

# Comparison Results for Gromov-Wasserstein and Gromov-Monge Distances

Facundo Mémoli<sup>1</sup> and Tom Needham<sup>2</sup>

<sup>1</sup>Department of Mathematics, The Ohio State University, [facundo.memoli@gmail.com](mailto:facundo.memoli@gmail.com)

<sup>2</sup>Department of Mathematics, Florida State University, [tneedham@fsu.edu](mailto:tneedham@fsu.edu)

## Abstract

Inspired by the Kantorovich formulation of optimal transport distance between probability measures on a metric space, Gromov-Wasserstein (GW) distances comprise a family of metrics on the space of isomorphism classes of metric measure spaces. In previous work, the authors introduced a variant of this construction which was inspired by the original Monge formulation of optimal transport; elements of the resulting family are referred to Gromov-Monge (GM) distances. These GM distances, and related ideas, have since become a subject of interest from both theoretical and applications-oriented perspectives. In this note, we establish several theoretical properties of GM distances, focusing on comparisons between GM and GW distances. In particular, we show that GM and GW distances are equal for non-atomic metric measure spaces. We also consider variants of GM distance, such as a Monge version of Sturm's  $L_p$ -transportation distance, and give precise comparisons to GW distance. Finally, we establish bi-Hölder equivalence between GM distance and an isometry-invariant Monge optimal transport distance between Euclidean metric measure spaces that has been utilized in shape and image analysis applications.

## 1 Introduction

Gromov-Wasserstein (GW) distance is a metric which allows one to compare probability measures defined over *different* metric spaces. This is a task which is necessary in many applications in shape analysis and machine learning, and GW distance has consequently become popular in these fields; see [35, 3, 12, 46, 17, 21], among many others. A primary reason for the usefulness of this framework is that computation of GW distance involves finding a soft correspondence between points of the two metric spaces being compared (more precisely, a probability coupling; see below), which gives a meaningful registration of the spaces for downstream analysis. On the other hand, it is sometimes desirable to require an exact point registration (i.e., a function) between the spaces, and this led us to define a variant of GW distance, called Gromov-Monge (GM) distance, in our previous paper [33]. Roughly, GM defines an optimization problem with the same objective as GW, but restricts the feasible set from soft correspondences to exact correspondences, so that  $\text{GW} \leq \text{GM}$ , in general. The GM distance, and related ideas, have since appeared in both theoretical and applied contexts [42, 38, 22, 26, 47] (see Section 2.3 for a detailed literature review). The purpose of the present paper is to collect some results on GM distance which have not appeared elsewhere<sup>1</sup>. The overarching theme is to address the following **main question**: what are conditions on the metric spaces/measures being compared which imply that their GW and GM distances agree?

In order to describe our results, we now introduce some precise definitions. A *metric measure space (mm-space)* is a triple  $\mathcal{X} = (X, d_X, \mu_X)$  such that  $(X, d_X)$  is a metric space and  $\mu_X$  is a fully supported Borel probability measure on  $X$ . For  $p \in [1, \infty)$ , the *Gromov-Wasserstein  $p$ -distance* between two such structures  $\mathcal{X}$  and  $\mathcal{Y}$  is given by

$$\text{GW}_p(\mathcal{X}, \mathcal{Y}) := \inf_{\pi} \left( \iint_{(X \times Y)^2} |d_X(x, x') - d_Y(y, y')|^p \pi(dx \times dy) \pi(dx' \times dy') \right)^{\frac{1}{p}},$$

where the infimum is over *couplings* of  $\mu_X$  and  $\mu_Y$ ; that is, joint probability measures  $\pi$  on  $X \times Y$  whose marginals are  $\mu_X$  and  $\mu_Y$ . This defines a metric on the space of mm-spaces up to a suitable notion of isomorphism, and extends in a natural way to  $p = \infty$ . As it is defined as an optimization problem over the space of couplings,

<sup>1</sup>Some of these results appeared in early versions of the arXiv version of our paper [33], but were removed from the published version for the sake of brevity.

