} : \Gamma \to \mathbb{R}$ and $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ respectively the reduction map (Proposition 6.19) and the push-forward map (Definition 6.20) associated to $\vec{e}$. Applying this reduction technique to $\mathbb{P}$ yields the measure $\mathbb{Q} := \mathcal{T}_{\#}\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$, whose unique barycenter is denoted by $\mu_{\mathbb{Q}}$.

The first step of our approach focuses on the potential mass concentration of $\mu_{\mathbb{P}}$ at vertices, specifically, determining if $\mu_{\mathbb{P}}(\{v\}) > 0$. As detailed in Section 6.5, this involves examining the support of $\mu_{\mathbb{Q}}$ by calculating $f^{-1}_{\mu_{\mathbb{Q}}}(0)$ and $f^{-1}_{\mu_{\mathbb{Q}}}(1)$, which in turn relies solely on the supports of measures in $\mathrm{supp}(\mathbb{P})$. Key results underpinning this step include:

1. Theorem 6.54 establishes that if $\mu_{\mathbb{Q}}(\mathbb{R}_-) = 0$, then $\mu_{\mathbb{P}}(\{v\}) = 0$.

2. Conversely, if $\mu_{\mathbb{Q}}(\mathbb{R}_-) = 1$, then Proposition 6.56 excludes the edges and vertices contained in $[T^{\vec{e}}]^{-1}(\mathbb{R}_+ \setminus \{0\})$ out of $\mathrm{supp}(\mu_{\mathbb{P}})$. This same proposition (Proposition 6.56) also enables the exclusion of parts of an edge from $\mathrm{supp}(\mu_{\mathbb{P}})$ by strategically adding a new vertex to that edge, a technique demonstrated in its proof.

3. Finally, Theorem 6.57 provides a fast and intuitive criterion for δ_v to be a (and thus the unique one) barycenter of $\mathbb{P}$.

The second step shifts focus to the behavior of $\mu_{\mathbb{P}}$ on the edges of Γ, i.e., analyzing its restriction $\mu_{\mathbb{P}}|_{\vec{e}}$. If $\mu_{\mathbb{P}}$ is supported in the edge $\vec{e}$, then Lemma 6.26 illustrates how $\mu_{\mathbb{P}}$ is fully determined by $\mu_{\mathbb{Q}}$. This connection allows us to extend several properties of Wasserstein barycenters from the real line to metric trees. For instance:

1. Proposition 6.29 contributes to the understanding of the (almost) absolute continuity of barycenters on metric trees. It implies that singularities of $\mu_{\mathbb{P}}$ are confined to vertices, provided that $\mathbb{P}$ assigns positive mass to absolutely continuous measures.

2. The concept of rigid property is proposed and explored in Section 6.4.2, with particular attention to various types of singularity. After a careful examination of the properties of $\mathcal{T}$, established rigid properties of barycenters on the real line, such as being a Dirac measure (Proposition 6.29), having negligible support (Theorem 6.48), and being singular (Theorem 6.51), can be effectively translated to criteria of certain singularities of $\mu_{\mathbb{P}}$.

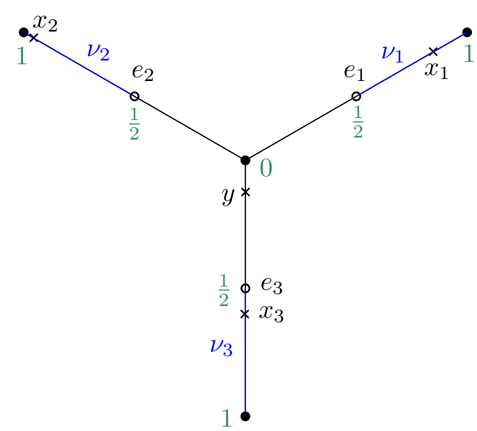
We point out a crucial aspect woven throughout our approach, which deserves special emphasis due to its subtle power. That is the *flexible application of the restriction property of Wasserstein barycenters* to gain deeper insights into $\mu_{\mathbb{P}}$. Recall that in Proposition 5.2, corresponding to a decomposition $\mu_{\mathbb{P}} = \lambda\,\mu^1 + (1-\lambda)\mu^2$, the construction of $F^i : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\Gamma)$ $(i = 1, 2)$ such that μ^i is a barycenter of $\mathbb{Q}^i := {F^i}_{\#}\mathbb{P}$ is not completely obscure. While a comprehensive understanding of optimal transport on metric trees would undoubtedly unlock the full potential of this construction, we can still derive valuable conclusions even with partial information, a strength demonstrated in the proof of Theorem 6.54.

To solidify the ideas discussed above, the remainder of this subsection presents several concrete examples. These examples have been carefully selected to illustrate the application of our approach and to showcase the distinguished properties of Wasserstein barycenters on metric trees.

Concrete examples

The following proposition presents the essential difference between the behaviors of Wasserstein barycenters on metric trees and on the real line. To achieve a comprehensive understanding of this difference, we provide two proofs with one based on direct calculations and the other one based on our results in section 6.5.

Proposition 6.59. *Let $\Gamma = (V, E, d_l)$ be the metric tree representing a tripod, where E is the union of three identical copies of the unit interval $[0, 1]$ and V consists of four points with one of them being the common end $0 \in V$ shared by all the three edges $e_1, e_2, e_3 \in E$. For $i = 1, 2, 3$, let ν_i be a probability measure supported in the interval $[\frac{1}{2}, 1]$ of edge e_i. The Dirac measure δ_0 at the vertex $0 \in V$ is the unique barycenter of the probability measure $\mathbb{P} := \sum_{i=1}^n \frac{1}{3}\delta_{\nu_i} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$.*



Proof. For three given points $x_1, x_2, x_3 \in [\frac{1}{2}, 1]$ in the supports of ν_1, ν_2, ν_3 respectively, we claim that the vertex 0 is the unique barycenter of the measure $\mu := \sum_{i=1}^n \frac{1}{3}\delta_{x_i}$. Fix a point $y \in \Gamma$. Without loss of generality, we assume that $y \in [0, 1]$ is on the edge e_3. Since $x_1 + x_2 \geq 1 \geq x_3$, we have the inequality

$$\begin{aligned}\int_\Gamma d_l(x, y)^2 \,\mathrm{d}\,\mu(x) &= \frac{1}{3}[d_l(x_1, 0) + d_l(0, y)]^2 + \frac{1}{3}[d_l(x_2, 0) + d_l(0, y)]^2 + \frac{1}{3}[d_l(x_3, 0) - d_l(0, y)]^2 \\ &= \sum_{i=1}^3 \frac{1}{3}x_i^2 + y^2 + \frac{2}{3}y(x_1 + x_2 - x_3) \geq \sum_{i=1}^3 \frac{1}{3}x_i^2 = \int_\Gamma d_l(x, 0)^2 \,\mathrm{d}\,\mu(x), \qquad (6.45)\end{aligned}$$

which is an equality if and only if $y = 0$. Hence, our claim is proven.

