Gromov-Monge Quasi-metrics and Distance Distributions*

Facundo Mémoli\(^1\) and Tom Needham\(^2\)

\(^1\)Department of Mathematics and Department of Computer Science and Engineering, The Ohio State University, memoli@math.osu.edu
\(^2\)Department of Mathematics, The Ohio State University, needham.71@osu.edu

October 24, 2018

Abstract

Applications in shape analysis and object classification often require maps between metric spaces which preserve geometry as much as possible. In this paper, we combine the Monge formulation of optimal transport with the Gromov-Hausdorff construction to define a measure of the minimum amount of geometric distortion required to map one metric measure space onto another. We show that the resulting quantity, called Gromov-Monge distance, defines an extended quasi-metric on the space of isomorphism classes of metric measure spaces and that it can be promoted to a true metric on certain subclasses of mm-spaces. We also give precise comparisons between Gromov-Monge distance and several other metrics which have appeared previously, such as the Gromov-Wasserstein metric and the continuous Procrustes metric. Finally, we derive polynomial-time computable lower bounds for Gromov-Monge distance. These lower bounds are expressed in terms of classical invariants of mm-spaces called distance distributions. In the second half of the paper, which may be of independent interest, we study the discriminative power of these lower bounds for simple subclasses of metric measure spaces. We first consider the case of planar curves, where we give a counterexample to the Curve Histogram Conjecture of Brinkman and Olver. Next we show that one of our lower bounds distinguishes metric trees locally—trees which lie sufficiently close to one another in Gromov-Hausdorff distance are always distinguished—and generically with respect to a natural measure on the space of trees.

1 Introduction

Optimal transport (OT) problems deal with finding an optimal way to match two probability measures \(\alpha\) and \(\beta\) defined on the same ambient metric space \(Z\). There are two historical formulations of OT: the original Monge formulation [27] and the Kantorovich formulation [21]. The former involves minimizing a certain cost over all measure preserving transformations between the measures, whereas the latter introduces a convex relaxation which enlarges the set of admissible mappings to those probability measures on the product space \(Z \times Z\) whose marginals coincide with \(\alpha\) and \(\beta\).

The convex relaxation corresponding to the Kantorovich formulation of OT has been adapted to the setting when one wishes to compare not just two probability measures defined on the same ambient space, but to the more general case when one wishes to compare triples of the form \(X = (X, d_X, \mu_X)\), where \((X, d_X)\) is a compact metric space and \(\mu_X\) is a Borel probability measure on \(X\). These triples are called metric measure spaces (mm-spaces for short), and the resulting notion of dissimilarity between pairs of such triples receives the name of Gromov-Wasserstein distance [26] [35].

While the Gromov-Wasserstein distance defines a metric on the space \(\mathcal{M}^m\) of isomorphism classes of mm-spaces, the distance between a pair of spaces is realized by a coupling of their measures. In

*We acknowledge funding from these sources: NSF AF 1526513, NSF DMS 1723003, NSF CCF 1740761.

arXiv:1810.09646v1 [math.MG] 23 Oct 2018
applications, one is frequently more interested in an actual mapping between spaces, as the mapping provides a tool for object registration, landmark propagation and feature comparison [10, 20]. In this paper we explore a quasi-metric on the space $\mathcal{M}^m$ obtained by a similar generalization of the Monge formulation of OT. We refer to the resulting quasi-metric as the Gromov-Monge distance, and prove that in certain scenarios it can be promoted to a legitimate metric on mm-spaces. Furthermore, we prove that the more standard Gromov-Wasserstein distance is always realized as a limit of Gromov-Monge distances (Theorem 3.2). Another benefit of extending the Monge formulation of OT is that we are easily able to construct variants of the Gromov-Monge distance by restricting to subclasses of measure-preserving mappings. We show that these variants generalize metrics for comparison of subsets of $\mathbb{R}^n$ which have appeared previously by showing that Gromov-Monge distance is equivalent to a natural isometry-invariant transport-based distance for Euclidean sets (Theorem 3.3 and the examples in Section 3.4). For example, the continuous Procrustes metric [2] is an OT-based metric which has been used successfully for classification of anatomical surfaces [10], and we show that it can be viewed as a restriction of a variant of Gromov-Monge distance.

We obtain polynomial-time computable lower bounds on Gromov-Monge distance that parallel existing lower bounds for the Gromov-Wasserstein distance [26]. These bounds involve the so called local shape distributions: given an mm-space $\mathcal{X} = (X, d_X, \mu_X)$, the local shape distribution of $X$ is a function $d_{hX} : X \to P_1(\mathbb{R})$ where $x \mapsto (d_X(x, \cdot))_{#\mu_X}$. When the underlying distance is the Gromov-Wasserstein distance—and the matching between shapes is done via measure couplings—these bounds are known to fail to discriminate between certain shapes. Due to the additional structure imposed by measure preserving maps, we are able prove that when the underlying metric is the Gromov-Monge distance, these bounds based on local distribution distance can be upgraded to actual metrics on certain classes of shapes. These classes of shapes include metric trees (Theorem 6.1). Metric trees are metric graphs (that is, geodesic metric spaces with Hausdorff dimension 1) whose hyperbolicity [12] is zero.

In order to prove this strengthening of the lower bound we establish a connection to a distance between metric trees first considered in the topological data analysis community called the interleaving distance [28, 1].

Along the way, by suitably adapting counterexamples to a conjecture by Blaschke [23], we give counterexamples to a conjecture made by Brinkmann and Olver in [11] about the discriminatory power of the so called global shape distributions. These counterexamples encompass planar curves as well as 3D shapes. The global shape distribution $dH_{\mathcal{X}}$ of the mm-space $\mathcal{X}$ is the average of the local shape distribution $d_{h\mathcal{X}}$: for every $A \subset \mathbb{R}$ measurable,

$$dH_{\mathcal{X}}(A) := \int_X d_{h\mathcal{X}}(x)(A) \mu_X(dx).$$

The paper is organized as follows. The definition of Gromov-Monge distance is given in Section 2 where it is shown to be an extended quasi-metric on $\mathcal{M}^m$. Various alternative representations of Gromov-Monge distance are explored in Section 3. Comparisons with Gromov-Wasserstein distance and the continuous Procrustes metric are also provided therein. Section 4 includes definitions of the local and global shape distributions and treats the derivation of lower bounds on Gromov-Monge distance in terms of these invariants. The efficacy of these lower bounds is explored in Sections 5 and 6 for plane curves and metric trees, respectively.

2 Gromov-Monge Quasi-Metrics

2.1 Preliminary Definitions

Let $\mathcal{M}$ denote the collection of isometry classes of compact metric spaces $(X, d_X)$—we will refer to a metric space by its underlying set $X$ when it is clear that a metric has been fixed. The classical notion of Hausdorff distance between compact subsets of an ambient metric space $Z$ can be vastly generalized.
to give a metric on $M$. The \textit{Gromov-Hausdorff metric} $d_{\text{GH}}$ is defined by

$$d_{\text{GH}}(X, Y) = \inf_{Z, \phi_X, \phi_Y} d_H^Z(\phi_X(X), \phi_Y(Y)),$$

where the infimum is taken over all ambient metric spaces $(Z, d_Z)$ and isometric embeddings $\phi_X : X \to Z$ and $\phi_Y : Y \to Z$, and where $d_H^Z$ denotes Hausdorff distance in $Z$.

The Gromov-Hausdorff distance can be reformulated in several ways, and we will make particular use of a reformulation in terms of the distortion of a correspondence. For compact metric spaces $X$ and $Y$, let $\Gamma_{X,Y} : X \times Y \times X \times Y \to \mathbb{R}$ denote the \textit{distortion map}, defined by

$$\Gamma_{X,Y}(x, y, x', y') = |d_X(x, x') - d_Y(y, y')|.$$  

A \textit{correspondence} between sets $X$ and $Y$ is a subset $R \subset X \times Y$ such that the coordinate projection maps $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ give surjections when restricted to $R$. The set of all correspondences between $X$ and $Y$ is denoted $\mathcal{R}(X, Y)$. We then have the following reformulation of Gromov-Hausdorff distance [12]:

$$2d_{\text{GH}}(X, Y) = \inf_{R \in \mathcal{R}(X, Y)} \sup_{(x, y, x', y') \in R} \Gamma_{X,Y}(x, y, x', y').$$  

One might notice that the righthand side of (2) takes the form of an infimum of instances of an $L^\infty$ norm of the function $\Gamma_{X,Y}$. It is tempting to relax this expression to an $L^p$ norm, but in order to do so we must consider more general spaces. Let $X = (X, d_X, \mu_X)$ denote a \textit{metric measure space} (mm-space)—that is, $(X, d_X)$ is a compact metric space and $\mu_X$ is a Borel probability measure on $X$. For convenience, we always assume that $\text{supp}[\mu_X] = X$. Let $\mathcal{P}_1(X)$ denote the set of all fully-supported Borel probability measures on $X$. For a measurable map $\phi : X \to Y$ between metric spaces spaces and a measure $\mu_X$ on $X$, we will use the notation $\phi\#\mu_X$ for the \textit{pushforward} of $\mu_X$ by $\phi$. This is defined on a Borel subset $A \subset Y$ by

$$\phi\#\mu_X (A) = \mu_X (\phi^{-1}(A)).$$

An \textit{isomorphism} between mm-spaces $X$ and $Y$, is a measure-preserving map $\phi : X \to Y$ which is also a metric isometry: i.e., $\phi\#\mu_X = \mu_Y$ and $d_X = d_Y \circ (\phi, \phi)$. When such an isomorphism exists, we write $X \approx Y$. The collection of isomorphism classes of metric measure spaces will be denoted $\mathcal{M}^m$.

For mm-spaces $X$ and $Y$, define

$$\mathcal{U}(\mu_X, \mu_Y) = \{ \mu \in \mathcal{P}_1(X \times Y) \mid (\pi_X)\#\mu = \mu_X \text{ and } (\pi_Y)\#\mu = \mu_Y \}$$

to be the set of all probability measures on $X \times Y$ with marginals $\mu_X$ and $\mu_Y$, which are called \textit{couplings} between $\mu_X$ and $\mu_Y$. Note that the product measure $\mu_X \otimes \mu_Y$ is always in $\mathcal{U}(\mu_X, \mu_Y)$ so that the set of couplings is never empty. In a similar vein, let

$$\mathcal{T}(\mu_X, \mu_Y) = \{ \phi : X \to Y, \phi\#\mu_X = \mu_Y \}$$

denote the set of all measure-preserving maps between $X$ and $Y$. Notice that it could be that $\mathcal{T} = \emptyset$. Indeed, there is never a measure-preserving map from the mm-space containing a single point to an mm-space with larger cardinality. Given any $\phi \in \mathcal{T}(\mu_X, \mu_Y)$, one can consider the probability measure $\mu_\phi$ on $X \times Y$ given by

$$\mu_\phi = (\text{id} \times \phi)\#\mu_X.$$  

It is straightforward to check that $\mu_\phi \in \mathcal{U}(\mu_X, \mu_Y)$.
2.2 Gromov-Monge Distance

One notion of an $L^p$ relaxation of (2) is given by the Gromov-Wasserstein $p$-distance $d_{GW,p}$, defined on mm-spaces $\mathcal{X}$ and $\mathcal{Y}$ by

$$d_{GW,p}(\mathcal{X}, \mathcal{Y}) = \inf_{\mu \in U(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \| \Gamma_{\mathcal{X}, \mathcal{Y}} \|_{L^p(\mu \otimes \mu)}$$

$$= \inf_{\mu \in U(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \left( \int_{X \times Y \times X \times Y} \left| d_X(x, x') - d_Y(y, y') \right|^p \mu \otimes \mu (dx \times dy \times dx' \times dy') \right)^{1/p}.$$

Gromov-Wasserstein distance was introduced and shown to be a metric on $\mathcal{M}^m$ in [26].

As was discussed in the introduction, the aim of many applications in imaging and shape classification is to obtain a registration of two objects via a mapping from one object to another. Gromov-Wasserstein distance defines a metric on $\mathcal{M}^m$, but it is typically realized by a measure coupling and the vital registration map is not obtained. In light of this, we introduce a variant of Gromov-Wasserstein which restricts the constraint set to only consider measure-preserving mappings. The Gromov-Monge $p$-distance between mm-spaces $\mathcal{X}$ and $\mathcal{Y}$ is the quantity

$$d_{GM,p}(\mathcal{X}, \mathcal{Y}) = \inf_{\phi \in T(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \| \Gamma_{\mathcal{X}, \mathcal{Y}} \|_{L^p(\mu_0 \otimes \mu_0)}$$

$$= \inf_{\phi \in T(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \left( \int_{X \times X} \left| d_X(x, x') - d_Y(\phi(x), \phi(x')) \right|^p \mu_X \otimes \mu_X (dx \times dx') \right)^{1/p},$$

with the understanding that if $T(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) = \emptyset$ then $d_{GM,p}(\mathcal{X}, \mathcal{Y}) = \infty$.

Remark 2.1. Clearly, the bound $d_{GW,p}(\mathcal{X}, \mathcal{Y}) \leq d_{GM,p}(\mathcal{X}, \mathcal{Y})$ always holds.

Referring to $d_{GM,p}$ as a distance is a slight abuse of terminology. Recall that a quasi-metric on a set $S$ is a function $S \times S \to \mathbb{R}$ satisfying all of the usual axioms of a metric except for symmetry. An extended quasi-metric is permitted to assign distance $\infty$ to some pairs of points.

Theorem 2.1. For any $p \geq 1$ the function $d_{GM,p}$ defines an extended quasi-metric on $\mathcal{M}^m$ up to isomorphism.

Proof. Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be mm-spaces. It follows easily from the definition of $d_{GM,p}$, Remark 2.1 and the fact $d_{GW,p}$ is a metric on $\mathcal{M}^m$ that $d_{GM,p}(\mathcal{X}, \mathcal{Y}) \geq 0$, with $d_{GM,p}(\mathcal{X}, \mathcal{Y}) = 0$ if and only if $\mathcal{X} \approx \mathcal{Y}$.