GW distance is reminiscent of Kantorovich's formulation of optimal transport, which gives rise to Wasserstein  $p$ -distances (see [44, 34] for general references on optimal transport theory). If we instead extend Monge's original formulation of optimal transport, we are led to the notion of *Gromov-Monge  $p$ -distance*,

$$\text{GM}_p(\mathcal{X}, \mathcal{Y}) := \inf_{\phi} \left( \iint_{X \times X} |d_X(x, x') - d_Y(\phi(x), \phi(x'))|^p \mu_X(dx) \mu_X(dx') \right)^{\frac{1}{p}},$$

where the infimum is now over functions  $\phi : X \rightarrow Y$  which push the measure  $\mu_X$  forward to  $\mu_Y$ . It is straightforward to show that  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) \leq \text{GM}_p(\mathcal{X}, \mathcal{Y})$ , in general (Remark 2.10).

The results of this paper explore connections between GW and GM distances (and related constructions) under various assumptions on the spaces being compared. In fact, we work in the more general setting of *measure networks*; that is, triples  $\mathcal{X} = (X, \omega_X, \mu_X)$  where the *network function*  $\omega_X : X \times X \rightarrow \mathbb{R}$  is an arbitrary measurable map (Definition 2.1). Our main contributions are:

- We show that  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) = \text{GW}_p(\mathcal{X}, \mathcal{Y})$  for measure networks with non-atomic probability measures (Theorem 2). This generalizes a result in the recent paper [25] and is analogous to results which show the equivalence of Kantorovich and Monge (classical) optimal transport for non-atomic measures [5, 36]. This is also related to a growing body of work which seeks to characterize classes of measure networks where  $\text{GW}_p$  is realized by a Monge map [40, 42, 38]; in particular, recent work of Dumont, Lacombe and Vialard gives a complete solution for certain measure networks over Euclidean spaces [22]. This line of work is inspired by a famous result of Brenier for classical optimal transport [10].
- For measure networks over finite, uniformly distributed probability spaces with symmetric positive definite network functions, we show  $\text{GM}_2 = \text{GW}_2$  (Proposition 3.8); in particular,  $\text{GW}_2$  is always realized by a Monge map in this setting. This has applications to, e.g., graph analysis, where  $\omega$  can be the heat kernel. Our result is in the spirit of a result in [42], which applies to conditionally negative definite network functions.
- In the mm-space setting, we show that  $\text{GW}_p$  is equal to a variant of  $\text{GM}_p$  which allows for a certain notion of *mass splitting* (Theorem 3). This refines various observations about GW distance that have appeared previously [40, 18, 16].
- We define an alternate version of GM distance for mm-spaces based on embeddings into a common ambient space, in analogy with Sturm's  $L_p$ -transportation distance [39], and show that the embedding GM distance and Sturm's distance coincide for non-atomic mm-spaces (Theorem 4). Moreover, we show that the GM (as defined above) gives a lower bound on the embedding version, in general (Theorem 5).
- Previous works in shape and image analysis consider isometry-invariant versions of Monge optimal transport [24, 9, 1]. We show that isometry-invariant Monge distance for Euclidean mm-spaces is bi-Hölder equivalent to the embedding formulation of GM distance (Theorem 6).

The structure of the paper is as follows. In Section 2, we give the necessary background definitions and survey recent literature. Section 3 presents results on equivalence of GW and GM distances for certain classes of measure networks. Finally, in Section 4, we introduce the embedding formulation of GM distance, give comparison results with the corresponding GW version and show bi-Hölder equivalence with isometry-invariant Monge distances for Euclidean spaces.

## 2 Gromov-Wasserstein and Gromov-Monge Distances

This section introduces our basic notation and gives precise definitions of various optimal transport-based distances.

### 2.1 Measure Networks and Metric Measure Spaces

As was described above, the original work in Gromov-Wasserstein (GW) distances [30] concerned comparison of mm-spaces, but the purview of the theory has since been extended [15, 41, 20, 14]. In this paper, we will work in the following setting, borrowing terminology from [15].

**Definition 2.1.** A *measure network (m-net)* is a triple  $\mathcal{X} = (X, \omega_X, \mu_X)$  such that  $X$  is a separable and completely metrizable topological space (or **Polish space**),  $\mu_X$  is a fully-supported Borel probability measure and  $\omega_X : X \times X \rightarrow \mathbb{R}$  is a measurable function, which we refer to as the **network function**. We use  $\mathbb{N}$  to denote the class of measure networks.