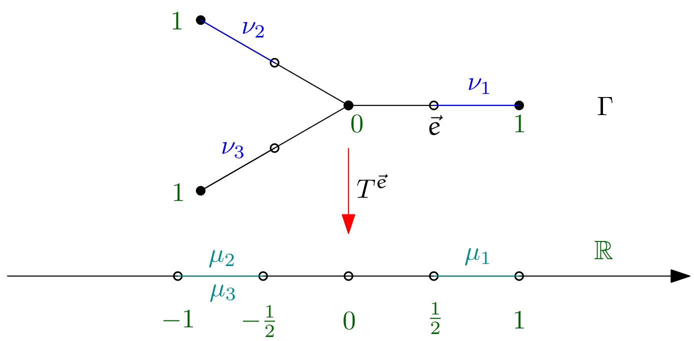
Let $\mu_{\mathbb{P}}$ be a barycenter of $\mathbb{P}$. Thanks to the gluing lemma [64, Lemma 7.1], there are random variables X, X_1, X_2, X_3 with laws $\mu_{\mathbb{P}}, \nu_1, \nu_2, \nu_3$ respectively such that $\mathbb{E}\, d_l(X, X_i)^2 = d_W(\mu_{\mathbb{P}}, \mu_i)^2$ for $i = 1, 2, 3$. By the claim proven in the previous paragraph,

$$\int_{\mathcal{W}_2(\Gamma)} d_W(\mu_{\mathbb{P}}, \nu)^2 \,\mathrm{d}\,\mathbb{P}(\nu) = \mathbb{E}\left[\sum_{i=1}^3 \frac{1}{3} d_l(X, X_i)^2\right] \geq \mathbb{E}\left[\sum_{i=1}^3 \frac{1}{3} d_l(0, X_i)^2\right] = \int_{\mathcal{W}_2(\Gamma)} d_W(\delta_0, \nu)^2 \,\mathrm{d}\,\mathbb{P}(\nu),$$

which must be an equality as $\mu_{\mathbb{P}}$ is a barycenter of $\mathbb{P}$. Hence, we have $X = 0$ almost everywhere and thus $\mu_{\mathbb{P}} = \delta_0$, which concludes the proof since $\mu_{\mathbb{P}}$ is arbitrarily chosen. ∎

We can also prove Proposition 6.59 using reduction maps.

Alternative proof of Proposition 6.59. Consider the order $0, 1$ for the two ends of the edge e_1, and denote by $\vec{e} = \overrightarrow{\{0,1\}}$ this oriented edge. Denote by $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ the push-forward map



associated to $\vec{e}$ (Definition 6.20), and define $\mu_i := \mathcal{T}(\nu_i)$ for $i = 1, 2, 3$. We claim that the barycenter $\mu_{\mathbb{Q}}$ of $\mathbb{Q} := \sum_{i=1}^{3} \frac{1}{3}\delta_{\mu_i}$ is supported in the interval $[-\frac{1}{2}, 0]$. Indeed, by assumptions, Lemma 1.33 implies $f_{\mu_1}^{-1}(1), f_{\mu_1}^{-1}(0) \in [\frac{1}{2}, 1]$ and $f_{\mu_i}^{-1}(1), f_{\mu_i}^{-1}(0) \in [-1, -\frac{1}{2}]$ for $i = 2, 3$. It follows from the formula of Wasserstein barycenters on $\mathbb{R}$ (Theorem 6.18) that

$$f_{\mu_{\mathbb{Q}}}^{-1}(1) = \frac{1}{3}(f_{\mu_1}^{-1}(1) + f_{\mu_2}^{-1}(1) + f_{\mu_3}^{-1}(1)) \le \frac{1}{3}(1 - \frac{1}{2} - \frac{1}{2}) \le 0, \quad \text{and}$$
$$f_{\mu_{\mathbb{Q}}}^{-1}(0) = \frac{1}{3}(f_{\mu_1}^{-1}(0) + f_{\mu_2}^{-1}(0) + f_{\mu_3}^{-1}(0)) \ge \frac{1}{3}(\frac{1}{2} - 1 - 1) \ge -\frac{1}{2},$$

which implies our claim according to Lemma 1.33. Since the assumptions of Proposition 6.59 are symmetric with respect to the three edges of Γ, our claim remains valid for all oriented edges corresponding to e_i with $i = 1, 2, 3$. Hence, Proposition 6.59 follows from Theorem 6.57. □

The endpoint $\frac{1}{2}$ for the interval $[\frac{1}{2}, 1]$ is optimal in the assumptions of Proposition 6.59. We now modify ν_3 to violate the assumptions, demonstrating how it forces a change of the barycenter $\mu_{\mathbb{P}}$.

Proposition 6.60. *Let $\Gamma = (V, E, d_l)$ be the tripod with three edges identified with the unit interval $[0, 1]$ such that $0 \in V$ is the common end shared by all the three edges $e_1, e_2, e_3 \in E$. Fix a positive number $0 < \theta < \frac{1}{2}$. Consider three points $x_1 := 1 \in e_1$, $x_2 := \frac{1}{2} \in e_2$, $x_3 := \frac{1}{2} - \theta \in e_3$ located at three different edges respectively, and define Dirac measures $\nu_i = \delta_{x_i}$ for $i = 1, 2, 3$. The barycenter of $\mathbb{P} := \sum_{i=1}^{n} \frac{1}{3}\delta_{\nu_i}$ is unique, and it is the Dirac measure at point $\frac{\theta}{3} \in e_1$.*

Proof. Define $\mu := \sum_{i=1}^{3} \frac{1}{3}\nu_i = \sum_{i=1}^{3} \frac{1}{3}\delta_{x_i}$. We claim that the point $z_\mu := \frac{\theta}{3} \in e_1$ located at edge e_1 is the unique barycenter of μ. To minimize the integral $I(y) := d_W(\delta_y, \mu)^2 = \int_\Gamma d_l(y, x)^2 \, \mathrm{d}\,\mu(x)$ for $y \in \Gamma$, we discuss the possibilities of y locating at different edges. For the case $y \in e_1$, with the same calculation as (6.45), the integral

$$I(y) = \sum_{i=1}^{3} \frac{1}{3}x_i^2 + y^2 + \frac{2}{3}y(x_2 + x_3 - x_1)$$

has a unique local minimizer $y = -\frac{1}{3}(x_2 + x_3 - x_1) = \frac{\theta}{3} = z_\mu \in e_1$. As for the cases that $y \in e_2$ or $y \in e_3$, since $x_1 + x_3 > x_2$ and $x_1 + x_2 > x_3$, $I(y)$ is locally minimized at $y = 0 \in e_2 \cap e_3$ according to similar calculations. Since 0 is a common vertex of all three edges, we conclude that $y = z_\mu$ is the unique global minimizer of $I(y)$ over $y \in \Gamma$, which proves our claim.