It only remains to show that $d_{GM,p}$ satisfies the triangle inequality $d_{GM,p}(\mathcal{X}, \mathcal{Z}) \leq d_{GM,p}(\mathcal{X}, \mathcal{Y}) + d_{GM,p}(\mathcal{Y}, \mathcal{Z})$. If $d_{GM,p}(\mathcal{X}, \mathcal{Y}) = \infty$, then $T(\mu_{\mathcal{X}}, \mu_{\mathcal{Z}}) = \emptyset$ and it follows that either $T(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) = \emptyset$ or $T(\mu_{\mathcal{Y}}, \mu_{\mathcal{Z}}) = \emptyset$, whence one of $d_{GM,p}(\mathcal{X}, \mathcal{Y})$ or $d_{GM,p}(\mathcal{Y}, \mathcal{Z})$ is infinity. The triangle inequality therefore holds in this case. The triangle inequality follows trivially if $d_{GM,p}(\mathcal{X}, \mathcal{Y})$ or $d_{GM,p}(\mathcal{Y}, \mathcal{Z})$ is infinite, so let us assume that all distances are finite. In this case we have

$$d_{GM,p}(\mathcal{X}, \mathcal{Z}) = \inf_{\phi \in T(\mu_{\mathcal{X}}, \mu_{\mathcal{Z}})} \| \Gamma_{\mathcal{X}, \mathcal{Z}} \|_{L^p(\mu_0 \otimes \mu_0)}$$

$$\leq \inf_{\phi_1 \in T(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \| \Gamma_{\mathcal{X}, \mathcal{Y}} \|_{L^p(\mu_0 \otimes \mu_0)} \| \phi \circ \phi_1 \|_{L^p(\mu_0 \otimes \mu_0)}$$

$$\leq \inf_{\phi_1 \in T(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \left( \| \Gamma_{\mathcal{X}, \mathcal{Y}} \|_{L^p(\mu_0 \otimes \mu_0)} + \| \Gamma_{\mathcal{Y}, \mathcal{Z}} \|_{L^p(\mu_0 \otimes \mu_0)} \right)^{1/p}$$

$$= \inf_{\phi_1 \in T(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \| \Gamma_{\mathcal{X}, \mathcal{Y}} \|_{L^p(\mu_0 \otimes \mu_0)} + \inf_{\phi_2 \in T(\mu_{\mathcal{Y}}, \mu_{\mathcal{Z}})} \| \Gamma_{\mathcal{Y}, \mathcal{Z}} \|_{L^p(\mu_0 \otimes \mu_0)}$$

$$= d_{GM,p}(\mathcal{X}, \mathcal{Y}) + d_{GM,p}(\mathcal{Y}, \mathcal{Z}).$$

The estimate (3) follows from the fact that

$$\| \Gamma_{\mathcal{X}, \mathcal{Z}} \|_{L^p(\mu_0 \otimes \mu_0)} \leq \| \Gamma_{\mathcal{X}, \mathcal{Y}} \|_{L^p(\mu_0 \otimes \mu_0)} + \| \Gamma_{\mathcal{Y}, \mathcal{Z}} \|_{L^p(\mu_0 \otimes \mu_0)}$$
for any fixed $\phi = \phi_2 \circ \phi_1$ with $\phi_1 \in T(\mu_X, \mu_Y)$ and $\phi_2 \in T(\mu_Y, \mu_Z)$. This holds by the definition of $\mu_\phi$, the Minkowski inequality and the fact that

$$\Gamma_{X,Z}(x, z, x', z') \leq \Gamma_{X,Y}(x, y, x', y') + \Gamma_{Y,Z}(y, z, y', z')$$

for all $x, x', y, y', z, z'$. \hfill \qed

**Example 2.1** ($d_{GM,p}$ is Not Symmetric). Let $X = \{u, v\}$ be endowed with empirical measure $\delta_u = \frac{1}{2}$ and metric determined by $d_X(u, v) = 1$. Let $Y = \{y\}$ with measure $\delta_y = 1$. Then the constant map $X \to Y$ is measure-preserving, but neither of the two possible maps $Y \to X$ preserves measure (we cannot split mass). Thus $d_{GM,p}(X, Y) = 1/2^{1/p}$ while $d_{GM,p}(Y, X) = \infty$

Even if finite mm-spaces have the same cardinality, it is possible that the set of measure-preserving transformations between them is empty. For example, take $X$ as above and let $Z = \{y, z\}$ denote the space with $Z = \{y, z\}$, $d_Z(y, z) = 1$ and $\mu_Z$ the measure with weights $\delta_y = 1/4$ and $\delta_z = 3/4$.

In applications, mm-spaces under consideration frequently arise by taking finite samples from continuous spaces and assigning uniform weights. The next proposition shows that $d_{GM,p}$ defines an extended metric when restricted to the class of finite mm-spaces of a fixed cardinality.

**Proposition 2.1.** Let $X$ and $Y$ be mm-spaces such that $X$ and $Y$ are finite sets. If $|X| = |Y|$ then $d_{GM,p}(X, Y) = d_{GM,p}(Y, X)$. Otherwise, one of $d_{GM,p}(X, Y)$ or $d_{GM,p}(Y, X)$ must be infinity and symmetry only holds when $d_{GM,p}(X, Y) = d_{GM,p}(Y, X) = \infty$.

**Proof.** A measure-preserving map $\phi : X \to Y$ between finite mm-spaces with fully supported measures must be surjective. Suppose that $|X| = n$ and $|Y| = m$ and assume without loss of generality that $n \geq m$. If $n > m$, then there is no measure-preserving map $Y \to X$, whence $d_{GM,p}(Y, X) = \infty$, and this proves the second claim.

On the other hand, suppose $n = m$. Let $\delta_{x_j}$ denote the weight assigned to $x_j \in X$ and $\delta_{y_j}$ the weight assigned to $y_j \in Y$. The sets $T(\mu_X, \mu_Y)$ and $T(\mu_Y, \mu_X)$ are both empty or both non-empty, and in the former case we have $d_{GM,p}(X, Y) = d_{GM,p}(Y, X) = \infty$. Supposing that a measure-preserving map $\phi : X \to Y$ exists, it is a bijection such that $\delta_{\phi(x_j)} = \delta_{x_j}$. A straightforward calculation then shows that the inverse of the minimizer realizing $d_{GM,p}(X, Y)$ is the minimizer realizing $d_{GM,p}(Y, X)$. \hfill \qed

At the other extreme, $d_{GM,p}$ induces a true metric on a broad class of mm-spaces with infinite cardinality. Let $\mathcal{M}_\infty^m$ denote the collection of isomorphism classes of mm-spaces $X = (X, d_X, \mu_X)$ such that $X$ is uncountable, $(X, d_X)$ is separable and $\mu_X$ has no atoms.

**Proposition 2.2.** For $X$ and $Y$ representing elements of $\mathcal{M}_\infty^m$, $d_{GM,p}(X, Y)$ is finite. It follows that

$$(X, Y) \mapsto \max \{d_{GM,p}(X, Y), d_{GM,p}(Y, X)\}$$

defines a genuine metric on $\mathcal{M}_\infty^m$.

**Proof.** Let $X$ and $Y$ represent elements of $\mathcal{M}_\infty^m$ and let $I$ denote the interval $[0, 1]$ endowed with Lebesgue measure. A classical result from real analysis says that there exists a bijection $\phi_X : X \to I$ such that both $\phi_X$ and $\phi_X^{-1}$ are measure-preserving \cite[Chapter 5, Theorem 16]{31}. Likewise, there is a map $\phi_Y : Y \to I$ with the same properties. Any measure-preserving map $\psi : I \to I$ (e.g. the identity map) defines a measure-preserving map

$$\phi_Y^{-1} \circ \psi \circ \phi_X : X \to Y.$$

This implies that the set $T(\mu_X, \mu_Y)$ is nonempty, and it follows that $d_{GM,p}$ is finite. \hfill \qed
3 Alternative Representations

3.1 Gromovization of the Mass Transport Problem

We can similarly define $L^p$ versions of the original embedding formulation of Gromov-Hausdorff distance [1]. These definitions are more clearly tied to ideas from optimal transport. Before making our definitions, we more precisely recall the two formulations of the optimal transport problem (see [36] for a general reference).

Given a compact metric space $(Z,d_Z)$ and Borel probability measures $\alpha$ and $\beta$ on $Z$, the Kantorovich formulation of the optimal transport problem seeks the quantity

$$d_{W,p}^Z(\alpha,\beta) = \left(\inf_{\mu \in U(\alpha,\beta)} \int_{Z \times Z} (d_Z(z,z'))^p \mu(dz \times dz')\right)^{1/p},$$

where $p \geq 1$. The function $d_{W,p}$ has come to be known as the Wasserstein metric on $\mathcal{P}_1(Z)$. As shown by Sturm [34,35], the Gromov-Hausdorff construction can be adapted to give a metric on $\mathcal{M}^m$, which we denote

$$d_{GM,p}^{emb}(\mathcal{X},\mathcal{Y}) = \inf_{Z,\phi_X,\phi_Y} d_{W,p}^Z((\phi_X)_#\mu_X, (\phi_Y)_#\mu_Y),$$

where the infimum is taken over all isometric embeddings $\phi_X : X \to Z$ and $\phi_Y : Y \to Z$ into some compact metric space $Z$.

On the other hand, the original Monge formulation of optimal transport seeks the quantity

$$d_{M,p}^Z(\alpha,\beta) = \left(\inf_{\phi \in \mathcal{T}(\alpha,\beta)} \int_Z (d_Z(\phi(z)))^p \alpha(dz)\right)^{1/p},$$

which is taken as $\infty$ when $\mathcal{T}(\alpha,\beta) = \emptyset$. Of course, $d_{M,p}^Z \geq d_{W,p}^Z$. We analogously define a new extended quasi-metric $d_{GM,p}^{emb}$ on $\mathcal{M}^m$ by

$$d_{GM,p}^{emb}(\mathcal{X},\mathcal{Y}) := \inf_{Z,\phi_X,\phi_Y} d_{M,p}^Z((\phi_X)_#\mu_X, (\phi_Y)_#\mu_Y),$$

where the infimum is once again taken over isometric embeddings into a common ambient metric space. These definitions extend to the case $p = \infty$ as usual.

**Theorem 3.1.** For all $\mathcal{X},\mathcal{Y} \in \mathcal{M}^m$ and for any $p \in [1,\infty]$,

$$\frac{1}{2} d_{GM,p}(\mathcal{X},\mathcal{Y}) \leq d_{GM,p}^{emb}(\mathcal{X},\mathcal{Y}).$$

(5)

In the case that $p = \infty$, we have

$$\frac{1}{2} d_{GM,\infty}(\mathcal{X},\mathcal{Y}) = d_{GM,\infty}^{emb}(\mathcal{X},\mathcal{Y}).$$

(6)

**Proof.** First we will show that $d_{GM,p}(\mathcal{X},\mathcal{Y}) = \infty$ if and only if $d_{GM,p}^{emb}(\mathcal{X},\mathcal{Y}) = \infty$. This follows because we can always isometrically embed $X$ and $Y$ into the metric space $Z = X \sqcup Y$ with metric $d_Z|_{X \times X} = d_X$, $d_Z|_{Y \times Y} = d_Y$, and $d_Z(x,y) = \max\{\text{diam}(X),\text{diam}(Y)\}$. Then there exists a measure-preserving map $X \to Y$ if and only if there exists a map $Z \to Z$ taking the pushforward of $\mu_X$ to the pushforward of $\mu_Y$.

Now we will assume that both Gromov-Monge distances are finite and show that (5) holds for $p \in [1,\infty]$. This is equivalent to showing that whenever $d_{GM,p}^{emb}(\mathcal{X},\mathcal{Y}) < r$, there exists $\phi \in \mathcal{T}(\mu_X,\mu_Y)$ such that $\|\Gamma_{\phi}\|_{L^p(\mu \times \mu)} \leq 2r$. If $d_{GM,p}^{emb}(\mathcal{X},\mathcal{Y}) < r$ then we can find isometries $\phi_X : X \to Z$ and $\phi_Y : Y \to Z$ into a metric space $Z$ such that $d_{M,p}^Z((\phi_X)_#\mu_X, (\phi_Y)_#\mu_Y) < r$. We may as well assume
that $X, Y \subset Z$ and that $\mu_X$ and $\mu_Y$ are measures on $Z$ with $\text{supp}[\mu_X] = X$ and $\text{supp}[\mu_Y] = Y$. By definition of $d_{GM,p}^Z$, there exists $\phi \in T(\mu_X, \mu_Y)$ such that

$$\|d_Z(\cdot, \phi(\cdot))\|_{L^p(\mu_X)} < r,$$

where $d_Z(\cdot, \phi(\cdot))$ denotes the map from $Z$ to $\mathbb{R}$ given by $z \mapsto d_Z(z, \phi(z))$. Now note that for all $x, x' \in X$, the triangle inequality in $Z$ implies that

$$|d_Z(x, x') - d_Z(\phi(x), \phi(x'))| \leq d_Z(x, \phi(x)) + d_Z(x', \phi(x')).$$

Putting this together with the triangle inequality for the $L^p$ norm, we have

$$\|\Gamma_{X,Y}\|_{L^p(\mu_0 \otimes \mu_0)} \leq 2\|d_Z(\cdot, \phi(\cdot))\|_{L^p(\mu_X)} < 2r.$$

This establishes (5).

Now we wish to show that $\frac{1}{2}d_{GM,\infty}(X,Y) \geq d_{GM,\infty}^{emb}(X,Y)$. Let $\phi_0 : X \rightarrow Y$ be any measure preserving map with

$$\|\Gamma_{X,Y}\|_{L^\infty(\mu_0 \otimes \mu_0)} = \sup_{x,x' \in X} \Gamma_{X,Y}(x,x', \phi_0(x), \phi_0(x')) = 2r.$$

The claim follows if we are able to construct a metric space $(Z, d_Z)$ and isometric embeddings $\phi_X$ and $\phi_Y$ such that

$$d_{GM,\infty}^Z((\phi_X)\#\mu_X, (\phi_Y)\#\mu_Y) \leq r.$$

Let $Z$ denote the disjoint union of $X$ and $Y$ and define a metric $d_Z$ on $Z$ by setting $d_Z|_{X \times X} = d_X$, $d_Z|_{Y \times Y} = d_Y$ and

$$d_Z(x, y) = d_Z(y, x) = \inf_{x' \in X} \{d_X(x, x') + r + d_Y(\phi_0(x'), y)\}$$

for any $x \in X$ and $y \in Y$. Then

$$d_{GM,\infty}^Z(\mu_X, \mu_Y) = \inf_{\phi \in T(\mu_X, \mu_Y)} \sup_{x \in X} d_Z(x, \phi(x)) \leq \sup_{x \in X} d_Z(x, \phi_0(x))$$

$$= \sup_{x \in X} \inf_{x' \in X} \{d_X(x, x') + r + d_Y(\phi_0(x'), \phi_0(x))\} = r.$$

This completes the proof.

**Example 3.1** ($d_{GM,p}$ and $d_{emb}^{GM,p}$ are Not BiLipschitz Equivalent). Consider the family of mm-spaces $\Delta_n$. Each $\Delta_n$ consists of the space $X_n = \{1, \ldots, n\}$ with metric $d_n(i, j) = 1 - \delta_{ij}$ and measure $\nu_n$ defined by $\nu_n(i) = 1/n$. For $p < \infty$, $d_{GM,\infty}^{emb}(\Delta_{2n}, \Delta_n) \geq d_{GM,p}^{emb}(\Delta_{2n}, \Delta_n) \geq \frac{1}{4}$, where the former bound holds generally and the latter is shown in [24 Claim 5.3]. On the other hand, for $p < \infty$, $d_{GM,p}(\Delta_{2n}, \Delta_n) = 1/(2n)^{1/p}$. This follows from the fact that $T(\nu_{2n}, \nu_n)$ is simply the set of 2-to-1 maps from $\{1, \ldots, 2n\}$ to $\{1, \ldots, n\}$ and all such maps have the same Monge cost, so that the quantity is obtained by a direct calculation. It follows that the quasi-metrics $d_{GM,p}$ and $d_{emb}^{GM,p}$ are not bi-Lipschitz equivalent for any $p \in [1, \infty)$. 