In the case that the network function is a metric on  $X$ , we will typically denote it as  $d_X$ , we will assume that  $d_X$  generates the topology of  $X$  and we will refer to the triple  $\mathcal{X} = (X, d_X, \mu_X)$  as a **metric measure space (mm-space)**. We denote the class of mm-spaces by  $\mathbf{M}$ .

**Examples 2.2.** There are many useful examples of  $m$ -nets which are not mm-spaces. Here are a few:

- It is frequently useful to consider network functions which satisfy a subset of the metric axioms. For instance, a **pseudo-metric** is a function  $\omega_X : X \times X \rightarrow \mathbb{R}$  satisfying all of the metric axioms, besides allowing  $\omega_X(x, x') = 0$  for  $x \neq x'$ . Pseudo-metric measure spaces arise naturally as elements of the completion of the space of metric measure spaces with respect to Gromov-Wasserstein distances [40].
- A graph  $G = (V, E)$  with (finite) vertex set  $V$  and edge set  $E$  can be represented as an  $m$ -net by setting  $X = V$ , taking  $\mu_X$  to be uniform and defining  $\omega(x, x') \in \{0, 1\}$  to be 1 if and only if  $\{x, x'\} \in E$  [45].
- The previous example can be generalized by taking  $\omega$  to be some other graph kernel, such as the graph Laplacian or a heat kernel [19].

There are two notions of equivalence of measure networks which are relevant to the GW framework. We recall that a measurable function  $\phi : X \rightarrow Y$  between measure spaces  $(X, \mu_X)$  and  $(Y, \mu_Y)$  is **measure-preserving** if  $\phi_{\#}\mu_X = \mu_Y$ , where  $\phi_{\#}\mu_X(A) := \mu_X(\phi^{-1}(A))$ , for any measurable  $A \subset Y$ .

**Definition 2.3.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be  $m$ -nets. A **strong isomorphism** from  $\mathcal{X}$  to  $\mathcal{Y}$  is a measure-preserving bijective map  $\phi : X \rightarrow Y$  with measure-preserving inverse, such that  $\omega_Y(\phi(x), \phi(x')) = \omega_X(x, x')$  for all  $(x, x') \in X \times X$ .

**Definition 2.4.** Measure networks  $\mathcal{X}$  and  $\mathcal{Y}$  are **weakly isomorphic** if there exists a Borel probability space  $(Z, \mu_Z)$ , together with measure-preserving maps  $\phi_X : Z \rightarrow X$  and  $\phi_Y : Z \rightarrow Y$  such that  $\omega_X(\phi_X(z), \phi_X(z')) = \omega_Y(\phi_Y(z), \phi_Y(z'))$  holds for  $\mu_Z \otimes \mu_Z$ -almost every  $(z, z') \in Z \times Z$ .

Trivially, if  $\mathcal{X}$  and  $\mathcal{Y}$  are strongly isomorphic then they are weakly isomorphic. The converse holds if  $\mathcal{X}$  and  $\mathcal{Y}$  are mm-spaces, but not in general (see Theorem 1 below).

## 2.2 Distances Between Measures and Measure Networks

Optimal transport distances are defined in terms of certain joint probability measures called couplings, which we now recall.

**Definition 2.5.** Let  $(X, \mu_X)$  and  $(Y, \mu_Y)$  be probability spaces. A **coupling** of  $\mu_X$  and  $\mu_Y$  is a probability measure  $\pi$  on  $X \times Y$  whose marginals are  $\mu_X$  and  $\mu_Y$ , respectively. That is, let  $\rho_X : X \times Y \rightarrow X$  and  $\rho_Y : X \times Y \rightarrow Y$  be coordinate projections; then a coupling satisfies  $(\rho_X)_{\#}\pi = \mu_X$  and  $(\rho_Y)_{\#}\pi = \mu_Y$ . The collection of all couplings of  $\mu_X$  and  $\mu_Y$  will be denoted  $\mathcal{C}(\mu_X, \mu_Y)$ .

For a measure-preserving map  $\phi : X \rightarrow Y$ , let  $\text{id}_X \times \phi : X \rightarrow X \times Y$  be defined by  $(\text{id}_X \times \phi)(x, x') = (x, \phi(x'))$ . We define the **coupling induced by  $\phi$**  to be  $\pi_{\phi} := (\text{id}_X \times \phi)_{\#}\mu_X \in \mathcal{C}(\mu_X, \mu_Y)$ . Let  $\mathcal{T}(\mu_X, \mu_Y)$  denote the collection of all measure-preserving mappings from  $(X, \mu_X)$  to  $(Y, \mu_Y)$  and let

$$\mathcal{T}_{\#}(\mu_X, \mu_Y) := \{\pi_{\phi} \mid \phi \in \mathcal{T}(\mu_X, \mu_Y)\} \subset \mathcal{C}(\mu_X, \mu_Y)$$

denote the set of couplings induced by measure-preserving mappings.