To show that $\mu_{\mathbb{P}} := \delta_{z_\mu}$ is the unique barycenter of $\mathbb{P}$, observe that for $\eta \in \mathcal{W}_2(\Gamma)$,

$$\begin{aligned}\int_{\mathcal{W}_2(\Gamma)} d_W(\eta,\nu)^2 \,\mathrm{d}\,\mathbb{P}(\nu) &= \sum_{i=1}^{3} \frac{1}{3} \int_\Gamma d_l(y,x_i)^2 \,\mathrm{d}\,\eta(y) = \int_\Gamma \int_\Gamma d_l(y,x)^2 \,\mathrm{d}\,\mu(x)\,\mathrm{d}\,\eta(y) \\ &\geq \int_\Gamma d_l(z_\mu,x)^2 \,\mathrm{d}\,\mu(x) = \int_{\mathcal{W}_2(\Gamma)} d_W(\delta_{z_\mu},\nu)^2 \,\mathrm{d}\,\mathbb{P}(\nu),\end{aligned}$$

where the equality is reached if and only if $y = z_\mu$ for η-almost every y, i.e., $\eta = \delta_{z_\mu}$. □

Regarding the singularity of Wasserstein barycenters, the rigid properties proven for the case of real line can be naturally extended to the general case of metric trees. For simplicity, in the following proposition, we only apply the restriction property of Wasserstein barycenters (Proposition 5.2) for finitely many times.

Proposition 6.61. *Let $\Gamma = (V,E,d_l)$ be a metric tree. Let $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\Gamma))$ be a probability measure. Assume that $\mathbb{P}$ has a barycenter $\mu_{\mathbb{P}}$ such that $\mu_{\mathbb{P}}$ has compact support, $\mu_{\mathbb{P}}$ gives no mass to the set of vertices (i.e., $\mu(V) = 0$), and $\mu_{\mathbb{P}}$ is singular with respect to the canonical reference measure $\mathcal{H}$ on Γ. Then for $\mathbb{P}$-almost every ν,* $\operatorname{supp}(\nu)$ *is compact and ν is singular with respect to $\mathcal{H}$.*

Proof. Since the set $\operatorname{supp}(\mu_{\mathbb{P}})$ is bounded and $\inf_{e\in E} l(e) > 0$, there exists at almost finitely many edges e_k, $k = 1,2,3,\ldots,n$, such that $\mu_{\mathbb{P}}(e_k) > 0$. As $\mu_{\mathbb{P}}(V) = 0$, we obtain the following decomposition,

$$\mu_{\mathbb{P}} = \lambda_1\,\mu^1 + \lambda_2\,\mu^2 + \cdots + \lambda_n\,\mu^n,$$

where for $1 \leq k \leq n$, $0 < \lambda_k := \mu_{\mathbb{P}}(e_k) \leq 1$, $\mu^k \in \mathcal{W}_2(\Gamma)$ assigns full mass to the interior of the edge e_k. We shall first show how to reduce our proposition to the case $n = 1$. Assume now $n > 1$.

By applying the restriction property (Proposition 5.2) for $n-1$ times, we construct n measurable maps $F^k : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\Gamma)$ such that μ^k is a barycenter of $\mathbb{Q}^k := F^k{}_{\#}\mathbb{P}$ and for $\nu \in \mathcal{W}_2(\Gamma)$,

$$\nu = \lambda_1\,F^1(\nu) + \lambda_2\,F^2(\nu) + \cdots + \lambda_n\,F^n(\nu). \tag{6.46}$$

The construction of $F^k, 1 \leq k \leq n$, is done as follows. We first apply Proposition 5.2 to the decomposition $\mu_{\mathbb{P}} = \lambda_1\,\mu^1 + (1-\lambda_1)\eta^1$ with $\eta^1 := \frac{1}{1-\lambda_1}\sum_{k=2}^n \lambda_k\,\mu^k$, and thus obtain two maps $A^1, B^1 : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\Gamma)$ such that μ^1 is a barycenter of $\mathbb{Q}^1 := A^1{}_{\#}\mathbb{P}$ and η^1 is a barycenter of $\mathbb{T}^1 := B^1{}_{\#}\mathbb{P}$. Moreover, for $\nu \in \mathcal{W}_2(\Gamma)$, we have $\nu = \lambda_1\,A^1(\nu) + (1-\lambda_1)B^1(\nu)$. Then we apply Proposition 5.2 to the decomposition $\eta^1 = \frac{\lambda_2}{1-\lambda_1}\mu^2 + (1-\frac{\lambda_2}{1-\lambda_1})\eta^2$ with $\eta^2 := \frac{1}{1-\lambda_1-\lambda_2}\sum_{k=3}^n \lambda_k\,\mu^k$. Hence, we obtain two measurable maps $A^2, B^2 : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\Gamma)$ such that μ^2 is a barycenter of $\mathbb{Q}^2 := A^2{}_{\#}\mathbb{T}^1 = [A^2 \circ B^1]_{\#}\mathbb{P}$ and η^2 is a barycenter of $\mathbb{T}^2 := B^2{}_{\#}\mathbb{T}^1$. Moreover, for $\nu \in \mathcal{W}_2(\Gamma)$,

$$\begin{aligned}\nu &= \lambda_1\,A^1(\nu) + (1-\lambda_1)B^1(\nu) \\ &= \lambda_1\,A^1(\nu) + (1-\lambda_1)\left[\frac{\lambda_2}{1-\lambda_1}A^2(B^1(\nu)) + (1-\frac{\lambda_2}{1-\lambda_1})B^2(B^1(\nu))\right] \\ &= \lambda_1\,A^1(\nu) + \lambda_2\,A^2 \circ B^1(\nu) + (1-\lambda_1-\lambda_2)B^2 \circ B^1(\nu).\end{aligned}$$

By repeating the above application of Proposition 5.2 for $n-1$ times, we obtain maps $A^k, B^k : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\Gamma)$ for $1 \le k \le n-1$. To complete the construction, it suffices to set $F^n := B^{n-1} \circ B^{n-2} \circ \cdots B^1$ and $F^k := A^k \circ B^{k-1} \circ \cdots \circ B^1$ for $1 \le k \le n-1$.

Thanks to the decomposition (6.46), for the proof of our proposition, it suffices to show that the following statement holds for any $1 \le k \le n$: for $\mathbb{Q}^k$-almost every ν, $\operatorname{supp}(\nu)$ is compact and ν is singular with respect to $\mathcal{H}$. Hence, the proposition can be reduced the case that $\mu_{\mathbb{P}}$ assigns full mass to the interior of some edge.