7
3.2 Gromov-Wasserstein Distance as a Gromov-Monge Distance

A pseudo-metric space is a pair \((X, d_X)\) such that the pseudo-metric \(d_X : X \times X \to \mathbb{R}\) satisfies all of the axioms of a metric, except that it is possible for \(d_X(x, x') = 0\) for distinct \(x, x' \in X\). The pseudo-metric defines a topology on \(X\) by taking a basis of open balls; the topology is non-Hausdorff if \(d_X\) is not actually a metric. A pseudo-metric measure space (pmm-space) is a triple \(X = (X, d_X, \mu_X)\), where \((X, d_X)\) is a compact pseudo-metric space and \(\mu_X\) is a Borel probability measure with full support. The definition of Gromov-Monge distance extends to pseudo-metric measure spaces without modification.

Example 3.2 (Pullback Pseudometric). Let \((X, d_X)\) be a metric space, let \(Z\) be a set and let \(f : Z \to X\) be a function. The pullback of \(d_X\) by \(f\) is the map \(f^*d_X : Z \times Z \to \mathbb{R}\) defined by

\[
(f^*d_X)(z, z') = d_X(f(z), f(z')).
\]

It is easy to check that \(f^*d_X\) satisfies the axioms of a pseudo-metric.

Let \(X\) be a mm-space. A mass splitting of \(X\) is a pmm-space \(Z\) such that there exists a measure-preserving map \(\pi : Z \to X\) with the property that \(d_Z = \pi^*d_X\).

Theorem 3.2. Let \(X\) and \(Y\) be metric measure spaces. Then

\[
d_{GW,p}(X, Y) = \inf_{Z \to X} d_{GM,p}(Z, Y),
\]

where the infimum is taken over mass-splittings of \(X\).

Proof. Given a measure coupling \(\mu\) of \(\mu_X\) and \(\mu_Y\), define a pmm-space \(Z\) by setting \(Z = X \times Y, \mu_Z = \mu, \pi = \pi_X\) (projection \(X \times Y \to X\) onto the first coordinate) and \(d_Z = \pi^*d_X\). Then \(\phi = \pi_Y : Z \to Y\) is a measure-preserving map with

\[
\|\Gamma_{Z,Y}\|_{L^p(\mu \otimes \mu)}^p = \int_{Z \times Z} \Gamma_{Z,Y}(z, \phi(z), z', \phi(z'))^p \mu_Z(\text{dz} \times \text{dz}')
\]

\[
= \int_{X \times Y} \int_{X \times Y} \Gamma_{Z,Y}((x, y), (x', y'), y'^{\prime})^p \mu_X(\text{dx} \times \text{dy} \times \text{dx}' \times \text{dy}')
\]

\[
= \int_{X \times Y} \int_{X \times Y} \Gamma_{X,Y}(x, y, x', y')^p \mu(\text{dx} \times \text{dy} \times \text{dx'} \times \text{dy'}),
\]

where the last line follows by \(d_Z = \pi_X^*d_X\). We conclude that

\[
d_{GW,p}(X, Y) \geq \inf_{Z \to X} d_{GM,p}(Z, Y).
\]

Conversely, let \(Z\) satisfy the conditions and let \(\phi : Z \to Y\) be a measure-preserving map. We define a probability measure \(\mu\) on \(X \times Y\) as

\[
\mu = (\pi \times \phi)^\#Z,
\]

where \(\pi \times \phi : Z \to X \times Y\) is the map taking \(z\) to \((\pi(z), \phi(z))\). Then \(\mu\) defines a measure coupling of \(\mu_X\) and \(\mu_Y\). Indeed, for any set \(A \subset X\),

\[
(\pi_X)^\#\mu(A) = \mu(\pi_X^{-1}(A)) = (\pi \times \phi)^\#Z(A \times Y) = \mu_Z((\pi \times \phi)^{-1}(A \times Y))
\]

\[
= \mu_Z(\pi^{-1}(A)) = \pi^\#Z(A) = \mu_X(A)
\]

8
and \((\pi_Y)_\# \mu = \mu_Y\) follows by a similar argument. Consider the following calculation (we suppress all function arguments to condense notation):

\[
\int_{(X \times Y)^2} (\Gamma_{X,Y})^p \mu \otimes \mu = \int_{(X \times Y)^2} (\Gamma_{X,Y})^p (\pi \times \phi)_\# \mu_Z \otimes (\pi \times \phi)_\# \mu_Z
\]

(7)

\[
= \int_{(X \times Y)^2} (\Gamma_{X,Y})^p ((\pi \times \phi) \times (\pi \times \phi))_\# \mu_Z \otimes \mu_Z
\]

(8)

\[
= \int_{Z \times Z} (\Gamma_{Z,Y} \circ ((\pi \times \phi) \times (\pi \times \phi)))^p \mu_Z \otimes \mu_Z
\]

(9)

Equality (7) follows from the fact that \((\pi \times \phi)_\# \mu_Z \otimes (\pi \times \phi)_\# \mu_Z\) and \(((\pi \times \phi) \times (\pi \times \phi))_\# \mu_Z \otimes \mu_Z\) define equivalent measures on \(X \times Y \times X \times Y\). Equality (8) follows from the change-of-variables property of pushforward measures. The final equality (9) follows once again from the change-of-variables property and the fact that

\[
\Gamma_{X,Y}(\pi(z), \phi(z), \pi(z'), \phi(z')) = \Gamma_{Z,Y}(z, \phi(z), z', \phi(z')),
\]

since \(d_Z = \pi^* d_X\). This calculation shows that

\[
d_{GW,p}(X, Y) \leq \inf_{Z \to X} d_{GM,p}(Z, Y)
\]

and the proof is therefore complete. \(\square\)

### 3.3 Gromov-Monge Distances for Euclidean Spaces

A *Euclidean mm-space* is an mm-space \(X\) such that \(X \subset \mathbb{R}^n\) for some \(n\) and \(d_X\) is the restriction of the Euclidean distance (we do not impose any extra conditions on \(\mu_X\)). For Euclidean mm-spaces \(X\) and \(Y\) with the same ambient space \(\mathbb{R}^n\), we can consider the *isometry-invariant Monge p-distance*

\[
d_{M,p,iso}^{\mathbb{R}^n}(X, Y) = \inf_{T \in E(n)} \left( \inf_{\phi \in T(X, \mu_X)} \int_X \|T(x) - \phi(x)\|^p \mu_X(dx) \right)^{1/p},
\]

where we use \(E(n)\) to denote the group of Euclidean isometries and \(\| \cdot \|\) for the Euclidean norm.

**Theorem 3.3.** Let \(X\) and \(Y\) be Euclidean mm-spaces in the same ambient space \(\mathbb{R}^n\), let \(p \geq 1\) and let \(M = \max\{\text{diam}(X), \text{diam}(Y)\}\). Then

\[
d_{GM,p}^{\text{emb}}(X, Y) \leq d_{M,p,iso}^{\mathbb{R}^n}(X, Y) \leq M^{1/2} \cdot c_n \cdot d_{GM,p}^{\text{emb}}(X, Y)^{1/2},
\]

where \(c_n\) is a constant depending only on \(n\).

**Remark 3.1.** It is easy to show that, for any mm-spaces \(X\) and \(Y\), if \(d_{GM,p}^{\text{emb}}(X, Y) < \infty\) then

\[
d_{GM,p}^{\text{emb}}(X, Y) \leq \max\{\text{diam}(X), \text{diam}(Y)\}.
\]

The proof idea is essentially the same as [25] Theorem 4, except that we require a technical lemma which is specific to our setup. We are also able to obtain slightly better bounds.

**Lemma 3.1.** Let \(X\) and \(Y\) be mm-spaces and let \(M = \max\{\text{diam}(X), \text{diam}(Y)\}\). If

\[
d_{GM,p}^{\text{emb}}(X, Y) \leq \epsilon \cdot M
\]

for \(\epsilon \leq 1\), then there exist \(X_\epsilon \subset X\), \(Y_\epsilon \subset Y\), \(R_\epsilon \in \mathcal{R}(X_\epsilon, Y_\epsilon)\), and a measure-preserving map \(\phi : X \to Y\) satisfying \(\phi(X_\epsilon) = Y_\epsilon\) such that

\[
\min\{\mu_X(X_\epsilon), \mu_Y(Y_\epsilon), \mu_\phi(R_\epsilon)\} \geq 1 - \epsilon^{p/2}
\]

(10)

and

\[
d_{GH}(X_\epsilon, Y_\epsilon) \leq \sup_{(x,y),(x',y') \in R_\epsilon} \Gamma_{X,Y}(x, y, x', y') \leq \epsilon^{1/2} M.
\]

(11)
Proof. Without loss of generality, suppose that $X$ and $Y$ are subsets of an ambient metric space $(Z, d_Z)$ and let $\phi \in \mathcal{S}$ such that

$$\int_Z d_Z(z, \phi(z))^p \mu_X(dz) \leq \epsilon^p M^p.$$  

Define a subset $R_\epsilon \subset X \times Y$ by

$$R_\epsilon = \{(x, y) \in X \times Y \mid y = \phi(x) \text{ and } d_Z(x, y) \leq \epsilon^{1/2} M\}.$$  

We then define $X_\epsilon = \pi_X(R_\epsilon)$ and $Y_\epsilon = \pi_Y(R_\epsilon)$, so that $R_\epsilon \in \mathcal{R}(X_\epsilon, Y_\epsilon)$. Also note that $\phi(X_\epsilon) = Y_\epsilon$. A short calculation using the triangle inequality shows that $R_\epsilon$ satisfies (11). Moreover,

$$\epsilon^p M^p \geq \int_Z d_Z(z, \phi(z))^p \mu_X(dz) \geq \int_{X \setminus X_\epsilon} d_Z(z, \phi(z))^p \mu_X(dz) \geq \epsilon^{p/2} M^p \mu_X(X \setminus X_\epsilon).$$

Rearranging, we obtain $\mu_X(X_\epsilon) \geq 1 - \epsilon^{p/2}$. Since $Y_\epsilon = \phi(X_\epsilon)$ and $\phi$ is measure-preserving, $\mu_Y(Y_\epsilon) \geq 1 - \epsilon^{p/2}$ as well. Finally, we have

$$\mu_X(X_\epsilon) = (\pi_X)_\# \mu_{\phi}(\pi_X(R_\epsilon)) = \mu_{\phi}(\pi_X^{-1} \circ \pi_X(R_\epsilon)) = \mu_{\phi}(R_\epsilon),$$

and (10) is satisfied. \qed

Proof of Theorem 3.3. The inequality on the left is obvious, so let us consider the inequality on the right. If either quantity is infinite, then both are, so assume not. By Remark 3.1, we suppose without loss of generality that $d_{\mathcal{M}^{\text{emb}}}(X, Y) = \epsilon M$ for some $\epsilon \leq 1$. Let $X_\epsilon, Y_\epsilon, \phi$ and $R_\epsilon$ be as in Lemma 3.1 and let $R_\epsilon^c = (X \times Y) \setminus R_\epsilon$. Then $d_{\mathcal{G}H}(X_\epsilon, Y_\epsilon) \leq \epsilon^{1/2} M$ and [12] Corollary 7.3.28 implies that there exists a map $\psi : X_\epsilon \to Y_\epsilon$ such that

$$\sup_{x, x' \in X_\epsilon} \Gamma(x, \psi(x), x', \psi(x')) \leq 2\epsilon^{1/2} M$$

and such that $\psi(X_\epsilon)$ is a $2\epsilon^{1/2} M$-net for $Y_\epsilon$. We apply [3, Theorem 2.2] to conclude that there is an isometry $T \in E(n)$ such that $\sup_{x \in X_\epsilon} \|T(x) - \psi(x)\| \leq \epsilon^{1/2} \cdot a_n \cdot M$, where $a_n$ is a constant depending only on $n$.

Applying the triangle inequality and the general inequality $(a + b)^p \leq 2^{p-1} (a^p + b^p)$ (for $a, b \geq 0$), we have

$$d_{\mathcal{M}^{\text{emb}}}(X, Y)^p \leq \int_{R_\epsilon^c} \|T(x) - y\|^p \mu_{\phi}(dx \times dy)$$

$$\quad + 2^{p-1} \int_{R_\epsilon} \|T(x) - \psi(x)\|^p \mu_{\phi}(dx \times dy)$$

$$\quad + 2^{p-1} \int_{R_\epsilon} \|\psi(x) - y\|^p \mu_{\phi}(dx \times dy).$$

We bound each term separately. First note that

$$\|T(x) - y\|^p \leq 2^{p-1} (\|T(x)\|^p + \|y\|^p) \leq 2^p \max_x \|T(x)\|^p, \max_y \|y\|^p = (2M)^p,$$

where we have used isometry invariance of $d_{\mathcal{M}^{\text{emb}}}$ to assume without loss of generality that the circumcenters of $X$ and $Y$ are at the origin in order to obtain the last equality. This implies that

$$\text{(12)} \leq (2M)^p \mu_{\phi}(R_\epsilon^c) \leq 2^p \cdot M^p \cdot \epsilon^{p/2}.$$

The bounds on (13) and (14) follow from our assumptions on $\psi$:

$$\text{(13)} \leq 2^{p-1} \max_x \|T(x) - \psi(x)\|^p \cdot \mu_{\phi}(R_\epsilon) \leq 2^{p-1} \cdot \epsilon^{p/2} \cdot a_n^p \cdot M^p$$

and

$$\text{(14)} \leq 2^{p-1} \max_y \|\psi(x) - y\|^p \cdot \mu_{\phi}(R_\epsilon) \leq 2^{p-1} \cdot \epsilon^{p/2} \cdot a_n^p \cdot M^p.$$
and
\[ \left( 14 \right) \leq 2^{p-1} \cdot \max_{(x,y) \in R_n} \| \psi(x) - y \|^p \mu_\phi(R_n) \leq 2^{p-1} \cdot 2^p \cdot e^{p/2} \cdot M^p. \]

Combining these estimates, we conclude
\[ d_{M,p,iso}^n(X, Y)^p \leq 2^p M^p e^{p/2} (1 + 2^{-1} + a_n^p) \leq 4^p M^p e^{p/2} \max\{2, a_n\}^p, \]
so that
\[ d_{M,p,iso}^n(X, Y) \leq M \cdot c_n \cdot e^{1/2} = M^{1/2} \cdot c_n \cdot d_{GM,p}(X, Y)^{1/2}, \]
with \( c_n = 4 \max\{2, a_n\} \).