**Remark 2.6.** The set  $\mathcal{C}(\mu_X, \mu_Y)$  is always nonempty, since it contains the product measure  $\mu_X \otimes \mu_Y$ . However, the set  $\mathcal{T}(\mu_X, \mu_Y)$  (or  $\mathcal{T}_{\#}(\mu_X, \mu_Y)$ ) can be empty. For example, if  $\mu_X$  is a Dirac measure on a single point and  $\mu_Y$  is a sum of Dirac measures  $\frac{1}{2}(\delta_y + \delta_{y'})$  for  $y \neq y'$ , then there is no measure-preserving map  $\phi : X \rightarrow Y$ .

With the notion of coupling in hand, we can now define some classical optimal transport distances between measures over the same metric space. We do so in the setting of metric spaces, since this level of generality will be sufficient for our purposes. For background on classical optimal transport, see, e.g., [44]. In the following, we allow our distances to take the value  $\infty$ ; that is, they may be **extended metrics**.

**Definition 2.7.** Let  $(X, d_X)$  be a metric space and let  $\mu$  and  $\nu$  be Borel probability measures on  $X$ . The **Wasserstein  $p$ -distance** is defined for  $p \in [1, \infty)$  by

$$W_p(\mu, \nu)^p := \inf_{\pi \in \mathcal{C}(\mu, \nu)} \int_{X \times X} d_X(x, x')^p \pi(dx \times dx') = \inf_{\pi \in \mathcal{C}(\mu, \nu)} \|d_X\|_{L^p(\pi)}^p$$

and for  $p = \infty$  by

$$W_\infty(\mu, \nu) := \inf_{\pi \in \mathcal{C}(\mu, \nu)} \sup\{d_X(x, x') \mid (x, x') \in \text{supp}(\pi)\} = \inf_{\pi \in \mathcal{C}(\mu, \nu)} \|d_X\|_{L^\infty(\text{supp}(\pi))},$$

where  $\text{supp}(\pi)$  is the support of  $\pi$ .

We likewise define the **Monge  $p$ -distance** by restricting the feasible set of couplings; that is,

$$M_p(\mu, \nu) := \begin{cases} \inf_{\pi \in \mathcal{T}_\#(\mu, \nu)} \|d_X\|_{L^p(\pi)} & p < \infty \\ \inf_{\pi \in \mathcal{T}_\#(\mu, \nu)} \|d_X\|_{L^\infty(\text{supp}(\pi))} & p = \infty, \end{cases}$$

where we take the infimum over the empty set to be infinity.

When the underlying metric space needs to be emphasized, we will write  $W_p^X = W_p$  and  $M_p^X = M_p$ .

These distances have analogues which measure the distance between pairs of mm-spaces or measure networks. Since the underlying spaces are allowed to vary, we need to measure the quality of a coupling in a relative manner, leading to the following definition.

**Definition 2.8.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be measure networks and  $p \in [1, \infty)$ . The  **$p$ -distortion** of a coupling  $\pi \in \mathcal{C}(\mu_X, \mu_Y)$  is

$$\begin{aligned} \text{dis}_p(\pi)^p &= \text{dis}_{p, \mathcal{X}, \mathcal{Y}}(\pi)^p := \iint_{(X \times Y)^2} |\omega_X(x, x') - \omega_Y(y, y')|^p \pi(dx \times dy) \pi(dx' \times dy') \\ &= \|\omega_X - \omega_Y\|_{L^p(\pi \otimes \pi)}^p. \end{aligned}$$

Similarly, for  $p = \infty$ , we define

$$\text{dis}_\infty(\pi) = \text{dis}_{\infty, \mathcal{X}, \mathcal{Y}}(\pi) := \sup\{|\omega_X(x, x') - \omega_Y(y, y')| \mid (x, y), (x', y') \in \text{supp}(\pi)\}.$$

We define the  **$p$ -distortion** of a measure-preserving map  $\phi \in \mathcal{T}(\mu_X, \mu_Y)$  to be  $\text{dis}_p(\phi) = \text{dis}_p(\pi_\phi)$ , with  $\pi_\phi$  as in Definition 2.5. The expression for the distortion simplifies in the  $p < \infty$  case as

$$\text{dis}_p(\phi)^p = \iint_{X \times X} |\omega_X(x, x') - \omega_Y(\phi(x), \phi(x'))|^p \mu_X(dx) \mu_X(dx'),$$

and similarly in the  $p = \infty$  case.