Fix an oriented edge $\vec{\mathfrak{e}}$ of Γ and denote by $\mathring{\mathfrak{e}}$ its interior. We now prove the proposition under the assumption that $\mu_{\mathbb{P}}(\mathring{\mathfrak{e}}) = 0$. Denote by $T^{\vec{\mathfrak{e}}} : \Gamma \to \mathbb{R}$ and $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ respectively the reduction map (Proposition 6.19) and the push-forward map (Definition 6.20) associated to $\vec{\mathfrak{e}}$. Define $\mathbb{Q} := \mathcal{T}_{\#}\mathbb{P}$ and denote by $\mu_{\mathbb{Q}}$ the unique barycenter of $\mathbb{Q}$. According to Corollary 6.27, $\mu_{\mathbb{Q}} = \mathcal{T}(\mu_{\mathbb{P}})$. By definition of $\mathcal{H}$, since $T^{\vec{\mathfrak{e}}}|_{\vec{\mathfrak{e}}} : \vec{\mathfrak{e}} \to [0, l(\vec{\mathfrak{e}})]$ is an isometry and $\operatorname{supp}(\mu_{\mathbb{P}}) \subset \vec{\mathfrak{e}}$, $\mu_{\mathbb{P}}$ is singular if and only if $\mu_{\mathbb{Q}} = T^{\vec{\mathfrak{e}}}{}_{\#}\mu_{\mathbb{P}}$ is singular. Since $\mu_{\mathbb{Q}}$ is singular, Theorem 6.51 implies that for $\mathbb{Q}$-almost every $\nu \in \mathcal{W}_2(\mathbb{R})$, ν is singular. That is to say, for $\mathbb{P}$-almost every $\nu \in \mathcal{W}_2(\Gamma)$, $\mathcal{T}(\nu)$ is singular. We claim that

$$\mathcal{T}(\nu) \in \mathcal{W}_2(\mathbb{R}) \text{ is singular } \implies \nu \in \mathcal{W}_2(\Gamma) \text{ is singular .}$$

We prove the claim by contradiction and assume that $\nu \in \mathcal{W}_2(\Gamma)$ is not singular. Then we can write $\nu = \theta\,\nu_1 + (1-\theta)\nu_2$ with $0 < \theta < 1$, $\nu_1, \nu_2 \in \mathcal{W}_2(\Gamma)$ such that ν_1 is absolutely continuous. As $\mathcal{T}$ is a push-forward map, $\mathcal{T}(\nu) = T^{\vec{\mathfrak{e}}}{}_{\#}\nu = \theta\,\mathcal{T}(\nu_1) + (1-\theta)\mathcal{T}(\nu_2)$. According to Lemma 6.21, $\mathcal{T}(\nu_1)$ is absolutely continuous, which implies that $\mathcal{T}(\nu)$ is not singular, contradicting the assumption. Hence, the claim is proven, and for $\mathbb{P}$-almost every $\nu \in \mathcal{W}_2(\Gamma)$, ν is singular with respect to $\mathcal{H}$.

We are left to prove the conclusion concerning the compactness of $\operatorname{supp}(\nu)$. According to Proposition 6.29, since $\mu_{\mathbb{Q}}$ has compact support, for $\mathbb{P}$-almost every ν, $\mathcal{T}(\nu)$ has compact support. Note that $T^{\vec{\mathfrak{e}}}$ maps the metric ball $B(\vec{\mathfrak{e}}_0, r) \subset \Gamma$ centered at $\vec{\mathfrak{e}}_0$ with radius $r > 0$ to the metric ball $B(0, r) \subset \mathbb{R}$ centered at 0 with the same radius r. Hence, $\mathcal{T}(\nu) \in \mathcal{W}_2(\mathbb{R})$ has compact support if and only if $\nu \in \mathcal{W}_2(\Gamma)$ has compact support, which concludes the proof. ☐

We can readily demonstrate that absolute continuity is not a rigid property of Wasserstein barycenters on $\mathbb{R}$. For instance, if we consider $\mathbb{P} := \frac{1}{2}\delta_\nu + \frac{1}{2}\delta_{\delta_x} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}))$, where ν is an absolutely continuous measure, then its barycenter $\mu_{\mathbb{P}}$ is absolutely continuous, while δ_x, a singular measure also in $\operatorname{supp}(\mathbb{P})$, is not. Nevertheless, it is particularly interesting to construct an absolutely continuous barycenter $\mu_{\mathbb{P}}$ with $\operatorname{supp}(\mathbb{P})$ being a subset of singular measures, such as those supported on merely two fixed points. This construction, which also helps to further illuminate Proposition 6.61, is presented in the following proposition. To build this example, it will suffice to set $\mu := \mathcal{L}^1|_{[0,1]}$.

Proposition 6.62. *Let μ be a probability measure supported in the unit interval $[0,1]$. Denote by $\widetilde{\mu}$ its dual measure (Definition 6.30). Consider the map $F : [0,1] \to \mathcal{W}_2(\mathbb{R})$ defined by*

$$F(x) := x\,\delta_0 + (1-x)\delta_1.$$

Then $\mu = \mu_{\mathbb{P}}$ is the unique barycenter of $\mathbb{P} := F_{\#}\widetilde{\mu}$.

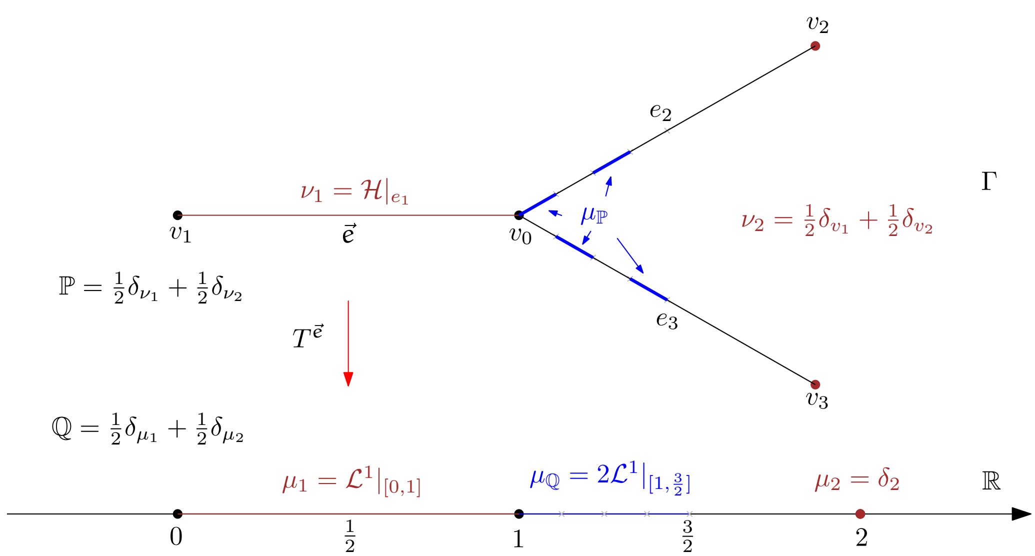
Proof. As a result of direct calculations, we observe that the dual measure of $F(x) = x\,\delta_0 + (1-x)\delta_1$ is the Dirac measure δ_x at x. According to Theorem 6.18, the unique barycenter $\mu_{\mathbb{P}}$ of $\mathbb{P}$ satisfies that, for $t \in (0,1)$,

$$f_{\mu_{\mathbb{P}}}^{-1}(t) = \int_{\mathcal{W}_2(\mathbb{R})} f_\nu^{-1}(t)\, \mathrm{d}\,\mathbb{P}(\nu) = \int_0^1 f_{F(x)}^{-1}(t)\, \mathrm{d}\,\widetilde{\mu}(x) = \int_0^1 f_{\delta_x}(t)\, \mathrm{d}\,\widetilde{\mu}(x)$$
$$= \int_0^1 \mathbb{1}_{[0,t]}(x)\, \mathrm{d}\,\widetilde{\mu}(x) = \widetilde{\mu}([0,t]) = f_{\widetilde{\mu}}(t).$$