### 3.4 Variants of Gromov-Monge Distances

A benefit of using mappings (as opposed to couplings) in the definition of the Gromov-Monge \( p \)-distances is that it makes the definition amenable to restricting to various convenient subclasses of maps. In general for mm-spaces \( X \) and \( Y \), let \( S = S(X, Y) \subset \mathcal{T}(\mu_X, \mu_Y) \) denote some prescribed class of measure-preserving mappings \( \phi : X \to Y \). We then define the restricted Gromov-Monge quasi-\( p \)-distance
\[ d_{GM,p,S}(X, Y) = \inf_{\phi \in S} \| \Gamma_{\chi_X, \chi_Y} \|_{L^p(\mu_\phi \otimes \mu_\phi)}, \]
with the distance taken to be \( \infty \) when \( S = \emptyset \). The proof of Theorem 2.1 still applies to show that \( d_{GM,p,S} \) is an extended quasi-metric on \( M^m \).

**Example 3.3.** Taking \( S \) to be the set of bijective measure-preserving mappings, \( d_{GM,p,S} \) is symmetric and therefore defines an extended metric on \( M^m \). It restricts to a (finite-valued) metric on the space \( M^\infty \) of uncountable, separable, nonatomic mm-spaces.

**Example 3.4.** We could take \( S \) to be the set of continuous measure-preserving mappings. Restricting to the subclass of mm-spaces \( X \) such that \( X \) is a smooth manifold, we could also take \( S \) to be the set of measure-preserving mappings of higher regularity. We will show below (Examples 3.5 and 3.6) that restricting to these sets of mappings allows us to generalize metrics which have appeared previously in the literature.

We also define
\[ d_{GM,p,S}^{emb}(X, Y) = \inf_{Z, \phi_X, \phi_Y} d_{M,p,S}^Z((\phi_X)\#\mu_X, (\phi_Y)\#\mu_Y), \]
where the infimum is taken over isometric embeddings into a compact metric space \( (Z, d_Z) \), and where
\[ d_{M,p,S}^Z(\alpha, \beta) = \left( \inf_{\phi \in S} \int_Z (d_Z(z, \phi(z)))^p \alpha(dz) \right)^{1/p}, \]
with \( S = S(\text{supp}[\alpha], \text{supp}[\beta]) \). Finally, we extend the definition of \( d_{M,p,iso}^{emb} \) analogously to obtain the isometry-invariant quasi-metric \( d_{M,p,iso,S}^{emb} \).

Going through the proofs, one observes that it is possible to extend most of our results to the restricted quasi-metrics. For example, Theorem 3.1 can be adapted to give the general bound
\[ d_{GM,p,S}(X, Y) \leq d_{GM,p,S}^{emb}(X, Y). \]
Likewise, rewriting the arguments and replacing \( \mathcal{T} \) with \( S \), Theorem 3.3 can be extended to show that Euclidean mm-spaces satisfy
\[ d_{GM,p,S}^{emb}(X, Y) \leq d_{M,p,iso,S}^{emb}(X, Y) \leq M^{1/2} \cdot c_n \cdot d_{GM,p,S}^{emb}(X, Y)^{1/2}. \]

(15)
Example 3.5. For smooth surfaces $\mathcal{X}$ and $\mathcal{Y}$ in $\mathbb{R}^3$, the continuous Procrustes distance is given in our notation by $d_{M,2,\text{iso},S}^{\mathcal{X},\mathcal{Y}}$, where $S$ is the collection of continuous measure preserving maps $X \to Y$. Theoretical aspects of this metric are studied in [2], where it is shown that optimal mappings are close to being conformal. The metric was used in [10] to classify anatomical surfaces (shapes of primate teeth), where its effectiveness at classification was shown to be roughly on par with that of a trained morphologist. The bound (15) shows that continuous Gromov-Monge distances can be seen as a generalization of the continuous Procrustes metric.

Example 3.6. The metric $d_{M,2,\text{iso},S}^{\mathcal{X},\mathcal{Y}}$ with $S$ the collection of measure-preserving diffeomorphisms is studied in [20] for applications to 2D image registration. Once again, (15) shows that this is generalized by a variant of Gromov-Monge distance.

4 Lower Bounds

In order to use Gromov-Monge quasi-distances in classification tasks, we seek lower bounds which are efficiently computable. In this section we introduce two such bounds, defined in terms of distance distributions of the metric spaces. The subsequent sections give rigorous results on the effectiveness of these lower bounds at distinguishing metric measure spaces. Distance distributions have appeared frequently as summary shape descriptors, so these effectiveness results are of independent interest.

4.1 Global Distance Distributions

A simple measure-preserving isometry-invariant descriptor of a mm-space $\mathcal{X}$ is its distance distribution

$$H_\mathcal{X} : \mathbb{R}^\geq \to \mathbb{R}^\geq \quad r \mapsto \mu_\mathcal{X} \otimes \mu_\mathcal{X}(\{(x,x') \in X \times X | d_\mathcal{X}(x,x') \leq r\}) .$$

Distance distributions (sometimes referred to as shape distributions or distance histograms) are a standard tool for summarizing metric spaces. They have been used for classification of geometric objects [29, 8] and their mathematical properties have been studied in a variety of contexts [9, 11, 6]. The distance distribution can be used to give a lower bound on $d_{GM,p}$ as follows.

**Proposition 4.1.** For mm-spaces $\mathcal{X}$ and $\mathcal{Y}$ and $p \in [1, \infty)$

$$d_{GM,p}(\mathcal{X},\mathcal{Y}) \geq \left( \int_0^1 |H^{-1}_\mathcal{X}(u) - H^{-1}_\mathcal{Y}(u)| \; du \right)^{1/p} ,$$

where for $u \in [0,1]$,

$$H^{-1}_\mathcal{X}(u) = \inf\{r \geq 0 | H_\mathcal{X}(r) > u\} .$$

For $p = 1$ this simplifies to

$$d_{GM,1}(\mathcal{X},\mathcal{Y}) \geq \int_0^\infty |H_\mathcal{X}(r) - H_\mathcal{Y}(r)| \; dr .$$

**Proof.** This follows immediately from $d_{GM,p}(\mathcal{X},\mathcal{Y}) \geq d_{GW,p}(\mathcal{X},\mathcal{Y})$ and a similar lower bound on Gromov-Wasserstein distances derived in [26, Section 6].

4.2 Local Distance Distributions

The lower bound derived in Proposition 4.1 involves no optimization and one would therefore not expect it to be very tight. In order to achieve a lower bound which involves an optimization component, we
introduce another classical mm-space invariant. The local distance distribution of a mm-space \( X \) is the function

\[
h_X : X \times \mathbb{R}_+ \to [0, 1]
\]

\[
(x, t) \mapsto \mu_X(B_t(x)).
\]

This quantity is related to the global distance distribution by

\[
H_X(r) = \int_X h_X(x, r) \mu_X(dx).
\]

Local distance distributions (and closely related invariants, such as shape contexts) have also appeared frequently in the shape analysis literature (e.g., \([17, 5, 33]\)).

Given \( X, Y \in \mathcal{M}^m \) one defines the following cost function \( c_{X, Y} : X \times Y \to \mathbb{R}_+ \) by

\[
c_{X, Y}(x, y) = \int_0^\infty |h_X(x, t) - h_Y(y, t)| \, dt.
\]

**Proposition 4.2.** Let \( X, Y \in \mathcal{M}^m \). For each \( p \in [1, \infty) \), we have

\[
d_{GM,p}(X, Y) \geq \inf_{\phi \in \mathcal{T}(\mu_X, \mu_Y)} \left( \int_X \left( \int_0^1 |h_X^{-1}(x, u) - h_Y^{-1}(\phi(x), u)|^p \, du \right)^{1/p} \mu_X(dx) \right),
\]

where

\[
h_X^{-1}(x, u) = \inf\{r \geq 0 \mid h_X(x, r) > u\}.
\]

In particular, for \( p = 1 \), this simplifies to

\[
d_{GM,1}(X, Y) \geq \inf_{\phi \in \mathcal{T}(\mu_X, \mu_Y)} \int_X c_{X, Y}(x, \phi(x)) \mu_X(dx).
\]

We require a technical lemma, which appears as [26, Lemma 6.1] and is restated here for convenience. It follows by a change of variables and standard facts about optimal transportation on the real line.

**Lemma 4.1.** Let \( X \) and \( Y \) be mm-spaces, \( f : X \to \mathbb{R} \) and \( g : Y \to \mathbb{R} \) continuous maps and \( \xi : \mathbb{R} \to [0, \infty) \) a convex function. Then

\[
\inf_{\mu \in \mathcal{U}(\mu_X, \mu_Y)} \int_{X \times Y} \xi(f(x) - g(y)) \, \mu(dx \times dy) \geq \int_0^1 \xi(F^{-1}(t) - G^{-1}(t)) \, dt,
\]

where for \( t \geq 0 \),

\[
F(t) = \mu_X\{x \in X \mid f(x) \leq t\} \quad \text{and} \quad G(t) = \mu_Y\{y \in Y \mid g(y) \leq t\}
\]

have generalized inverses

\[
F^{-1}(t) = \inf\{u \in \mathbb{R} \mid F(u) > t\} \quad \text{and} \quad G^{-1}(t) = \inf\{u \in \mathbb{R} \mid G(u) > t\}.
\]

When \( \xi(u) = |u| \), this simplifies to

\[
\inf_{\mu \in \mathcal{U}(\mu_X, \mu_Y)} \int_{X \times Y} |f(x) - g(y)| \, \mu(dx \times dy) \geq \int_0^\infty |F(u) - G(u)| \, du.
\]

**Corollary 4.1.** Let \( X \) and \( Y \) be mm-spaces and let \( p \in [1, \infty) \). For any map \( \phi \in \mathcal{T}(\mu_X, \mu_Y) \) and any fixed points \( x \in X \) and \( y \in Y \), the following estimate holds:

\[
\|\Gamma_{X, Y}(x, y, \cdot, \cdot)\|_{L^p(\mu_\phi)} \geq \left( \int_0^1 |h_X^{-1}(x, u) - h_Y^{-1}(y, u)|^p \, du \right)^{1/p}.
\]

When \( p = 1 \), this simplifies to

\[
\|\Gamma_{X, Y}(x, y, \cdot, \cdot)\|_{L^1(\mu_\phi)} \geq c_{X, Y}(x, y).
\]
Proof. We apply Lemma 4.1 with \( f(\cdot) = d_X(x, \cdot), g(\cdot) = d_Y(y, \cdot) \) and \( \xi(u) = |u|^p \) to obtain
\[
\| \Gamma_{X,Y}(x, y, \cdot) \|^p_{L^p(\mu)} \geq \inf_{\phi \in T(\mu_X, \mu_Y)} \int_{X \times Y} |d_X(x, x') - d_Y(y, y')|^p \mu_\phi(dx \times dy) \\
\geq \inf_{\mu \in \mathcal{L}(\mu_X, \mu_Y)} \int_{X \times Y} |d_X(x, x') - d_Y(y, y')|^p \mu(dx \times dy) \\
\geq \int_0^1 |h_X^{-1}(x, u) - h_Y^{-1}(y, u)|^p du.
\]
The lemma also provides the claimed simplification in the \( p = 1 \) case. \( \square \)

Proof of Proposition 4.2 For any \( \phi \in T(\mu_X, \mu_Y) \),
\[
\| \Gamma_{X,Y} \|^p_{L^p(\mu \otimes \mu)} = \int \int_{X \times Y \times X \times Y} |\Gamma_{X,Y}(x, y, x', y')|^p \mu_\phi(dx \times dy) \mu_\phi(dx' \times dy') \\
\geq \int_{X \times Y} \| \Gamma_{X,Y}(x, y, \cdot, \cdot) \|^p_{L^p(\mu)} \mu_\phi(dx \times dy) \\
\geq \int_{X \times Y} \int_0^1 |h_X^{-1}(x, u) - h_Y^{-1}(y, u)|^p \mu_\phi(dx \times dy) \quad \text{(Corollary 4.1)} \\
= \int_X \int_0^1 |h_X^{-1}(x, u) - h_Y^{-1}(\phi(x), u)|^p \mu_X(dx).
\]
Since \( \phi \) was arbitrary, the general result follows. The simplified expression for \( p = 1 \) follows similarly from Corollary 4.1. \( \square \)

4.3 Distinguishing Power of the Lower Bounds

Given our derived lower bounds on \( d_{GM,p} \), we arrive at a natural question: is it possible for one of our lower bounds on \( d_{GM,p}(X, Y) \) to return the value zero for non-isomorphic spaces \( X \) and \( Y \)? For simplicity, we will focus on the \( p = 1 \) bounds and use the notation
\[
L_H(X, Y) = \int_0^\infty |H_X(r) - H_Y(r)| \, dr
\]
and
\[
L_h(X, Y) = \inf_{\phi \in T(\mu_X, \mu_Y)} \int_X c_{X,Y}(x, \phi(x)) \mu_X(dx).
\]
We say that \( L_H \) (respectively, \( L_h \)) distinguishes spaces in a fixed subclass of mm-spaces if \( L_H(X, Y) = 0 \) (respectively, \( L_h(X, Y) = 0 \)) implies that \( X \approx Y \) for all spaces \( X \) and \( Y \) in the subclass. The rest of the paper is devoted to proving rigorous results on the extent to which \( L_H \) and \( L_h \) distinguish spaces in simple subclasses of mm-spaces; namely, the subclasses of plane curves and metric trees.

5 Lower Bounds for Plane Curves

5.1 Point Clouds

We begin by reviewing some results on the simplest subclass of mm-spaces. Let \( \mathcal{M}^{m}_{N,k} \) denote the collection of isomorphism classes of mm-spaces \( X \) such that \( X \) is a point cloud of \( N \) points in \( \mathbb{R}^k \), \( d_X \) is Euclidean distance and \( \mu_X \) is uniform measure. The problem of reconstructing \( X \) from its collection of interpoint distances (i.e., from its distribution \( X' \)) is classical and has applications to DNA sequencing and X-ray crystallography [22].
Figure 1: A counterexample to Conjecture 5.1.

It is well known that $L_H$ does not distinguish elements of $\mathcal{M}_{N,k}^m$ for any $(N,k)$ with $N > 3$. A simple counterexample appears in [7] in the context of difference sets of integers and is given by mm-spaces $X, Y \in \mathcal{M}_{6,1}^m$ with $X = \{0, 1, 4, 10, 12, 17\}$ and $Y = \{0, 1, 8, 11, 13, 17\}$. One can check that $X \not\approx Y$, yet $L_H(X, Y) = 0$. Counterexamples in higher dimensions were given by Boutin and Kemper in [9, Section 1.1]. Boutin and Kemper refined the question of whether $L_H$ distinguishes point clouds and proved the following theorem (which we state using our terminology).

**Theorem 5.1** ([9]). The lower bound $L_H$ distinguishes elements of $\mathcal{M}_{N,k}^m$

(a) Locally: For every point cloud $X$, there exists $\epsilon_X > 0$ such that if a point cloud $Y$ satisfies $d_{L_H}(X, Y) < \epsilon_X$ and $L_H(X, Y) = 0$ then $X$ and $Y$ differ by a rigid motion.

(b) Densely: For every point cloud $X$ and every $\epsilon > 0$, there exists a point cloud $X'$ such that $d_{L_H}(X, X') < \epsilon$ and $X'$ is distinguished from all other point clouds by $L_H$.

We will provide a similar result regarding the distinguishing power of $L_h$ on the space of metric trees in Section 6.