Finally, we define our distances between measure networks. These agree in the mm-space setting with the definitions given in the introduction.

**Definition 2.9.** For  $p \in [1, \infty]$ , the **Gromov-Wasserstein  $p$ -distance** is the function  $\text{GW}_p : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R} \cup \{\infty\}$  defined by

$$\text{GW}_p(\mathcal{X}, \mathcal{Y}) := \inf_{\pi \in \mathcal{C}(\mu_X, \mu_Y)} \text{dis}_p(\pi).$$

The **Gromov-Monge  $p$ -distance** is the function  $\text{GM}_p : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R} \cup \{\infty\}$  defined by

$$\text{GM}_p(\mathcal{X}, \mathcal{Y}) := \inf_{\phi \in \mathcal{T}(\mu_X, \mu_Y)} \text{dis}_p(\phi) = \inf_{\pi_\phi \in \mathcal{T}_\#(\mu_X, \mu_Y)} \text{dis}_p(\pi_\phi),$$

where we once again declare the infimum over the empty set to be infinity.

**Remark 2.10.** Since  $\text{GW}_p$  involves optimization over a larger set of couplings than  $\text{GM}_p$ , the inequality  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) \leq \text{GM}_p(\mathcal{X}, \mathcal{Y})$  always holds.

**Remark 2.11.** The infimum in the definition of  $\text{GW}_p$  is actually a minimum. This is proved in [15, Theorem 2.2] for m-nets whose network functions are bounded and whose measures are fully supported, and in [40, Lemma 1.7] for m-nets whose network functions are metrics, i.e., for mm-spaces (without any boundedness or support assumptions). Working through the proofs, one sees that the various assumptions are not necessary, and the result goes through in the general setting of Definition 2.1.

In this paper, we consider m-nets with potentially unbounded network functions, meaning that  $\text{GW}_p$  is not guaranteed, in general, to be finite. It is sometimes useful to restrict to subspaces of m-nets with additional control on the network functions. It is straightforward to see that  $\text{GW}_p$  is finite when restricted to  $\mathbb{N}_p \times \mathbb{N}_p$ , where  $\mathbb{N}_p$  is as defined below.

**Definition 2.12.** For  $p \in [1, \infty)$  and an  $m$ -net  $\mathcal{X}$ , define the  $p$ -size of  $\mathcal{X}$ ,  $\text{size}_p(\mathcal{X}) \in \mathbb{R} \cup \{\infty\}$ , by

$$\text{size}_p(\mathcal{X})^p := \iint_{X \times X} |\omega_X(x, x')|^p \mu_X(dx) \mu_X(dx').$$

Similarly, let

$$\text{size}_\infty(\mathcal{X}) := \sup_{x, x' \in X} |\omega_X(x, x')|.$$

We denote the class of  $m$ -nets (respectively,  $mm$ -spaces)  $\mathcal{X}$  with  $\text{size}_p(\mathcal{X}) < \infty$  by  $\mathbb{N}_p$  (respectively,  $\mathbb{M}_p$ ).

**Remark 2.13.** The terminology above is used in, e.g., [40]. In the setting of  $mm$ -spaces,  $\text{size}_p(\mathcal{X})$  is sometimes referred to as the  $p$ -diameter of  $\mathcal{X}$ . It appears in certain estimates of GW distance—see [32, Theorem 5.1]. In particular, letting  $*$  denote the 1-point  $mm$ -space,

$$\text{GW}_p(\mathcal{X}, *) = \text{size}_p(\mathcal{X}) = \text{GM}_p(\mathcal{X}, *),$$

where the first equality appears in [32, Theorem 5.1 (f)] and the second follows because the unique coupling of  $\mathcal{X}$  and  $*$  is induced by a measure-preserving map.