It follows that $0 \le f_{\mu_{\mathbb{P}}}^{-1} \le 1$, which further implies $\mathrm{supp}(\mu_{\mathbb{P}}) \subset [0,1]$ according to Lemma 1.33. Therefore, the above equality shows that $\mu_{\mathbb{P}}$ is the dual measure of $\widetilde{\mu}$, which implies our conclusion $\mu = \mu_{\mathbb{P}}$ by Lemma 6.31. $\square$

The uniqueness of Wasserstein barycenters on the real line is guaranteed by Theorem 6.18. Moreover, given a Riemannian manifold $(M, d_{\mathfrak{g}})$, the barycenter of $\mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(M))$ is unique if $\mathbb{P}$ assigns positive mass to the set of absolutely continuous measures (Section 2.3). However, as illustrated by the following example, such uniqueness fails to hold on metric trees due to their inherent branching structure.

Proposition 6.63. *Let $\Gamma = (V, E, d_l)$ be the tripod with three edges of unit length, i.e., $V = \{v_0, v_1, v_2, v_3\}$, $E = \{e_1, e_2, e_3\}$ with $e_i = \{v_0, v_i\}$ and $l(e_i) = 1$ for $i = 1,2,3$. Let ν_1 be the uniform probability measure on e_1 and $\nu_2 = \frac{1}{2}\delta_{v_2} + \frac{1}{2}\delta_{v_3}$ be the averaged sum of two Dirac measures at vertices v_2 and v_3. Define $\mathbb{P} := \frac{1}{2}\delta_{\nu_1} + \frac{1}{2}\delta_{\nu_2}$. Then $\mathbb{P}$ has infinitely many absolutely continuous barycenters, and optimal transport maps pushing forward ν_1 to ν_2 are not unique.*



Proof. Consider the oriented edge $\vec{e} = \overrightarrow{\{v_1, v_0\}}$ and let $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ be the push-forward map associated to $\vec{e}$ (Definition 6.20). Define $\mu_1 := \mathcal{T}(\nu_1) = \mathcal{L}^1|_{[0,1]}$, $\mu_2 := \mathcal{T}(\nu_2) = \delta_2$ and $\mathbb{Q} := \mathcal{T}_\# \mathbb{P} = \frac{1}{2}\delta_{\mu_1} + \frac{1}{2}\delta_{\mu_2}$. According to the formula of Wasserstein barycenters on $\mathbb{R}$ (Theorem 6.18), $\mu_{\mathbb{Q}} := 2\mathcal{L}^1|_{[1,\frac{3}{2}]}$ is the barycenter of $\mathbb{Q}$.

By direct calculation using Theorem 1.37, we obtain $d_W(\mu_1,\mu_{\mathbb{Q}}) = d_W(\mu_{\mathbb{Q}},\mu_2) = \frac{1}{2}d_W(\mu_1,\mu_2) = \sqrt{7}/(2\sqrt{3})$. Since ν_1 is supported in the oriented edge $\vec{\mathfrak{e}}$, Proposition 6.19 implies that for any measure $\nu \in \mathcal{W}_2(\Gamma)$, $d_W(\nu_1,\nu) = d_W(\mu_1,\mathcal{T}(\nu))$. We prove that any measure $\mu_{\mathbb{P}}$ constructed as follows is a barycenter of $\mathbb{P}$. Define $\mu_{\mathbb{P}} := \frac{1}{2}\eta_2 + \frac{1}{2}\eta_3$ with η_2 (respectively η_3) being a probability measure supported in the edge e_2 (respectively e_3) such that

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

Since $\mu_{\mathbb{Q}} = \mathcal{T}(\mu_{\mathbb{P}})$ is the barycenter of $\mathbb{Q}$, we have

$$d_W(\nu_1,\mu_{\mathbb{P}}) = d_W(\mu_1,\mathcal{T}(\mu_{\mathbb{P}})) = d_W(\mu_1,\mu_{\mathbb{Q}}) = \frac{1}{2}d_W(\mu_1,\mu_2) = \frac{1}{2}d_W(\nu_1,\nu_2).$$

Since $e_2 \cup e_3$ is isometric to a segment of length 2, we can deduce from Theorem 1.37 the following optimal transport plan γ between $\mu_{\mathbb{P}} = \frac{1}{2}\eta_2 + \frac{1}{2}\eta_3$ and ν_2,

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

Using the fact that both the restrictions $T^{\vec{\mathfrak{e}}}|_{e_2}$ and $T^{\vec{\mathfrak{e}}}|_{e_3}$ are isometric, we obtain

$$\begin{aligned}
d_W(\mu_{\mathbb{P}},\nu_2)^2 &= \int_\Gamma d_l(x,y)^2 \,\mathrm{d}\,\gamma(x,y) \\
&= \frac{1}{2}\int_{e_2} d_l(x,v_2)^2 \,\mathrm{d}\,\eta_2(x) + \frac{1}{2}\int_{e_3} d_l(x,v_3)^2 \,\mathrm{d}\,\eta_3(x) \\
&= \frac{1}{2}d_W(\mathcal{T}(\eta_2),\mu_2)^2 + \frac{1}{2}d_W(\mathcal{T}(\eta_3),\mu_2)^2 \\
&= d_W(\frac{1}{2}\mathcal{T}(\eta_2) + \frac{1}{2}\mathcal{T}(\eta_3),\mu_2)^2 = d_W(\mathcal{T}(\mu_{\mathbb{P}}),\mu_2)^2 \\
&= d_W(\mu_{\mathbb{Q}},\mu_2)^2 = \left[\frac{1}{2}d_W(\mu_1,\mu_2)\right]^2 = \frac{1}{4}d_W(\nu_1,\nu_2)^2.
\end{aligned}$$

Therefore, $d_W(\nu_1,\mu_{\mathbb{P}}) = d_W(\mu_{\mathbb{P}},\nu_2) = \frac{1}{2}d_W(\nu_1,\nu_2)$, which implies that $\mu_{\mathbb{P}}$ is a barycenter of $\mathbb{P} = \frac{1}{2}\delta_{\nu_1} + \frac{1}{2}\delta_{\nu_2}$ according to Lemma 2.7 (c.f. proof of Proposition 6.39). From our construction of $\mu_{\mathbb{P}}$, there are infinitely many possible choices and all of them are absolutely continuous.