5.2 The Curve Histogram Conjecture

In [11], Brinkman and Olver consider the class of mm-spaces $X$ with $X$ a plane curve, $d_X$ extrinsic Euclidean distance between points, and $\mu_X$ normalized arclength measure along the curve. They study global distance distributions (which they refer to as distance histograms) for elements of this class. They obtain convergence results for discrete approximations of plane curves and pose the following conjecture.

**Conjecture 5.1** (The Curve Histogram Conjecture [11]). The distance histogram $H_X$ determines a fully regular plane curve up to isometry.

*Fully regular* is a regularity class of plane curves including polygons and generic smooth curves—see [11] for details. Using the terminology of this paper, the Curve Histogram Conjecture says that $L_H$ distinguishes plane curves.

Consider the curves shown in Figure 1. Each curve is obtained as the boundary of a polygon, each constructed by appending four congruent isosceles triangles to a regular octagon. These curves are clearly non-isometric and will serve as our counterexample to the Curve Histogram Conjecture. It is interesting to note that these curves first appeared in [23] as a counterexample to a conjecture of Blaschke that linear-intercept distributions distinguish convex polygons (see [32]).

To prove that the curves give a counterexample to the curve histogram conjecture, we will derive a general criterion for mm-spaces to have the same global distance distribution. For a mm-space $X$, let $X = X_1 \cup X_2 \cup \cdots \cup X_n$ be a partition. For each $i, j \in \{1, \ldots, n\}$ let $H_X^{ij}$ be the function defined by

$$H_X^{ij}(r) = \mu_X \otimes \mu_X \left(\{(x, x') \in X_i \times X_j \mid d_X(x, x') \leq r\}\right).$$
Denoting by \( \mathbb{1}_r : X \times X \to \mathbb{R} \) the indicator function for the set
\[
\{(x, x') \in X \times X \mid d_X(x, x') \leq r\},
\]
we have
\[
H_X(r) = \int_{X \times X} \mathbb{1}_r(x, x') \mu_X \otimes \mu_X(dx \times dx')
= \sum_{i,j=1}^n \int_{X \times X_j} \mathbb{1}_r(x, x') \mu_X \otimes \mu_X(dx \times dx') = \sum_{i,j=1}^n H_X^{ij}(r).
\]

The next theorem then follows immediately.

**Theorem 5.2.** Let \( X \) and \( Y \) be mm-spaces. If there exist partitions \( X = X_1 \cup X_2 \cup \cdots \cup X_n \) and \( Y = Y_1 \cup Y_2 \cup \cdots \cup Y_n \) such that for each pair \( (X_i, X_j) \), there is a pair \( (Y_k, Y_ℓ) \) with \( H_X^{ij} = H_Y^{kℓ} \), then \( H_X = H_Y \).

**Proposition 5.1.** The curves in Figure 7 satisfy the hypotheses of Theorem 5.2 and therefore have the same distance distribution.

**Proof.** Denote the curve on the left by \( X \) and the curve on the right by \( Y \). We partition each curve into triangular pieces \( T_i \) (respectively, \( T'_i \)) and straight pieces \( S_j \) (respectively, \( S'_j \)) according to the figure. For each pair \( (X_i, X_j) \) of pieces \( X \) (here \( X_j \) stands as a placeholder for either an \( S_j \) or a \( T_j \)), we find a corresponding pair \( (Y_k, Y_ℓ) \) in \( Y \) so that there is a rigid motion taking the first pair onto the second. Since the functions \( H_X^{ij} \) and \( H_Y^{kℓ} \) are invariant under rigid isometries, the hypotheses of Theorem 5.2 are then satisfied.

Pairs of the form \( (S_i, S_j) \) and \( (T_i, T_ℓ) \) are matched with \((S'_i, S'_j)\) and \((T'_i, T'_j)\), respectively. Moreover, note that curve \( Y \) is obtained form curve \( X \) by “swapping” \( S_1 \) with \( T_1 \), it is clear that \( (S_1, T_1) \) should be matched with \((S'_1, T'_1)\). It also follows that we have obvious matchings \((S_i, S_j) \leftrightarrow (S'_i, S'_j), (T_i, T_ℓ) \leftrightarrow (T'_i, T'_j)\) and \((S_i, T_ℓ) \leftrightarrow (S'_i, T'_j)\) for \( i, j \neq 1 \). It remains to show that the desired matchings exist for pairs which include \( S_1 \) or \( T_1 \). These are given by
\[
\begin{align*}
(S_1, S_2) &\leftrightarrow (S'_1, S'_2) & (S_1, T_2) &\leftrightarrow (S'_1, T'_2) & (T_1, T_2) &\leftrightarrow (T'_1, T'_2) & (T_1, S_2) &\leftrightarrow (T'_1, S'_2) \\
(S_1, S_3) &\leftrightarrow (S'_1, S'_3) & (S_1, T_3) &\leftrightarrow (S'_1, T'_3) & (T_1, T_3) &\leftrightarrow (T'_1, T'_3) & (T_1, S_3) &\leftrightarrow (T'_1, S'_3) \\
(S_1, S_4) &\leftrightarrow (S'_1, S'_4) & (S_1, T_4) &\leftrightarrow (S'_1, T'_4) & (T_1, T_4) &\leftrightarrow (T'_1, T'_4) & (T_1, S_4) &\leftrightarrow (T'_1, S'_4) \\
\end{align*}
\]

**Remark 5.1.** By smoothing corners for the curves in our example in a symmetric manner, we could produce a pair of \( C^∞ \) curves satisfying the hypotheses of Theorem 5.2. This would therefore produce a counterexample to the Curve Histogram Conjecture involving \( C^∞ \) curves.

### 5.3 More Counterexamples for Curves

The construction used in the previous section can be used to generate counterexamples to the curve histogram conjecture starting with any regular polygon with an even number of sides. Starting from such a polygon, partition its edges into two sets of equal size such that there is no isometry taking one of the sets onto the other. We then construct a pair of curves by appending congruent isosceles triangles to each edge partition set—see Figure 2. One is able to show that the resulting curves satisfy the hypotheses of Theorem 5.2. We note that a similar idea was used recently in [16] to construct more counterexamples to Blaschke’s conjecture.

By starting with a regular \( 2n \)-gon with \( n \) sufficiently large and by appending isosceles triangles with sufficiently small height, we can construct pairs of curves in this manner which are arbitrarily close to a circle in Gromov-Hausdorff distance. We therefore get the following corollary demonstrating a lack of distinguishing power of \( L_H \) locally (cf. the main result of [4] demonstrating injectivity in a neighborhood of the unit circle for the *circular integral invariant* for plane curves).
Corollary 5.1. The lower bound $L_H$ does not distinguish plane curves in any Gromov-Hausdorff neighborhood of the unit circle.

The global histogram does distinguish the unit circle, denoted $S^1$, from all other simple closed curves.

Proposition 5.2. Assume a smooth simple planar closed curve $C$ (endowed with euclidean distance and normalized length measure) satisfies $H_C(r) = H_{S^1}(r)$ for all $r \geq 0$. Then, $C$ is isometric to $S^1$.

Proof. By [11], one has that for $r > 0$ small enough the Taylor expansion,

$$H_C(r) = \frac{2r}{\ell(C)} + \frac{r^3}{12\ell(C)^2} \int_C \kappa^2(s) \, ds + O(r^5)$$

where $\kappa$ denotes curvature (and $ds$ denotes arc length), and $\ell(C)$ is the length of $C$. Then, for $C = S^1$ this yields $H_{S^1}(r) = \frac{r}{\pi} + \frac{r^3}{24\pi} + O(r^5)$. Equating these two Taylor expansions gives $\ell(C) = 2\pi$ and $\int_C \kappa^2(s) \, ds = 2\pi$. By assumption, $\int_C \kappa(s) \, ds = 2\pi$, so applying the Cauchy-Schwarz inequality, one has

$$4\pi^2 = \left( \int_C \kappa(s) \cdot 1 \, ds \right)^2 \leq \left( \int_C \kappa^2(s) \, ds \right) \left( \int_C 1 \, ds \right) = 2\pi \cdot \ell(C) = 2\pi \cdot 2\pi = 4\pi^2.$$

Since equality in (*) is then forced, the equality conditions in the Cauchy-Schwarz inequality imply that $\kappa$ is constant. Since $\ell(C) = 2\pi$, then $\kappa(s) = 1$ for all $s \in [0, 2\pi]$. This finishes the proof.

5.4 Counterexamples for Surfaces

A similar construction can be used to produce nonisometric polyhedral surfaces which have the same distance distributions. Starting with a dodecahedron or icosahedron, we partition the faces into two sets so that there is no isometry mapping one set onto the other. For each set, we construct a new polyhedral surface by attaching a symmetric pyramid along the faces in the set. The resulting pair of polyhedral surfaces are nonisometric by construction, but one can show that they satisfy the hypotheses of Theorem 5.2. An example is shown in Figure 3. Similar constructions are also explored in [16] as counterexamples to a higher dimensional analogue of Blaschke’s conjecture.

6 Lower Bounds for Metric Trees

A metric tree is a compact metric space which is homeomorphic to a contractible one-dimensional simplicial complex. We treat a metric tree as a mm-space by endowing it with the uniform probability measure and denote it by $T = (T, d_T, \mu_T)$. The space of isomorphism classes of measured metric trees is denoted $\text{MTrees}$. Metric trees arise naturally in several applications; for instance, they appear in computational anatomy as models for blood vessels [14, 15] and in data visualization and shape analysis as merge trees [19, 28]. The goal of this section is to prove the following theorem.
Figure 3: Nonisometric polyhedral surfaces with the same distributions of distance. The surfaces are constructed by adding pyramids according to the partition described by the net on the left.

Theorem 6.1. The lower bound $L_h$ distinguishes metric trees

(a) Locally: For every metric tree $T$, there exists $\epsilon_T > 0$ such that if a metric tree $S$ satisfies $d_{GH}(T, S) < \epsilon_T$ and $L_h(T, S) = 0$, then $S$ is isomorphic to $T$.

(b) Densely and Generically: For every metric tree $T$ and every $\epsilon > 0$, there exists a metric tree $T'$ such that $d_{GH}(T, T') < \epsilon$ and $T'$ is distinguished from all other metric trees by $L_h$. Moreover, a metric tree is distinguished from all other metric trees by $L_h$ almost surely with respect to a natural measure on $\text{MTrees}$.

Part (a) suggests that $L_h$ can itself be upgraded to an intrinsic metric on $\text{MTrees}$ after symmetrizing as in Proposition 2.2 (cf. [13]). Part (b) implies that the symmetrization of $L_h$ defines a metric on a full measure subset of $\text{MTrees}$.

Theorem 6.1 result is analogous to Boutin and Kemper’s theorem on distance distributions of point clouds, described above in Theorem 5.1. It is also similar to the main result of [30], where the authors considered injectivity properties of a more complicated invariant of metric graphs based on persistent homology. Parts (a) and (b) of our theorem will follow from the stronger results Proposition 6.2 and Proposition 6.3, respectively, together with Proposition 6.1.

6.1 A Condition for Nonvanishing $L_h$

For a metric tree $T$, let $\mathcal{N}(T)$ denote its collection of vertices of valence $\neq 2$, called nodes (this concept is well-defined, as it doesn’t depend on the choice of homeomorphism onto a 1-complex). Also consider the finite multiset of functions

$$h_{\mathcal{N}}(T) = \{h_T(x, \cdot) \mid x \in \mathcal{N}(T)\},$$

which we refer to as the node multiset. The node multiset is an isomorphism invariant of a metric tree. These invariants give a convenient condition guaranteeing that $L_h$ does not vanish.

Proposition 6.1. Let $T$ and $S$ be metric trees. If $h_{\mathcal{N}}(T) \neq h_{\mathcal{N}}(S)$ as multisets of functions, then $L_h(T, S) \neq 0$.

The proof uses a technical lemma, whose proof is given in Section 7.1. For $k = 1$ or $k \geq 3$, let $\mathcal{N}_k(T)$ denote the set of valence-$k$ nodes of a metric tree $T$ (once again, this concept is well-defined).

Lemma 6.1. Let $T$ and $S$ be metric trees. If $L_h(T, S) = 0$ then $\#\mathcal{N}_k(T) = \#\mathcal{N}_k(S)$ for all $k$. It follows that $\#\mathcal{N}(T) = \#\mathcal{N}(S)$.

Proof of Proposition 6.1. We first consider the case that there is some $v \in \mathcal{N}(T)$ such that $h_T(v, \cdot) \neq h_S(w, \cdot)$ for all $w \in \mathcal{N}(S)$. If this is the case, then $h_T(v, \cdot) \neq h_S(y, \cdot)$ for any point $y$ in $S$, as $h_T(v, r)$ will differ from any $h_S(y, r)$ corresponding to a non-node $y$ for small values of $r$. It follows that the
continuous function $S \to \mathbb{R}_{\geq 0}$ on the compact space $S$ defined by $y \mapsto c_{T,S}(v,y)$ achieves its minimum $m_v > 0$. We claim that there exists an open neighborhood $U$ of $v$ in $T$ such that for any $x \in U$,\[
inf_{y \in S} c_{T,S}(x,y) \geq \frac{m_v}{2}.
\]
If this is not the case then there is a point $x \in T$ such that $c_{T,T}(v,x) < m_v/2$ and $c_{T,S}(x,y) < m_v/2$ for some $y \in S$. It is not hard to see that the cost function satisfies a triangle inequality-like relation for all $v, x \in T$ and $y \in S$:\[
c_{T,S}(v,y) \leq c_{T,T}(v,x) + c_{T,S}(x,y).
\]
Applying this to our particular points $v, x$ and $y$, it follows that\[
c_{T,S}(v,y) \leq c_{T,T}(v,x) + c_{T,S}(x,y) < m_v,
\]
which is a contradiction. For any measure preserving map $\phi : T \to S$,\[
\int_T c_{T,S}(x,\phi(x)) \mu_T(dx) \geq \frac{m_v}{2} \cdot \mu_T(U),
\]
whence $L_h(T,S) > 0$.

Now suppose that for each $v \in \mathcal{N}(T)$ there exists some $w \in \mathcal{N}(S)$ with $h_T(v,\cdot) = h_S(w,\cdot)$. Since we have assumed that $h_N(T) \neq h_N(S)$, Lemma 6.1 implies that there is some $w \in \mathcal{N}(S)$ such that $h_T(v,\cdot) \neq h_S(w,\cdot)$ for all $v \in \mathcal{N}(T)$. Let $m_w > 0$ denote the minimum of the map $x \to c_{T,S}(x,w)$. By a similar argument to the one given above, there exists an open neighborhood $V$ of $w$ such that for all $y \in V$, $\inf_x c_{T,S}(x,y) > m_w/2$. For any measure-preserving map $\phi : T \to S$, we have\[
\int_T c_{T,S}(x,\phi(x)) \mu_X(dx) \geq \frac{m_w}{2} \cdot \mu_X(\phi^{-1}(V)) = \frac{m_w}{2} \cdot \mu_Y(V),
\]
and it follows that $L_h(T,S) > 0$. \hfill $\square$

### 6.2 Counterexamples

Before proving Theorem 6.1 we provide counterexamples for some stronger statements about the distinguishing powers of $L_H$ and $L_h$ for combinatorial and metric trees.