The following theorem summarizes some basic results on metric properties of GM and GW distances in the literature. We recall some terminology from metric geometry. For a set  $Z$ , a function  $d : Z \times Z \rightarrow \mathbb{R} \cup \{\infty\}$  is a **Lawvere metric** if it satisfies the triangle inequality and  $d(z, z) = 0$  for all  $z \in Z$ —this terminology refers to classic work of Lawvere, which characterizes such structures in the language of enriched category theory [27]. A Lawvere metric is a **pseudometric** if it is, in addition, finite-valued and symmetric (i.e., it satisfies the axioms of a metric, except it is possible that  $d(z, z') = 0$  for  $z \neq z'$ ), as in Example 2.2.

**Theorem 1.** Let  $p \in [1, \infty]$ .

- On the space  $\mathbb{M}_p$ , the Gromov-Wasserstein  $p$ -distance defines a pseudometric such that  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) = 0$  if and only if  $\mathcal{X}$  and  $\mathcal{Y}$  are strongly isomorphic [30, 40].
- On the space  $\mathbb{M}$ , the Gromov-Monge  $p$ -distance defines a Lawvere metric such that  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) = 0$  if and only if  $\mathcal{X}$  and  $\mathcal{Y}$  are strongly isomorphic. It may take the value  $\infty$  even on  $\mathbb{M}_p$  [33].
- On the space  $\mathbb{N}_p$ , the Gromov-Wasserstein  $p$ -distance defines a pseudometric such that  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) = 0$  if and only if  $\mathcal{X}$  and  $\mathcal{Y}$  are weakly isomorphic [15].

**Remark 2.14.** Let us comment on some subtleties of the results referenced in Theorem 1. When  $\text{GW}_p$  was first introduced, it was shown to define a metric up to strong isomorphism on the space of compact  $mm$ -spaces [30, Proposition 6]. This was later extended to  $\mathbb{M}_p$  in [40, Proposition 1.12]; the proof is essentially the same as in the compact case. The case of  $\text{GW}_p$  for  $m$ -nets was formalized in [15], where it was shown to define a metric on  $\mathbb{N}_\infty$  ( $m$ -nets with bounded network functions), considered up to weak isomorphism; going through the proof there, one sees that it extends immediately to  $\mathbb{N}_p$ . Similarly, [33, Theorem 3] shows that  $\text{GM}_p$  defines a Lawvere metric on the space of strong isomorphism classes of compact  $mm$ -spaces, but the proof extends to all of  $\mathbb{M}$  without change.

The metric properties for  $\text{GM}_p$  on the space of measure networks are more subtle.

**Proposition 2.15.** On the space  $\mathbb{N}$ ,  $\text{GM}_p$  defines a Lawvere metric such that if  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) = 0$  then  $\mathcal{X}$  and  $\mathcal{Y}$  are weakly isomorphic and if  $\mathcal{X}$  and  $\mathcal{Y}$  are strongly isomorphic, then  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) = 0$ . However, it is possible for weakly isomorphic  $\mathcal{X}$  and  $\mathcal{Y}$  to have  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) > 0$ .

*Proof.* That  $\text{GM}_p(\mathcal{X}, \mathcal{X}) = 0$  is obvious and the proof of the triangle inequality follows as in the  $mm$ -space case [33], hence  $\text{GM}_p$  is a Lawvere metric. By Remark 2.10,  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) = 0$  implies  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) = 0$ , hence  $\mathcal{X}$  and  $\mathcal{Y}$  are weakly isomorphic. It is also easy to show that if  $\mathcal{X}$  and  $\mathcal{Y}$  are strongly isomorphic, then  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) = 0$ .

It remains to show that weak isomorphism of  $\mathcal{X}$  and  $\mathcal{Y}$  does not necessarily imply  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) = 0$ . This can be done by example: take  $X = \{x\}$  with uniform measure and  $\omega_X$  identically zero, and  $Y = \{y, y'\}$  with uniform measure and  $\omega_Y$  identically zero. It follows by a (trivial) computation that  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) = 0$ , hence  $\mathcal{X}$  and  $\mathcal{Y}$  are weakly isomorphic, but the set of measure preserving maps  $\mathcal{T}(\mu_X, \mu_Y)$  is empty, hence  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) = \infty$ .  $\square$