As for the optimal transport maps pushing forward ν_1 to ν_2, we can find multiple of them by dividing the edge e_1 into two parts with equal lengths. For example, set $I_1 := [T^{\vec{\mathfrak{e}}}]^{-1}([0,\frac{1}{2}])$ to be the pre-image of $[0,\frac{1}{2}]$ under the map $T^{\vec{\mathfrak{e}}}$ and set $I_2 := e_1 \backslash I_1$. Define two maps $F_1, F_2 : e_1 \to \{v_2, v_3\}$ as follows: F_1 sends points of I_1 to v_2 and sends points of I_2 to v_3; F_2 sends points of I_1 to v_3 and sends points of I_2 to v_2. Since $\nu_1(I_1) = \nu_1(I_2) = \frac{1}{2}$, both F_1 and F_2 push forward ν_1 to ν_2. By direct calculation, we have

$$\int_{e_1} d_l(F_1(x),x)^2 \,\mathrm{d}\,\nu_1(x) = \int_{e_2} d_l(F_2(x),x)^2 \,\mathrm{d}\,\nu_1(x) = \int_{[0,\frac{1}{2}]\cup(\frac{1}{2},1]} |1-x+1|^2 \,\mathrm{d}\,x = \frac{7}{3},$$

which shows that F_1 and F_2 are both optimal transport maps since $d_W(\nu_1,\nu_2)^2 = d_W(\mu_1,\mu_2)^2 = 7/3$. $\square$

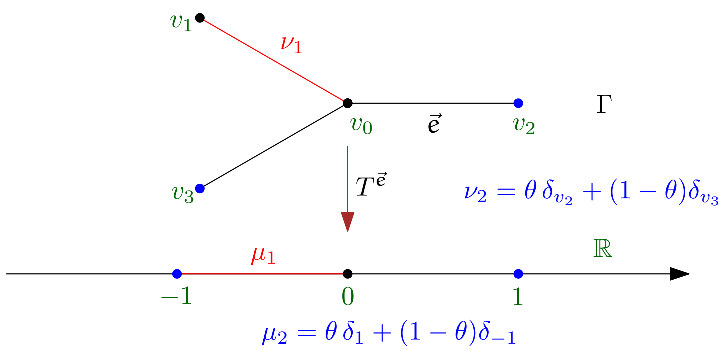
While Proposition 6.63 demonstrates that $\mathbb{P}$ can have infinitely many barycenters, a key question remains: does any barycenter $\mu_{\mathbb{P}}$ assigns positive mass to the common vertex v_0? To address this, we can, as noted earlier in this subsection, apply results from Section 6.5. The upcoming example highlights a situation where Theorem 6.54 alone is insufficient to conclude that $\mu_{\mathbb{P}}(\{v_0\}) = 0$. This therefore serves as an ideal example to illustrate a flexible application of the restriction property of Wasserstein barycenters.

Proposition 6.64. *Let $\Gamma = (V, E, d_l)$ be the tripod with three edges of unit length, i.e., $V = \{v_0, v_1, v_2, v_3\}$, $E = \{e_1, e_2, e_3\}$ with v_0, v_i being the two ends of e_i and $l(e_i) = 1$ for $i = 1, 2, 3$. Consider two probability measures $\nu_1, \nu_2 \in \mathcal{W}_2(\Gamma)$ such that ν_1 is supported in e_1, $\nu_1(\{v_1\}) = 0$, and $\nu_2 = \theta\, \delta_{v_2} + (1-\theta)\delta_{v_3}$ with $\theta \in [0,1]$. Define $\mathbb{P} := \frac{1}{2}\delta_{\nu_1} + \frac{1}{2}\delta_{\nu_2}$. Then any barycenter $\mu_{\mathbb{P}}$ of $\mathbb{P}$ gives no mass to the vertices of Γ, i.e., $\mu_{\mathbb{P}}(V) = 0$.*

Moreover, if ν_1 is absolutely continuous (with respect to $\mathcal{H}$), then all barycenters of $\mathbb{P}$ are absolutely continuous.

Proof. We first prove the claim that δ_{v_i} is not a barycenter of $\mathbb{P}$ for any $i = 0, 1, 2, 3$.

We begin with the case for v_0 since it is more complicated than others. Assume without loss of generality that $\nu_2(\{v_2\}) > 0$, i.e., $0 < \theta \le 1$. Consider the oriented edge $\vec{e} = \overrightarrow{\{v_0, v_2\}}$ and denote by $\mathcal{T} : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\mathbb{R})$ the push-forward map associated to $\vec{e}$ (Definition 6.20). Define $\mu_1 := \mathcal{T}(\nu_1)$,



$\mu_2 := \mathcal{T}(\nu_2) = \theta\, \delta_1 + (1-\theta)\delta_{-1}$, and $\mathbb{Q} := \mathcal{T}_{\#}\mathbb{P} = \frac{1}{2}\delta_{\mu_1} + \frac{1}{2}\delta_{\mu_2}$. By Lemma 1.33, $f_{\mu_1}^{-1}(1) > -1$ since $\nu_1(\{v_1\}) = 0$, and $f_{\mu_2}^{-1}(1) = 1$ as $\theta = \nu_2(\{v_2\}) > 0$. Therefore, according to Theorem 6.18, the unique barycenter $\mu_{\mathbb{Q}}$ of $\mathbb{Q}$ satisfies

$$f_{\mu_{\mathbb{Q}}}^{-1}(1) = \frac{1}{2} f_{\mu_1}^{-1}(1) + \frac{1}{2} f_{\mu_2}^{-1}(1) > 0.$$

Hence, by Lemma 1.33, $\mu_{\mathbb{Q}}$ is not supported in $\mathbb{R}_-$. It follows from Lemma 6.52 that δ_{v_0} is not a barycenter of $\mathbb{P}$.

As for the case v_i with $i = 1, 2, 3$, we consider the oriented edge $\overrightarrow{\{v_i, v_0\}}$ and the corresponding reduced Wasserstein barycenter problem on $\mathbb{R}$ as above. Note that the whole graph is mapped to a subset of $\mathbb{R}_+$ by the reduction map associated to $\overrightarrow{\{v_i, v_0\}}$ (Proposition 6.19). Hence, if δ_{v_i} is a barycenter of $\mathbb{P}$, then $\nu_1 = \nu_2 = \delta_{v_i}$ by Lemma 6.52, which is impossible under our assumptions. This concludes the proof of our previous claim.

We prove by contradiction $\mu_{\mathbb{P}}(V) = 0$. Otherwise, we can decompose a barycenter $\mu_{\mathbb{P}}$ as follows,

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

where i is one of the indices $0, 1, 2, 3$, $0 < \lambda < 1$ and $\mu^2 \in \mathcal{W}_2(\Gamma)$. By Proposition 5.2, there exists a measurable map $F^1 : \mathcal{W}_2(\Gamma) \to \mathcal{W}_2(\Gamma)$ such that $F^1(\nu)$ is absolutely continuous with respect to ν and δ_{v_i} is a barycenter of $\mathbb{Q}^1 := F^1{}_{\#}\mathbb{P}$. However, $\mathbb{Q}^1 = \frac{1}{2}\delta_{F^1(\nu_1)} + \frac{1}{2}\delta_{F^1(\nu_2)}$ satisfies the same assumption as $\mathbb{P}$, which leads to a contradiction due to the previous claim.

If ν_1 is absolutely continuous, Proposition 6.29 implies that any $\mu_{\mathbb{P}}$ is absolutely continuous when restricted to the interior of each edge. It follows from $\mu_{\mathbb{P}}(V) = 0$ that v_1 is absolutely continuous. $\square$

Bibliography

[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] C. Aliprantis and K. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer, 2007.