**Example 6.1 (LH Does Not Distinguish Trees).** For combinatorial trees, the global distance distribution is equal to the path sequence, a classical graph invariant. It is easy to construct simple examples of combinatorial trees which are non-isomorphic, but which share the same distribution. Figure 7 shows an example, taken from [24], of non-isomorphic combinatorial trees $T_1$ and $T_2$ with $L_H(T_1,T_2) = 0$. Taking edge lengths to be 1, we produce $T_1, T_2 \in \text{MTrees}$ which are not distinguished by $L_H$.

One can check that the node multisets for the metric trees are not equal, $h_N(T_1) \neq h_N(T_2)$. It follows from Proposition 6.1 that these metric trees are distinguished by $L_h$. 

![Figure 4: A pair of non-isomorphic combinatorial trees with the same global distance distribution—see [24].](image)
Example 6.2 (\(L_h\) Does Not Distinguish Trees). Consider the tree structure shown in Figure 5 where we are free to assign the number of branches \(A_j, B_j\) and \(C_j\) in each lobe. Let \(T_1\) and \(T_2\) be the combinatorial trees with numbers of branches given, respectively, by

\[
\begin{pmatrix}
A_1 & A_2 & A_3 \\
B_1 & B_2 & B_3 \\
C_1 & C_2 & C_3
\end{pmatrix} = \begin{pmatrix}
5 & 10 & 5 \\
3 & 3 & 14 \\
1 & 7 & 12
\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}
A_1 & A_2 & A_3 \\
B_1 & B_2 & B_3 \\
C_1 & C_2 & C_3
\end{pmatrix} = \begin{pmatrix}
5 & 14 & 1 \\
3 & 7 & 10 \\
5 & 3 & 12
\end{pmatrix}.
\]

The trees \(T_1\) and \(T_2\) are nonisomorphic. However, one can check explicitly that \(L_h(T_1, T_2) = 0\).

Moreover, the metric trees \(T_1\) and \(T_2\) obtained by taking all edgelengths to be one are nonisomorphic, but satisfy \(L_h(T_1, T_2) = 0\). A measure-preserving map realizing this value is obtained by interpolating the map on nodes which extends the node-pairing map of the combinatorial tree.

Example 6.3 (\(L_h\) Does Not Distinguish Trees in Any Gromov-Hausdorff Neighborhood). Fix any metric tree \(T\). Let \(S'_j\) denote the metric tree obtained from the combinatorial tree \(T_j\) defined in Example 6.2. Let \(S'_{j,\delta}\) denote the same metric tree with all edges scaled to have length \(\delta\). Fix a leaf \(v \in T\). We form the metric tree \(S_{j,\delta}\) by attaching \(S'_{j,\delta}\) (say, along its central node) to \(T\) (say, at \(v\)). Then \(L_h(S_{1,\delta}, S_{2,\delta}) = 0\); indeed, this is realized by defining a map \(\phi : S_{1,\delta} \to S_{2,\delta}\) taking nodes of \(T \subset S_{1,\delta}\) to themselves, nodes of \(S_{1,\delta}\) to nodes of \(S_{2,\delta}\) according to the matching from Example 6.2, and extending across edges by interpolation. Moreover, \(S_{j,\delta}\) can be made arbitrarily close in Gromov-Hausdorff distance to \(T\) by taking \(\delta\) to be sufficiently small.

6.3 Merge Trees and the Interleaving Distance

To study the properties of MTrees as a metric space under the Gromov-Hausdorff metric, we use the formalism of merge trees and their interleaving distances [28]. For a connected topological space \(X\) endowed with a continuous function \(f : X \to \mathbb{R}\), denote the epigraph of \(f\) by

\[\text{epi}(f) = \{(x, r) \in X \times \mathbb{R} \mid r \geq f(x)\}.$

The height function \(\tilde{f} : \text{epi}(f) \to \mathbb{R}\) is given by \(\tilde{f}(x, r) = r\). One forms a topological space \(M_f = \text{epi}(f)/\sim\), where \((x, r) \sim (x', r')\) if and only if \(r = r'\) and \((x, r), (x', r') \in \tilde{f}^{-1}(r)\) lie in the same connected component of \(\text{epi}(f)\). Then \(M_f\) has the structure of a tree and \(\tilde{f}\) induces a well-defined function \(M_f \to \mathbb{R}\) by \([x, r] \to r\). We abuse notation and continue to denote this induced map by \(\tilde{f}\).

Given a pair of merge trees \(M_f\) and \(M_g\), one computes the interleaving distance \(d_i(M_f, M_g)\) as follows. For any \(\varepsilon > 0\), there is a shift map \(\sigma^{\varepsilon} : M_f \to M_f\) which takes a point \([x, r]\) at height \(r\) to its unique ancestor \([x', r + \varepsilon]\) at height \(r + \varepsilon\). An \(\varepsilon\)-morphism is a continuous map \(\phi : M_f \to M_g\) such
that $\overline{\gamma} \circ \phi = T + c$. An $\epsilon$-interleaving of $M_f$ and $M_g$ is a pair of $\epsilon$-morphisms $\phi : M_f \to M_g$ and $\psi : M_g \to M_f$ such that
\[ \psi \circ \phi = \sigma^2_{\epsilon} \quad \text{and} \quad \phi \circ \psi = \sigma^2_{g}. \]

When there exists such an $\epsilon$-interleaving we say that $M_f$ and $M_g$ are $\epsilon$-interleaved. We then define
\[ d_{\epsilon}(M_f, M_g) = \inf \{ \epsilon \mid M_f \text{ and } M_g \text{ are } \epsilon\text{-interleaved} \}. \]

Let $T$ be a metric tree. For each $x \in T$, let $f^T_x : T \to \mathbb{R}$ be defined by $f^T_x(y) = -d_T(x, y)$. We can then form the merge tree $M_{fT}$, which we denote more succinctly as $T^x$. In this case, the structure of $T$ is largely unchanged as we pass to the merge tree $T^x$; indeed, if $x$ is not a leaf of $T$, then $T^x$ is a rooting of $T$ at $x$, with an extra edge extending to $\infty$ from $x$. If $x$ is a leaf of $T$, then $T^x$ is a rooting at the unique neighbor of $x$, once again with an infinite edge extending from the root.

For metric trees $T$ and $S$, define
\[ \Delta(T, S) = \min_{\tau \in \mathcal{N}(T), s \in \mathcal{N}(S)} d_t(T^\tau, S^s). \]

Recent results of Agarwal et. al. [1] relate this quantity to the Gromov-Hausdorff distance between $T$ and $S$ and give a short list of candidate values for $\Delta(T, S)$. These are summarized below.

**Theorem 6.2 (1).** For metric trees $T$ and $S$,
1. $\frac{1}{2} d_{GH}(T, S) \leq \Delta(T, S) \leq 14 d_{GH}(T, S)$;
2. for each $t \in \mathcal{N}(T)$ and $s \in \mathcal{N}(S)$, $d_t(T^t, S^s) \in \Lambda_{1,1} \cup \Lambda_{1,2} \cup \Lambda_{2,2}$, where
\[
\begin{align*}
\Lambda_{1,1} &= \left\{ \frac{1}{2} | \overline{f}_t^T(u) - \overline{f}_t^T(v) | \mid u, v \in \mathcal{N}(T^t) \right\}, \\
\Lambda_{1,2} &= \left\{ \frac{1}{2} | \overline{f}_s^S(u) - \overline{f}_s^S(v) | \mid u, v \in \mathcal{N}(S^s) \right\}, \\
\Lambda_{2,2} &= \left\{ | \overline{f}_t^T(u) - \overline{f}_s^S(v) | \mid u \in \mathcal{N}(T^t), v \in \mathcal{N}(S^s) \right\}.
\end{align*}
\]

### 6.4 $L_h$ Distinguishes Metric Trees Locally

Part (a) of Theorem 6.1 follows by combining Proposition 6.1 with the following result.

**Proposition 6.2.** For any metric tree $T$, there exists $\epsilon_T > 0$ such that if $h_{\mathcal{N}}(T) = h_{\mathcal{N}}(S)$ and $d_{GH}(T, S) < \epsilon_T$, then $T$ and $S$ are isomorphic.

We will introduce some notation and a technical lemma in order to prove the theorem. Let $T$ be a metric tree and fix some point $x \in T$. Let $\partial B_{d_T}(x, r)$ denote the discrete set of boundary points of the metric ball $B_{d_T}(x, r)$. For fixed $x$, consider the function $h'_T(x, r)$ defined by
\[ h'_T(x, r) = \lim_{\epsilon \to 0^+} \frac{h_T(x, r + \epsilon) - h_T(x, r)}{\epsilon}. \]

The function is piecewise constant with a finite number of discontinuities at some radii $r^1_T < r^2_T < \cdots < r^r_{M_T}$ depending on $x$. Moreover, it is the case that
\[ h'_T(x, r) = \# \partial B_{d_T}(x, r + \epsilon) \]
for some sufficiently small $\epsilon > 0$. For each of the $r^j_T$, $\partial B_{d_T}(x, r)$ necessarily contains a node. We claim that for $r \notin \{ r^1_T, \ldots, r^r_{M_T} \}$, $\partial B_{d_T}(x, r)$ can only contain a node if it, in particular, contains a leaf. Indeed, if $r$ lies in an interval $(r^j_T, r^{j+1}_T)$, $j = 0, 1, \ldots, M_T$ (taking $r^0_T = 0$ and $r^{M_T+1}_T = \infty$, for notational convenience), and $\partial B_{d_T}(x, r)$ contains only non-leaf nodes, then, for sufficiently small $\epsilon > 0$, the number of non-leaf elements of $\partial B_{d_T}(x, r + \epsilon)$ is strictly greater than the number of non-leaf elements of $\partial B_{d_T}(x, r)$. This contradicts the fact that the function $h'_T(x, r)$ is constant on such intervals.

Now suppose that $x$ is a node of $T$. The discussion of the previous paragraph allows us to distinguish a finite list of candidate values $r$ where $\partial B_{d_T}(x, r)$ can contain a node.
Lemma 6.2. Let \( T \) be a metric tree and let \( r_1, \ldots, r_M \) be as defined above. Let \( \ell_1, \ldots, \ell_N \) denote the lengths of all leaf edges of \( T \). The sphere \( \partial B_{d_T}(x, r) \) can only contain a node if \( r \) lies in the set

\[
\Sigma_T(x) = \{ \lambda_1 r_1 + \cdots + \lambda_M r_M^x + \mu_1 \ell_1 + \cdots + \mu_N \ell_N | \lambda_j, \mu_k \in \{0,1\}, \text{ at most one } \lambda_j \neq 0 \}.
\]

Proof. Suppose that \( \partial B_{d_T}(x, r) \) contains a node and that \( r \not\in \{ r_1^x, \ldots, r_M^x \} \). By the above discussion, \( \partial B_{d_T}(x, r) \) must contain a leaf \( v_1 \). Let \( \gamma_1 \) denote the unique path joining \( x \) to \( v_1 \). Let \( x_1 \) denote the node lying on \( \gamma_1 \) which immediately precedes \( v_1 \), let \( s_1 = d_T(x, x_1) \) and let \( \ell_{j_1} = d_T(x_1, v_1) \). Since \( v_1 \) is a leaf, \( \ell_{j_1} \) is the length of its leaf edge. There are several cases to consider. If \( x_1 = x \), then we are clearly finished, as \( r = \ell_{j_1} \). If \( s_1 = r_1^x \in \{ r_1^x, \ldots, r_M^x \} \), then we are finished because this implies \( r = r_1^x + \ell_{j_1} \in \Sigma_T(x) \). We therefore assume that we are in neither of these situations and iterate the process. That is, we have a node \( x_1 \in \partial B_{d_T}(x, s_1) \), where \( s_1 \not\in \{ r_1^x, \ldots, r_M^x \} \), and it follows that \( \partial B_{d_T}(x, s_1) \) contains a leaf \( v_2 \). Let \( \gamma_2 \) denote the path from \( x \) to \( v_2 \), let \( x_2 \) denote the node on \( \gamma_2 \) preceding \( v_2 \), let \( s_2 = d_T(x, x_2) \) and let \( \ell_{j_2} = d_T(x_2, v_2) \). The algorithm terminates at this stage if \( x_2 = x \) (in which case \( r = \ell_{j_2} + \ell_{j_1} \)) or if \( s_2 = r_1^x \in \{ r_1^x, \ldots, r_M^x \} \) (in which case \( r = r_1^x + \ell_{j_2} + \ell_{j_1} \)) and otherwise iterates again. The algorithm must eventually terminate, since the distances \( s_j \) in each step are strictly decreasing. \( \square \)

Proof of Proposition 6.2. From the multiset \( h_{\mathcal{N}}(T) \), we are able to extract all possible values of \( r \) for which the function \( h_T^r(x, \cdot) \) is discontinuous, for each \( x \in \mathcal{N}(T) \). That is, using the notation of Lemma 6.2, we can determine the set

\[
\{ r_1, \ldots, r_M \} = \bigcup_{x \in \mathcal{N}(T)} \{ r_1^x, \ldots, r_M^x \}
\]

Moreover, we can extract the set of lengths \( \{ \ell_1, \ldots, \ell_N \} \) of all leaf edges of \( T \) from the multiset \( h_{\mathcal{N}}(T) \). Indeed, first note that a function in \( h_{\mathcal{N}}(T) \) can be distinguished as the function of a leaf by the observation that \( h_T^r(x, 0) = 1 \) if and only if \( x \) is a leaf. Next use the fact that the length of the leaf edge of a leaf \( x \) is the smallest value of \( r > 0 \) such that \( h_T^r(x, r) \neq 1 \). The multiset \( h_{\mathcal{N}}(T) \) therefore determines the set

\[
\Sigma_T = \{ \lambda_1 r_1 + \cdots + \lambda_M r_M + \mu_1 \ell_1 + \cdots + \mu_N \ell_N | \lambda_j, \mu_k \in \{0,1\} \}.
\]

In particular, \( \Sigma_T \) contains the union of all \( \Sigma_T(x) \) for \( x \in \mathcal{N}(T) \). In general, this containment is strict.