**Example 2.16.** With slightly more work, one can construct weakly isomorphic spaces  $\mathcal{X}$  and  $\mathcal{Y}$  such that  $\text{GM}_p(\mathcal{X}, \mathcal{Y})$  is finite but nonzero. Indeed, take

$$X = \{x, x', x''\}, \quad \mu_X(x) = \frac{1}{2}, \mu_X(x') = \mu_X(x'') = \frac{1}{4}, \quad \omega_X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{matrix} x \\ x' \\ x'' \end{matrix}$$

and

$$Y = \{y, y', y''\}, \quad \mu_Y(y) = \mu_Y(y') = \frac{1}{4}, \mu_Y(y'') = \frac{1}{2}, \quad \omega_Y = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{matrix} y \\ y' \\ y'' \end{matrix}.$$

Then one can show that  $X$  and  $Y$  are weakly isomorphic, but both of the two possible preserving maps  $\phi : X \rightarrow Y$  have positive distortion.

In Section 3, we give more precise comparisons between Gromov-Wasserstein and Gromov-Monge distances under additional assumptions on the measure networks.

### 2.3 Related Work

We defined Gromov-Monge distances between mm-spaces in our previous paper [33]. There, the focus was on restrictions to various subcategories of mm-spaces, and GM distances were mainly used as a framing device to motivate certain inverse problems for mm-spaces. Gromov-Monge distances have since arisen in theoretical and applied contexts, and we survey those appearances here.

Several articles consider the following problem, which is inspired by the well known work of Brenier in the classical optimal transport setting [10].

**Problem 2.17** (Monge Map Problem for GW). *Given a class of measure networks, determine whether Gromov-Wasserstein distance between any two elements is realized by a Monge map. That is, for a class  $\mathcal{C}$  of measure networks, we would like to show whether or not it is the case that for all  $\mathcal{X}, \mathcal{Y} \in \mathcal{C}$ , there exists  $\phi \in \mathcal{T}(\mu_X, \mu_Y)$  such that  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) = \text{disp}_p(\phi)$  (for some  $p \in [1, \infty]$ ).*

Finding a class of mm-spaces with an affirmative answer to Problem 2.17 was first posed by Sturm as a “Challenge” in [40, Challenge 3.6]. Moreover, [40, Challenge 5.27] asks one to solve the problem specifically for the class of finite mm-spaces of fixed cardinality, with uniform measures. The first solution to Problem 2.17 appears as [40, Theorem 9.21] for the class of measure networks  $\mathcal{X}$  with  $X \subset \mathbb{R}^n$ ,  $\omega_X(x, x') = \|x - x'\|^2$  (squared Euclidean distance) and  $\mu_X$  absolutely continuous with respect to Lebesgue measure and rotationally invariant. This result, and all other results described in the remainder of this subsection, are valid for  $\text{GW}_2$ , specifically.

A solution to the Monge map problem for the class measure networks  $\mathcal{X}$  with  $X$  a finite subset of the real line of some fixed finite cardinality,  $\omega_X(x, x') = |x - x'|^2$  and  $\mu_X$  uniform was proposed in [43], with a view toward a “sliced” approximation of GW distance (similar to **sliced Wasserstein distances** [8]). In particular, it was claimed that the optimal Monge map is always order-preserving or order-reversing. However, a counterexample to this stronger claim was recently demonstrated [6]. The solution to Problem 2.17 nonetheless holds in this setting, and was shown in the thesis of Vayer to also hold for higher-dimensional finite Euclidean mm-spaces [42, Theorem 4.1.2]. That is, for the class of mm-spaces  $\mathcal{X}$  with  $X$  a subset of a Euclidean space (of arbitrary finite dimension) of fixed finite cardinality,  $\omega_X(x, x') = \|x - x'\|^2$  and  $\mu_X$  uniform, it was shown that GW distance is always realized by a permutation (i.e., a Monge map).

In [42], Vayer also considered Euclidean measure networks whose measures are assumed to be compactly supported and to have density with respect to Lebesgue measure, and whose network functions are squared Euclidean distance—we refer to this as the **quadratic Euclidean class**—and Euclidean measure networks whose network functions are (standard) inner products—we refer to this as the **inner product Euclidean class**. Vayer proves the existence of Monge maps under additional assumptions (the existence of an optimal coupling with certain properties) in both of these classes [42, Theorem 4.2.3 and Proposition 4.2.4]. Recently, general detailed solutions to Problem 2.17 were obtained in the quadratic and inner product Euclidean classes by Dumont, Lacombe, and Vialard in [22, Theorems 3.2 and 3.6], as consequences of technical theorems on existence of Monge maps for optimal transport costs defined by submersions. Finally, in [38], Salmona, Delon and Desolneux study GW distances between measure networks in the quadratic and inner product Euclidean classes whose measures are Gaussians.