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

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

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

[7] L. Ambrosio and N. Gigli, *A user's guide to optimal transport*, in Modelling and optimisation of flows on networks, Springer, 2013, pp. 1–155.

[8] 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.

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

[10] J. J. Benedetto and W. Czaja, *Integration and modern analysis*, vol. 15, Springer, 2009.

[11] Y. M. Berezansky, Z. G. Sheftel, and G. F. Us, *Functional Analysis: Vol. I*, vol. 85, Birkhäuser, 2012.

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

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

[14] J. Bertrand and B. R. Kloeckner, *A geometric study of Wasserstein spaces: Hadamard spaces*, Journal of Topology and Analysis, 4 (2012), pp. 515–542.

[15] ——, *A geometric study of Wasserstein spaces: an addendum on the boundary*, in International Conference on Geometric Science of Information, Springer, 2013, pp. 405–412.

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

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

[18] J. A. Bondy and U. S. R. Murty, *Graph theory*, Springer Publishing Company, Incorporated, 2008.

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

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

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

[22] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, vol. 319, Springer Science & Business Media, 2013.

[23] D. Burago, I. D. Burago, Y. Burago, S. Ivanov, S. V. Ivanov, and S. A. Ivanov, *A course in metric geometry*, vol. 33, American Mathematical Soc., 2001.

[24] 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.

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

[26] E. Çinlar, *Probability and Stochastics*, Graduate Texts in Mathematics, Springer New York, 2011.

[27] 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.

[28] S. N. Cohen and R. J. Elliott, *Stochastic calculus and applications*, vol. 2, Springer, 2015.

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

[30] 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.

[31] ——, *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.

[32] R. Diestel, *Graph Theory*, Graduate Texts in Mathematics, Springer Berlin Heidelberg, 2018.

[33] N. Dinculeanu, *Vector-valued stochastic processes. V. Optional and predictable variation of stochastic measures and stochastic processes*, Proceedings of the American Mathematical Society, 104 (1988), pp. 625–631.

[34] M. Erbar, D. L. Forkert, J. Maas, and D. Mugnolo, *Gradient flow formulation of diffusion equations in the Wasserstein space over a metric graph*, Networks and Heterogeneous Media, 17 (2022).

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

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

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

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

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

[40] ——, *Broad Foundations*, vol. 2 of Measure Theory, Torres Fremlin, Colchester CO3 3AT, England, 2009.

[41] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian Geometry*, Universitext, Springer Science & Business Media, Berlin, 2004.

[42] 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.

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

[44] 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.

[45] N. Grossman, *Hilbert manifolds without epiconjugate points*, Proceedings of the American Mathematical Society, 16 (1965), pp. 1365–1371.

[46] B.-X. Han, D. Liu, and Z. Zhu, *On the geometry of Wasserstein barycenter I*, arXiv preprint arXiv:2412.01190, (2024).

[47] B.-X. Han, D.-y. Liu, and Z.-n. Zhu, *Barycenter curvature-dimension condition for extended metric measure spaces*, arXiv preprint arXiv:2502.06793, (2025).

[48] S.-w. He, J. Wang, and J.-A. Yan, *Semimartingale Theory and Stochastic Calculus*, Taylor & Francis, 1992.

[49] C. Heil, *Introduction to real analysis*, vol. 280, Springer, 2019.

[50] T. Hotz, S. Skwerer, S. Huckemann, H. Le, J. Marron, J. C. Mattingly, E. Miller, J. Nolen, M. Owen, and V. Patrangenaru, *Sticky central limit theorems on open books*, (2013).

[51] J. Jacod and A. Shiryaev, *Limit theorems for stochastic processes*, vol. 288, Springer Science & Business Media, 2003.

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

[53] D. Kannan and V. Lakshmikantham, *Handbook of stochastic analysis and applications*, CRC Press, 2001.

[54] R. Kannan and C. K. Krueger, *Advanced analysis: on the real line*, Springer Science & Business Media, 2012.

[55] A. Kechris, *Classical Descriptive Set Theory*, Graduate Texts in Mathematics, Springer New York, 2012.

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

[57] Y.-H. Kim and B. Pass, *Multi-marginal optimal transport on Riemannian manifolds*, American Journal of Mathematics, 137 (2015), pp. 1045–1060.

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

[59] B. Kloeckner, *A geometric study of Wasserstein spaces: Euclidean spaces*, Annali della Scuola Normale Superiore di Pisa-Classe di Scienze, 9 (2010), pp. 297–323.

[60] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry, Volume 1*, vol. 1, John Wiley & Sons, Hoboken, 1996.

[61] T. Le, M. Yamada, K. Fukumizu, and M. Cuturi, *Tree-sliced variants of Wasserstein distances*, Advances in neural information processing systems, 32 (2019).

[62] T. Le Gouic and J.-M. Loubes, *Existence and consistency of Wasserstein barycenters*, Probability Theory and Related Fields, 168 (2017), pp. 901–917.

[63] R. Leadbetter, S. Cambanis, and V. Pipiras, *A Basic Course in Measure and Probability: Theory for Applications*, Cambridge university press, 2014.

[64] J. M. Lee, *Introduction to topological manifolds*, vol. 202, Springer Science & Business Media, 2010.

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

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

[67] G. Leoni, *A First Course in Sobolev Spaces*, American Mathematical Soc., 2 ed., 2017.

[68] 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.

[69] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Annals of Mathematics, 169 (2009), pp. 903–991.

[70] M. Mathey-Prevot and A. Valette, *Wasserstein distance and metric trees*, L'Enseignement Mathématique, 69 (2023), pp. 315–333.

[71] J. M. Mazón, J. D. Rossi, and J. Toledo, *Optimal mass transport on metric graphs*, SIAM Journal on Optimization, 25 (2015), pp. 1609–1632.

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

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

[74] G. J. Minty, *Monotone (nonlinear) operators in Hilbert space*, Duke Mathematical Journal, 29 (1962).

[75] I. P. Natanson, *Theory of functions of a real variable*, Courier Dover Publications, 2016.

[76] 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.

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

[78] V. M. Panaretos and Y. Zemel, *Statistical aspects of Wasserstein distances*, Annual review of statistics and its application, 6 (2019), pp. 405–431.

[79] ———, *An invitation to statistics in Wasserstein space*, SpringerBriefs in Probability and Mathematical Statistics, Springer, Cham, Switzerland, 2020.

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

[81] G. Peyré, M. Cuturi, et al., *Computational optimal transport: With applications to data science*, Foundations and Trends® in Machine Learning, 11 (2019), pp. 355–607.

[82] I. Rana, *An Introduction to Measure and Integration*, Alpha Science international, 2005.

[83] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, vol. 293, Springer Science & Business Media, 2013.

[84] F. Riesz and B. S. Nagy, *Functional analysis*, Courier Corporation, 2012.

[85] S. J. Robertson, *On discrete curvatures of trees*, arXiv preprint arXiv:2412.20661, (2024).

[86] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.