Fix a node \( t \in \mathcal{N}(T) \) and consider the set

\[
\Lambda_{1,1} = \left\{ \frac{1}{2} | f^T_t(u) - f^T_t(v) | \mid u, v \in \mathcal{N}(T^t) \right\}
\]

defined in Theorem 6.2. Lemma 6.2 implies that for any node \( u \in \mathcal{N}(T^t) \), the value \(-f^T_t(u)\) lies in the set \( \Sigma_T(t) \). It follows that

\[
\Lambda_{1,1} \subset \left\{ \frac{1}{2} | A - A' | \mid A, A' \in \Sigma_T(t) \right\}
\]

\[
\subset \left\{ \frac{1}{2} | A - A' | \mid A, A' \in \Sigma_T \right\} =: \Sigma_1
\]

If we assume that \( h_{\mathcal{N}}(T) = h_{\mathcal{N}}(S) \), then the discussion from the previous paragraph implies that \( \Sigma_T = \Sigma_S \), and we have similar statements for the other sets \( \Lambda_{i,j} \) from Theorem 6.2.

\[
\Lambda_{2,2} \subset \left\{ \frac{1}{2} | B - B' | \mid B, B' \in \Sigma_S(t) \right\} \subset \left\{ \frac{1}{2} | B - B' | \mid B, B' \in \Sigma_T \right\} = \Sigma_1,
\]

\[
\Lambda_{1,2} \subset \{ | A - B | \mid A \in \Sigma_T(t), B \in \Sigma_S(s) \} \subset \{ | A - B | \mid A, B \in \Sigma_T \} =: \Sigma_2
\]

Combining these inclusions with part 2 of Theorem 6.2, we have

\[
\Delta(T, S) = \min_{t \in \mathcal{N}(T), s \in \mathcal{N}(S)} d_t(T^t, S^s) \in \Sigma_1 \cup \Sigma_2.
\]
Finally, let
\[ \epsilon_T = \frac{1}{14} \min (\Sigma_1 \cup \Sigma_2) \setminus \{0\}. \]
If \( h_N(T) = h_N(S) \) and \( d_{GH}(T, S) < \epsilon_T \), then part 1 of Theorem 6.2 implies that
\[ \Delta(T, S) \leq 14d_{GH}(T, S) < \min (\Sigma_1 \cup \Sigma_2) \setminus \{0\}. \]

By the discussion of the previous paragraph, \( \Delta(T, S) \in \Sigma_1 \cup \Sigma_2 \), and it follows that \( \Delta(T, S) = 0 \).

Applying the lower bound of Theorem 6.2, we conclude that \( d_{GH}(T, S) = 0 \) and it follows that \( T \) and \( S \) are isomorphic.

\[ \Box \]

6.5 \( L_h \) Distinguishes Metric Trees Generically

The lower bound \( L_h \) distinguishes trees generically. To state this more precisely, we define a probability measure \( \rho \) on \( \text{MTrees} \), following the definition of [30]. For a fixed combinatorial tree \( T = (V, E) \in \text{CTrees} \), pick an arbitrary probability measure \( \rho_T \) on the orthonormal \( \mathbb{R}^{V,E} \) which has a density with respect to Lebesgue measure. This induces a probability measure on the subspace of \( \text{MTrees} \) containing metric trees whose underlying combinatorial tree is equal to \( T \). Next choose an arbitrary probability mass function \( f \) on the countable space of combinatorial trees. The probability measure \( \rho \) on \( \text{MTrees} \) is a mixture model: a random element of \( \text{MTrees} \) is drawn by first drawing a combinatorial tree \( T \) according to \( f(T) \), then promoting it to a metric tree according to \( \rho_T \).

Together with Proposition 6.1, the following proposition proves part (b) of Theorem 6.1.

Proposition 6.3. The set of metric trees \( T \) which are determined up to isomorphism by the multiset \( h_N(T) \) is dense in \( \text{MTrees} \) with respect to Gromov-Hausdorff distance and full measure with respect to \( \rho \).

Our proof relies on a technical lemma, which is proved in Section 7.2.

Lemma 6.3. If all functions in the multiset \( h_N(T) \) are distinct, then \( T \) is determined by \( h_N(T) \) up to isomorphism.

Proof of Proposition 6.3. Lemma 6.3 implies that the map \( T \mapsto h(T) \) is injective on the set of metric trees \( T \) such that the functions in \( h_N(T) \) are distinct. This set contains, in particular, the set of metric trees with the property that there are no nontrivial equalities between edgelengths involving only addition and subtraction. This property is referred to as Property P in [30], where the authors remark that the set of metric graphs lacking this property has zero measure with respect to \( \rho \). This proves the statement about generic injectivity.

To prove the statement about density, let \( T \) be a metric tree and let \( \epsilon > 0 \). We wish to construct a new tree \( T' \) which is \( \epsilon \)-close to \( T \) in Gromov-Hausdorff distance and which has the property that \( h_N(T') \) contains distinct functions. To do so, enumerate the elements of \( \mathcal{N}(T) \) as \( v_1, \ldots, v_N \). Let \( r_1, \ldots, r_N \) be a list of distinct real numbers which are also distinct from all edgelengths of \( T \). At each \( v_j \) we append a leaf edge of length \( r_j \), and the resulting metric tree is \( T' \). By rescaling the \( r_j \) uniformly as needed, we can ensure that \( d_{GH}(T, T') < \epsilon \). It therefore remains to show that the functions in \( h_N(T') \) are all distinct.

Let \( v \in \mathcal{N}(T') \) and let \( \delta > 0 \) be less than the shortest edgelength of \( T' \). If \( v \) is a leaf which was appended to the vertex \( v_j \in T \) in order to obtain \( T' \), then for \( r \leq r_j + \delta \),
\[
h'_{T'}(v, r) = \begin{cases} 1 & r \leq r_j, \\ 2 & r_j < r \leq r_j + \delta. \end{cases}
\]

By our choice of the \( r_j \), this is enough to characterize the function \( h_{T'}(v, \cdot) \) uniquely amongst those in \( h_N(T') \). On the other hand, suppose that \( v = v_j \) (that is, \( v \) corresponds to a node in the original tree \( T \)). Let \( d \geq 2 \) denote the degree of \( v_j \), as a node in \( T' \). Then for \( r \leq r_j + \delta \),
\[
h'_{T'}(v, r) = \begin{cases} d & r \leq r_j, \\ d-1 & r_j < r \leq r_j + \delta. \end{cases}
\]
and this is once again enough to characterize \( h_T(v, \cdot) \) uniquely.

\[\square\]

7 Appendix: Proofs of Technical Lemmas

7.1 Lemma 6.1

Our proof of Lemma 6.1 requires some auxiliary results.

**Lemma 7.1.** Let \( \mathcal{X} \) and \( \mathcal{Y} \) be mm-spaces and let \( f : X \rightarrow \mathbb{R} \) and \( g : Y \rightarrow \mathbb{R} \) be measurable maps. For each \( \mu \in \mathcal{M}(\mu_X, \mu_Y) \), we have

\[
\int_{X \times Y} |f(x) - g(y)| \mu(dx \times dy) \geq d_{W,1}^R (f \# \mu_X, g \# \mu_Y).
\]

**Proof.** Recall the definition of \( d_{W,1}^R \) from (4). First note that the Wasserstein distance is well-defined, since \( \mathbb{R} \) with the Euclidean distance is a Polish space and the measures \( f \# \mu_X \) and \( g \# \mu_Y \) are compactly supported. Let \( f \times g : X \times Y \rightarrow \mathbb{R} \times \mathbb{R} \) be the product map \( (f \times g)(x, y) = (f(x), g(y)) \) and consider the measure \( (f \times g) \# \mu \in \mathcal{P}(\mathbb{R}^2) \). It is easy to check that \( (f \times g) \# \mu \) is a measure coupling of \( f \# \mu_X \) and \( g \# \mu_Y \). The change of variables formula then implies that

\[
\int_{X \times Y} |f(x) - g(y)| \mu(dx \times dy) = \int_{\mathbb{R} \times \mathbb{R}} |u - v| \left( (f \times g) \# \mu \right)(du \times dv)
\geq d_{W,1}^R (f \# \mu_X, g \# \mu_Y).
\]

\[\square\]

For a mm-space \( \mathcal{X} \) and a fixed \( r > 0 \), let \( h^r_X : X \rightarrow \mathbb{R} \) denote the function \( h^r_X(x) = h_X(x, r) \). We will consider the one-parameter family of measures \( \{(h^r_X) \# \mu_X \}_{r \geq 0} \subset \mathcal{P}(\mathbb{R}) \). We note that this family of measures appears in the definition of the modulus of mass distribution studied in [18].

**Lemma 7.2.** For mm-spaces \( \mathcal{X} \) and \( \mathcal{Y} \), \( L_h(\mathcal{X}, \mathcal{Y}) = 0 \) implies that \( (h^r_X) \# \mu_X = (h^r_Y) \# \mu_Y \) for almost every \( r > 0 \).

**Proof.** If \( L_h(\mathcal{X}, \mathcal{Y}) = 0 \) then for every \( \epsilon > 0 \) there exists a measure-preserving map \( \phi : X \rightarrow Y \) which produces a coupling \( \mu_\phi \in \mathcal{M}(\mu_X, \mu_Y) \) such that

\[
\epsilon > \int_{X \times Y} c_{\mathcal{X}, \mathcal{Y}}(x, y) \mu_\phi(dx \times dy)
= \int_{X \times Y} \left( \int_0^\infty |h_X(x, r) - h_Y(y, r)| \ dr \right) \mu_\phi(dx \times dy)
= \int_0^\infty \left( \int_{X \times Y} |h^r_X(x) - h^r_Y(y)| \mu_\phi(dx \times dy) \right) \ dr
\geq \int_0^\infty d_{W,1}^R ((h^r_X) \# \mu_X, (h^r_Y) \# \mu_Y) \ dr,
\]

where (17) follows by Fubini’s Theorem and (18) follows from Lemma 7.1. Since this holds for arbitrarily small \( \epsilon \), it follows that \( d_{W,1}^R ((h^r_X) \# \mu_X, (h^r_Y) \# \mu_Y) = 0 \) for almost every \( r > 0 \). This completes the proof, since Wasserstein distance is a metric on the space of compactly supported probability measures on \( \mathbb{R} \).

\[\square\]

We now proceed with the proof of the main technical lemma.
Proof of Lemma 6.1 Suppose that \( L_h(T, S) = 0 \). Lemma 7.2 implies that we can choose an \( r \) which is less than half the length of the shortest edge in \( T \) or \( S \) such that \((h^r_T)\#\mu_T = (h^r_S)\#\mu_S\).

We first show that the number of leaves of \( T \) can be recovered from \((h^r_T)\#\mu_T\):

\[
(h^r_T)\#\mu_T(0, 2r) = \mu_T \left( \{ x \in T \mid \mu_T(B_T(x, r)) < 2r \} \right)
= \mu_T \left( \{ x \in T \mid d_T(x, N_1(T)) < r \} \right)
= \#N_1(T) \cdot r.
\]

It follows that \( T \) and \( S \) contain the same number of leaves.

A similar strategy works for nodes of higher valence. Let \( k_T \) denote the maximum node valence of \( T \). We claim that for each \( k \geq 3 \), the quantity \((h^r_T)\#\mu_T((k - 1) \cdot r, k \cdot r)\) is given by

\[
\frac{r}{k - 2} \cdot \#N_k(T) + \frac{r}{k + 1 - 2} \cdot \#N_{k+1}(T) + \cdots + \frac{r}{k_T - 2} \cdot \#N_{k_T}(T),
\]

when \( k \leq k_T \) and that it is equal to zero otherwise. Assuming that the claim holds, the number of nodes of each valence of \( T \) can be recovered recursively from \((h^r_T)\#\mu_T\), and this completes the proof.

It remains to derive (19), by definition,

\[
(h^r_T)\#\mu_T((k - 1) \cdot r, k \cdot r) = \mu_T \left( \{ x \in T \mid (k - 1) \cdot r < \mu_T(B_T(x, r)) < k \cdot r \} \right).
\]

By our choice of \( r \), the maximum value of \( \mu_T(B_T(x, r)) \) is \( r \cdot k_T \), when \( x \) is a node of valence \( k_T \). It follows that \((h^r_T)\#\mu_T((k - 1) \cdot r, k \cdot r) = 0 \) when \( k \geq k_T + 1 \). We then consider \( k \) with \( 3 \leq k \leq k_T \). A point \( x \in T \) satisfies \((k - 1) \cdot r < \mu_T(B_T(x, r)) < k \cdot r\) if and only if it satisfies one of the mutually exclusive conditions, indexed by \( \ell = k, k + 1, \ldots, k_T \):

\[
\frac{\ell - k}{\ell - 2} r < d_T(x, N_\ell(T)) < \frac{\ell - k - 1}{\ell - 2} r.
\]

To see this, note that if \( \mu_T(B_T(x, r)) > 2r \), then it must lie within distance \( r \) from a (unique) node of valence at least 3. Suppose that \( x \) lies within distance \( \epsilon < r \) to a node of valence \( \ell \). Then

\[
\mu_T(B_T(x, r)) = r + \epsilon + (\ell - 1) \cdot (r - \epsilon) = \ell \cdot r - (\ell - 2) \epsilon.
\]

Our condition then becomes

\[
(k - 1) \cdot r < \ell \cdot r - (\ell - 2) \cdot \epsilon < k \cdot r,
\]

and solving for \( \epsilon \) shows that condition (C\( \ell \)) must hold. The set of points satisfying this condition has measure

\[
\#N_\ell(T) \cdot \left( \frac{\ell - k - 1}{\ell - 2} r - \frac{\ell - k}{\ell - 2} r \right) = \#N_\ell(T) \cdot \frac{r}{\ell - 2}.
\]

Adding up these measures for \( \ell = k, k + 1, \ldots, k_T \), we obtain (19). \( \square \)

7.2 Lemma 6.3

To prove Lemma 6.3, we will require some additional notation and preliminary lemmas. Let \( T = (T, d_T, \mu_T) \) be a metric tree. We will define a nested sequence of metric subforests \( F_k \) of \( T \). That is, each \( F_k = (F_k, d_{F_k}, \mu_{F_k}) \) consists of a subset \( F_k \subseteq T \) such that each connected component of \( F_k \) is a metric tree. For the sake of convenience, we will not assume the measure on each component of \( F_k \) is a probability measure. The metric \( d_{F_k} \) is geodesic distance in \( F_k \) with respect to the restriction of \( d_T \); in particular, the distance between points in distinct connected components of \( F_k \) is \( \infty \), as there is no geodesic joining them. The measure \( \mu_{F_k} \) is simply the restriction of \( \mu_T \).

The subset \( F_0 \) defining the metric forest \( F_0 \) consists of only the leaves of \( T \). To define \( F_1 \), we include the leaf edges of the leaves in \( F_0 \) and their opposite endpoints. That is, \( F_1 \) consists of leaves, leaf edges
Then we can choose some edges. An example of this sequence shown in Figure 6.

Proof. Since $T$ is assumed to be connected, one direction is clear. Conversely, suppose that $F_k \neq T$. We therefore have obtained a contradiction. We therefore assume without loss of generality that the path terminates at a leaf contained in $T_2$. The path must then pass through $v_2$, by construction of $F_k$. By our assumption, we can choose another edge which is incident with $v_2$ is not contained in the path from the previous step, and is not contained in $T_2$. Now choose a new path which starts at $v_2$, passes through the newly chosen edge, and continues by increasing distance from $v_2$. The new path will terminate in a leaf not contained in $T_2$. If its terminal leaf lies in $T_1$, then we concatenate with the path from the first step to produce a non-trivial cycle in $T$, thereby obtaining a contradiction. We therefore assume without loss of generality that the new path terminates at a leaf contained in $T_3$. Continuing with this process, a nontrivial cycle must eventually be formed because there are finitely many $T_j$, and we have obtained a contradiction.

Corollary 7.1. With $T$ and $F_k$ defined as above, there exists some $N$ such that $F_k = T$ for all $k \geq N$.

Proof. If $F_k \neq T$, then $V_k \neq \emptyset$, by Lemma 7.3. It follows that an edge is added to form $F_{k+1}$. Since there are finitely many edges in $T$, it must be that $F_N = T$ for some $N$.

We remark that it is possible to show that the number of trees in $F_{k+1}$ is strictly less than the number of trees in $F_k$ whenever $F_k \neq T$, and it follows that $N$ is at most the number of leaves in $T$.

Corollary 7.2. With $T$ and $F_k$ defined as above, $F_k = T$ if and only if $F_k$ is connected.

Proof. Since $T$ is assumed to be connected, one direction is clear. Conversely, suppose that $F_k \neq T$. Then we can choose some $v \in V_k$, by Lemma 7.3. Let $e$ denote the single edge which is incident
with $v$ but which is not contained in $F_k$ and let $v'$ denote the opposite node of $e$. If $v' \in F_k$, then $F_k$ is disconnected. Otherwise, choose a path which starts at $v$, passes through $e$ and continues with increasing distance from $v$. The path terminates in a leaf, and the leaf necessarily lies in a path component of $F_k$ which is different than that of $v$.

The proof of the main technical lemma now follows.

**Proof of Lemma 6.3.** Let $T$ and $S$ be metric trees and suppose that $h_N(T) = h_N(S)$ and that the functions contained in this common set are distinct. Let $F_k$ and $G_k$ denote the subforests for $T$ and $S$, respectively, as defined above. Let $V_k$ denote the subset of nodes of $F_k$ as defined above, and let $W_k \subset G_k$ be defined similarly. We will construct an isomorphism $\phi : T \rightarrow S$ by inductively defining isomorphisms $\phi_k : F_k \rightarrow G_k$ on the metric subforests—an *isomorphism of subforests* is a bijection whose restriction to each connected component is an isomorphism of metric trees.

First consider the metric forests $F_0$ and $G_0$ containing only the leaves of their respective trees. For each leaf $v \in F_0$, there is exactly one point $w_v \in S$ such that $h_T(v, \cdot) = h_S(w_v, \cdot)$. We claim that $w_v \in G_0$. This follows from the observation that the $h$-function of a point $x \in T$ has the property that $h_T(x, 0) = 1$ if and only if $x$ is a leaf. The isomorphism (in this case, simply a bijection) $\phi_0 : F_0 \rightarrow G_0$ is the map taking $v$ to $w_v$.

Next consider $F_1$ and $G_1$. Let $v \in N(F_1)$ be a leaf parent with child leaves as $v_1, \ldots, v_n$. Then

$$h_T(v, \cdot) = \begin{cases} h_T(v_j, r + r_j) - (r_j - r) & r \leq r_j \\ h_T(v_j, r + r_j) & r \geq r_j \end{cases}$$

for each leaf $v_j$, where $r_j$ is the length of its leaf edge. There is a unique vertex $w_v \in S$ such that $h_T(v, \cdot) = h_S(w_v, \cdot)$. Moreover, it must be the leaf parent of $\phi_0(v_1), \ldots, \phi_0(v_n)$, since the leaf parent of each $\phi_0(v_j)$ is determined by an explicit equation of the form (20). We extend $\phi_0$ to a map $\phi_1 : N(F_1) \rightarrow N(G_1)$ by sending each $v$ to $w_v$. The map extends to an isomorphism $\phi_1 : F_1 \rightarrow G_1$ by interpolation over edges. To see that this step is valid, consider a leaf $v_j$ with parent node $v$ in $F_1$. The length of the leaf edge of $v_j$ is given by the smallest value of $r > 0$ such that $h_T(v_j, r) \neq 1$, and it follows that the edge joining $\phi_1(v)$ and $\phi_1(v_j)$ must have the same length. We complete this step by noting that there is also an explicit formula for the function $h_{F_1}(v, \cdot)$. Assuming without loss of generality that the leaf edge lengths $r_j$ of the leaves $v_j$ are ordered increasingly as $r_1 < r_2 < \cdots < r_n$, the formula is determined by

$$h_{F_1}(v, r) = \begin{cases} n & 0 \leq r < r_1 \\ n - 1 & r_1 \leq r < r_2 \\ \vdots & \vdots \\ 1 & r_{n-1} \leq r < r_n \\ 0 & r_n \leq r \end{cases}$$

Moreover, we have $h_{G_1}(\phi_1(v), \cdot) = h_{F_1}(v, \cdot)$.

We now proceed inductively. Assume that we have constructed an isomorphism $\phi_k : F_k \rightarrow G_k$ and moreover that for each $v \in N(F_k)$ we have determined the function $h_{F_k}(v, \cdot) = h_{G_k}(\phi_k(v), \cdot)$. If $F_k = T$, then $G_k = S$ and we are done, since $F_k = T$ implies $\phi(F_k) = G_k$ is connected and $G_k = S$ follows from Corollary 7.2. We therefore assume that $F_k \neq T$. Lemma 7.3 implies that $V_k \neq \emptyset$, and we claim that $\phi_k(V_k) = W_k$. Indeed, $v \in V_k$ if and only if $\deg_T(v) = \deg_{F_k}(v) + 1$, and the quantities in this equation are given by $h_T(v, 0)$ and $h_{F_k}(v, 0)$, respectively. Since these quantities are preserved by $\phi$, the claim holds.

Now assume that there is some node $v \in N(F_{k+1})$ which is not contained in $F_k$. Let $v_1, \ldots, v_n \in V_k$ denote nodes which are connected to $v$ and let $r_j$ denote the length of the edge joining $v_j$ to $v$. For each $v_j$, we have

$$h_T(v, r) = \begin{cases} h_T(v_j, r + r_j) - h_{F_k}(v_j, r + r_j) - (r_j - r) & r \leq r_j \\ h_T(v_j, r + r_j) - h_{F_k}(v_j, r + r_j) & r \geq r_j \end{cases}$$

27
As shown above, each $v_j$ maps under $\phi_k$ into $W_k$. It follows that for each such $v$ there is a unique $w_v \in \mathcal{N}(G_{k+1}) \setminus \mathcal{N}(G_k)$ with the same $h$-function. We extend $\phi_k$ to the map on nodes $\phi_{k+1} : \mathcal{N}(\mathcal{F}_{k+1}) \rightarrow \mathcal{N}(G_{k+1})$ given by sending $v$ to $w_v$. Since a formula similar to (21) can be written for each $\phi_k(v_j) \in W_k$, it must be that the elements of $W_k$ which connect to $w_v$ are exactly $\phi_k(v_1), \ldots, \phi_k(v_n)$. Finally, note that the length of the edge connecting $v_j$ to $v$ is the smallest $r > 0$ such that $h'_{\mathcal{F}_k}(v_j, r) - h'_{\mathcal{F}_k}(v_j, r) \neq 1$ and it follows that we can extend $\phi_{k+1}$ to an isomorphism $\phi_{k+1} : \mathcal{F}_{k+1} \rightarrow G_{k+1}$ by interpolating over edges. Finally, we complete the inductive step in this case by deriving a formula for $h_{\mathcal{F}_{k+1}}(v, \cdot)$. Assuming without loss of generality that $r_1 \leq r_2 \leq \cdots \leq r_n$, we have

$$h'_{\mathcal{F}_{k+1}}(v, r) = \begin{cases} 
\frac{n}{n - 1 + h'_{\mathcal{F}_k}(v_1, r - r_1)} & 0 \leq r < r_1 \\
\vdots & \vdots \\
1 + h'_{\mathcal{F}_k}(v_1, r - r_1) + \cdots + h'_{\mathcal{F}_k}(v_1, r - r_{n-1}) & r_{n-1} \leq r < r_n \\
h'_{\mathcal{F}_k}(v_1, r - r_1) + \cdots + h'_{\mathcal{F}_k}(v_1, r - r_n) & r_n \leq r.
\end{cases}$$

It follows that $h_{G_{k+1}}(\phi_{k+1}(v), \cdot) = h_{\mathcal{F}_{k+1}}(v, \cdot)$.

There is one more case to consider: $\mathcal{F}_k \neq \mathcal{T}$, but $\mathcal{F}_k$ contains all nodes of $\mathcal{F}_{k+1}$. Let $v \in V_k$ and let $v'$ denote the opposite endpoint of the single edge which incident on $v$ but not contained in $\mathcal{F}_k$. As above, a formula for $h_{\mathcal{T}}(v', \cdot)$ can be derived explicitly in terms of $h_{\mathcal{T}}(v, \cdot)$ and $h_{\mathcal{F}_k}(v, \cdot)$. Moreover, the length of the edge joining $v$ to $v'$ can be extracted from these functions. This is enough to extend the map $\phi_k$ over this edge by interpolation. Performing this extension for each $v \in V_k$ produces the isomorphism $\phi_{k+1}$.

By Corollary 7.1 this inductive process stabilizes and produces an isomorphism $\phi : \mathcal{T} \rightarrow \mathcal{S}$. \hfill $\square$

References

[1] Pankaj K Agarwal, Kyle Fox, Abhinandan Nath, Anastasios Sidiropoulos, and Yusu Wang. Computing the Gromov-Hausdorff distance for metric trees. In *International Symposium on Algorithms and Computation*, pages 529–540. Springer, 2015.

[2] Reema Al-Aifari, Ingrid Daubechies, and Yaron Lipman. Continuous procrustes distance between two surfaces. *Communications on Pure and Applied Mathematics*, 66(6):934–964, 2013.

[3] P Alestalo, DA Trotsenko, and J Väisälä. Isometric approximation. *Israel Journal of Mathematics*, 125(1):61–82, 2001.

[4] Martin Bauer, Thomas Fidler, and Markus Grasmair. Local uniqueness of the circular integral invariant. *Inverse Problems & Imaging*, 7(1):107–122, 2013.

[5] Serge Belongie, Greg Mori, and Jitendra Malik. Matching with shape contexts. In *Statistics and Analysis of Shapes*, pages 81–105. Springer, 2006.

[6] José R Berrendero, Antonio Cuevas, and Beatriz Pateiro-López. Shape classification based on interpoint distance distributions. *Journal of Multivariate Analysis*, 146:237–247, 2016.

[7] Gary S Bloom. A counterexample to a theorem of s. piccard. *Journal of Combinatorial Theory, Series A*, 22(3):378–379, 1977.

[8] Marco Bonetti and Marcello Pagano. The interpoint distance distribution as a descriptor of point patterns, with an application to spatial disease clustering. *Statistics in medicine*, 24(5):753–773, 2005.

[9] Mireille Boutin and Gregor Kemper. On reconstructing n-point configurations from the distribution of distances or areas. *Advances in Applied Mathematics*, 32(4):709–735, 2004.
[10] Doug M Boyer, Yaron Lipman, Elizabeth St Clair, Jesus Puente, Biren A Patel, Thomas Funkhouser, Jukka Jernvall, and Ingrid Daubechies. Algorithms to automatically quantifying the geometric similarity of anatomical surfaces. *Proceedings of the National Academy of Sciences*, 2011.

[11] Daniel Brinkman and Peter J Olver. Invariant histograms. *American Mathematical Monthly*, 119(1):4–24, 2012.

[12] Dmitri Burago, Yuri Dmitrievich Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33. American Mathematical Soc., 2001.

[13] Mathieu Carrière and Steve Oudot. Local equivalence and intrinsic metrics between reeb graphs. *arXiv preprint arXiv:1703.02901*, 2017.

[14] Claire Chalopin, Gérard Finet, and Isabelle E Magnin. Modeling the 3d coronary tree for labeling purposes. *Medical image analysis*, 5(4):301–315, 2001.

[15] Arnaud Charnoz, Vincent Agnus, Grégoire Malandain, Luc Soler, and Mohamed Tajine. Tree matching applied to vascular system. In *International Workshop on Graph-Based Representations in Pattern Recognition*, pages 183–192. Springer, 2005.

[16] Ricardo García-Pelayo. Pairs of subsets of regular polyhedra with the same distribution of distance. *Applied Mathematical Sciences*, 10(26):1285–1297, 2016.

[17] Natasha Gelfand, Niloy J Mitra, Leonidas J Guibas, and Helmut Pottmann. Robust global registration. In *Symposium on geometry processing*, volume 2, page 5, 2005.

[18] Andreas Greven, Peter Pfaffelhuber, and Anita Winter. Convergence in distribution of random metric measure spaces (λ-coalescent measure trees). *Probability Theory and Related Fields*, 145(1-2):285–322, 2009.

[19] Charles Gueunet, Pierre Fortin, Julien Jomier, and Julien Tierny. Task-based augmented merge trees with fibonacci heaps. In *Large Data Analysis and Visualization (LDAV), 2017 IEEE 7th Symposium on*, pages 6–15. IEEE, 2017.

[20] Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent. Optimal mass transport for registration and warping. *International Journal of computer vision*, 60(3):225–240, 2004.

[21] Leonid V Kantorovich. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201, 1942.

[22] Paul Lemke, Steven S Skiena, and Warren D Smith. Reconstructing sets from interpoint distances. In *Discrete and computational geometry*, pages 597–631. Springer, 2003.

[23] CL Mallows and JMC Clark. Linear-intercept distributions do not characterize plane sets. *Journal of Applied Probability*, 7(1):240–244, 1970.

[24] Jeremy L Martin, Matthew Morin, and Jennifer D Wagner. On distinguishing trees by their chromatic symmetric functions. *Journal of Combinatorial Theory, Series A*, 115(2):237–253, 2008.

[25] Facundo Mémoli. Gromov-hausdorff distances in euclidean spaces. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW’08. IEEE Computer Society Conference on*, pages 1–8. IEEE, 2008.

[26] Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011.
[27] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l’Académie Royale des Sciences de Paris*, 1781.

[28] Dmitriy Morozov, Kenes Beketayev, and Gunther Weber. Interleaving distance between merge trees. *Discrete and Computational Geometry*, 49(22-45):52, 2013.

[29] Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Shape distributions. *ACM Transactions on Graphics (TOG)*, 21(4):807–832, 2002.

[30] Steve Oudot and Elchanan Solomon. Barcode embeddings, persistence distortion, and inverse problems for metric graphs. *arXiv preprint arXiv:1712.03630*, 2017.

[31] Halsey Lawrence Royden. *Real analysis*. Macmillan New York, 1988.

[32] Luis Antonio Santaló. *Introduction to integral geometry*, volume 1198. Hermann, 1953.

[33] Yonggang Shi, Paul M Thompson, Greig I de Zubicaray, Stephen E Rose, Zhuowen Tu, Ivo Dinov, and Arthur W Toga. Direct mapping of hippocampal surfaces with intrinsic shape context. *NeuroImage*, 37(3):792–807, 2007.

[34] Karl-Theodor Sturm. On the geometry of metric measure spaces. *Acta mathematica*, 196(1):65–131, 2006.

[35] Karl-Theodor Sturm. The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces. *arXiv preprint arXiv:1208.0434*, 2012.

[36] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.