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

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

[89] S. Saks, *Theory of the Integral*, Hafner publishing company, New York, 2 ed., 1937.

[90] 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.

[91] A. Schmeding, *An Introduction to Infinite-Dimensional Differential Geometry*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2022.

[92] W. Sierpiński, *Les ensembles projectifs et analytiques.* Mem. Sci. Math. 112. Paris: Gauthier-Villars. 80 p. (1950)., 1950.

[93] S. M. Srivastava, *A course on Borel sets*, vol. 180, Springer Science & Business Media, 2008.

[94] K.-T. Sturm, *Probability measures on metric spaces of nonpositive curvature*, SFB 611, 2003.

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

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

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

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

[99] T. Tao, *An introduction to measure theory*, vol. 126, American Mathematical Soc., 2011.

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

[101] A. van der Vaart, *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2000.

[102] J. van Neerven, *Functional analysis*, vol. 201, Cambridge University Press, 2022.

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

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

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

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

[107] R. Wilson, *Introduction to Graph Theory*, Longman, 2010.

[108] Q. Xia, *Optimal paths related to transport problems*, Communications in Contemporary Mathematics, 5 (2003), pp. 251–279.

[109] ——, *Ramified optimal transportation in geodesic metric spaces.*, Advances in Calculus of Variations, 4 (2011).

[110] ——, *Motivations, ideas and applications of ramified optimal transportation*, ESAIM: Mathematical Modelling and Numerical Analysis, 49 (2015), pp. 1791–1832.

[111] W. P. Ziemer and M. Torres, *Modern real analysis*, vol. 278, Springer, 2017.

Titre : Régularité des barycentres de Wasserstein

Mots clés : barycentre Wasserstein, théorie du transport, géométrie riemannienne, arbre métrique

Résumé : Cette thèse porte sur l'étude des conditions géométriques qui gouvernent la régularité des barycentres de Wasserstein. L'objectif central est de comprendre comment la géométrie sous-jacente d'un espace métrique — en particulier, la présence ou l'absence d'une minoration de la courbure de Ricci — détermine si un barycentre est absolument continu ou singulier. La recherche aboutit à deux contributions principales, l'une concernant les espaces lisses et non-branchants (variétés riemanniennes) et l'autre les espaces singuliers et branchants (arbres métriques).

La première contribution majeure établit la continuité absolue des barycentres de Wasserstein sur les variétés riemanniennes sous des hypothèses significativement plus faibles que celles connues auparavant. Cette thèse montre que sur toute variété riemannienne complète dotée d'une courbure de Ricci minorée, le barycentre de Wasserstein $\mu_{\mathbb{P}}$ est garanti d'être absolument continu, à condition que sa mesure de probabilité définissante $\mathbb{P}$ sur l'espace de Wasserstein assigne une masse positive à l'ensemble des mesures absolument continues. Ce résultat assouplit considérablement les conditions des travaux antérieurs, qui exigeaient à la fois la compacité de la variété et que la mesure $\mathbb{P}$ assigne une masse positive à l'ensemble des mesures dont les densités sont uniformément bornées. La preuve introduit une approche novatrice reposant sur plusieurs nouveaux outils analytiques. Nous développons un nouveau type de fonctionnelles de déplacement, dont les propriétés découlent d'une nouvelle égalité hessienne pour les barycentres. Ces fonctionnelles permettent d'établir une estimation cruciale garantissant la régularité. De plus, la preuve intègre des outils de la théorie des espaces sousliniens pour lier l'hypothèse générale sur $\mathbb{P}$ aux conditions topologiques spécifiques requises par nos estimations fonctionnelles.

La seconde contribution majeure est le développement d'un cadre systématique pour analyser et caractériser les barycentres de Wasserstein sur les arbres métriques, où l'absence d'une minoration synthétique de la courbure de Ricci peut engendrer un comportement complexe. Ce cadre, qui combine un principe de localisation novateur (la "propriété de restriction") avec des techniques qui ramènent les problèmes de transport sur les arbres au cadre plus simple de la droite réelle, fournit les outils nécessaires pour étudier en détail la structure de ces barycentres. Il permet une investigation méthodique de la manière dont la topologie de branchement de l'arbre génère des phénomènes tels que la singularité et la non-unicité, qui distinguent ces espaces des variétés.

En résumé, cette thèse fait progresser notre compréhension des barycentres de Wasserstein en :

1. Prouvant un résultat définitif sur leur continuité absolue sur les variétés à courbure de Ricci minorée, étayé par de nouveaux outils analytiques incluant des fonctionnelles de déplacement novatrices et l'application de la théorie des espaces sousliniens.
2. Développant un cadre systématique pour caractériser les barycentres sur les arbres métriques, fournissant les outils pour analyser leur structure et leur singularité.

Title: Regularity of Wasserstein barycenters

Key words: Wasserstein barycenter, optimal transport, Riemannian geometry, metric tree

Abstract: This thesis investigates the geometric conditions that govern the regularity of Wasserstein barycenters. The central goal is to understand how the underlying geometry of a metric space—specifically, the presence or absence of a lower Ricci curvature bound—determines whether a barycenter is absolutely continuous or singular. The research yields two main contributions, one concerning smooth, non-branching spaces (Riemannian manifolds) and the other concerning singular, branching spaces (metric trees).

The first key contribution establishes the absolute continuity of Wasserstein barycenters on Riemannian manifolds under significantly weaker assumptions than previously known. This thesis shows that on any complete Riemannian manifold with a lower Ricci curvature bound, the Wasserstein barycenter $\mu_{\mathbb{P}}$ is guaranteed to be absolutely continuous, provided its defining probability measure $\mathbb{P}$ on the Wasserstein space assigns positive mass to the set of absolutely continuous measures. This result significantly relaxes the conditions of previous work, which required both the manifold to be compact and the measure $\mathbb{P}$ to assign positive mass to the set of measures with uniformly bounded densities. The proof introduces a novel approach built on several new analytical tools. We develop a new class of displacement functionals, whose properties are derived from a novel Hessian equality for barycenters. These functionals allow us to establish a crucial regularity-enforcing estimate. Furthermore, the proof integrates tools from Souslin space theory to bridge the gap between the general assumption on $\mathbb{P}$ and the specific topological conditions required by our functional estimates.

The second key contribution is the development of a systematic framework to analyze and characterize Wasserstein barycenters on metric trees, where the lack of a lower synthetic Ricci curvature bound can lead to complex behavior. This framework, which combines a novel localization principle (the "restriction property") with techniques that relate transport problems on trees to the simpler setting of the real line, provides the necessary tools to study the structure of these barycenters in detail. It allows for a methodical investigation of how the tree's branching topology generates phenomena such as singularity and non-uniqueness, which distinguish these spaces from manifolds.

In summary, this thesis advances our understanding of Wasserstein barycenters by:

1. Proving a definitive result on their absolute continuity on manifolds with a lower Ricci curvature bound, supported by new analytical tools including novel displacement functionals and the application of Souslin space theory.
2. Developing a systematic framework to characterize barycenters on metric trees, providing the tools to analyze their structure and singularity.