An explicit formula is derived in the inner product class [38, Proposition 4.1] and it is shown that if one restricts the feasible set of couplings to those couplings which are themselves Gaussian, the same solution is optimal in the quadratic class [38, Theorem 4.1].

Gromov-Monge distances, or closely related variants, have recently appeared in the machine learning literature. In [26], the GM distance is symmetrized by defining a new distance, called **Reversible Gromov-Monge (RGM) distance**, which involves optimization over a pair of measure-preserving maps satisfying a certain consistency condition. In particular, one always has  $\text{GM}_2(\mathcal{X}, \mathcal{Y}) \leq \text{RGM}(\mathcal{X}, \mathcal{Y})$ . Several theoretical properties of RGM distance are derived, and applications to simulation-based inference are described, where the idea is to use optimal measure-preserving maps to design transform samplers. Similarly, a two-map variant of GM distance is introduced in [47], where the main distinction from RGM is the inclusion of a term which penalizes the maps according to how far they are from being inverses to one another. A benefit of GM-type distances, which is taken advantage of in both [26] and [47], is that the spaces of admissible mappings can be parameterized as neural networks. Learning optimal measure-preserving maps can then take advantage of efficient neural network training algorithms.

Finally, we note that isometry-invariant versions of Monge distance have appeared in the literature previously in the context of registering images [24] and anatomical surfaces [9, 1]. We show in Section 4.3 that these metrics are bi-Hölder equivalent to an alternative version of GM distance, as defined in Section 4.1. More details on these isometry-invariant metrics and their connections to GM distances are provided below in Remark 4.5.

### 3 Comparisons Between Gromov-Monge and Gromov-Wasserstein Distances

A natural relaxation of the Monge Map Problem 2.17 is:

**Problem 3.1** (Equality of GW and GM). *Given a class of measure networks  $\mathcal{C}$ , determine whether  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) = \text{GM}_p(\mathcal{X}, \mathcal{Y})$  for all  $\mathcal{X}, \mathcal{Y} \in \mathcal{C}$ .*

This is indeed a relaxation of Problem 2.17; if there exists a measure-preserving map  $\phi : X \rightarrow Y$  such that  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) = \text{dis}_p(\phi)$ , then  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) \geq \text{GM}_p(\mathcal{X}, \mathcal{Y})$  and Remark 2.10 implies that this is actually an equality. This section treats Problem 3.1 and other related problems.

#### 3.1 Non-Atomic Spaces

We first address the question of equality of Gromov-Wasserstein and Gromov-Monge distances in the setting of non-atomic spaces—recall that a measure is **non-atomic** if it assigns zero mass to any singleton.

**Theorem 2.** *Let  $\mathcal{X}$  and  $\mathcal{Y}$  be measure networks in  $\mathbb{N}_p$  such that  $\mu_X$  and  $\mu_Y$  are non-atomic. Then, for  $p \in [1, \infty)$ ,  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) = \text{GW}_p(\mathcal{X}, \mathcal{Y})$ .*

This result is analogous to a theorem of Pratelli [36], who showed equality of Wasserstein and Monge optimal transport for nonatomic measures when the underlying cost is assumed to be continuous, generalizing a result of Ambrosio [5, Theorem 2.1] (remarkably, Pratelli’s theorem allows for unbounded costs, which can even take the value  $\infty$ ). Our theorem generalizes recent work of Hur, Guo and Liang [25], which shows equality of  $\text{GW}_2$  and  $\text{GM}_2$  for non-atomic measure networks with continuous and bounded network functions whose underlying spaces are subsets of  $\mathbb{R}^d$  (this follows by combining Theorem 3 and Proposition 1 of [25]). Notably, we make no continuity or boundedness assumption on network functions in Theorem 2. In [26, Theorem 5.5], from the published version of [25], the authors give a refinement of Theorem 2, at a similar level of generality, citing a preprint version of the present paper as an inspiration for the proof of their result.

Let  $I = [0, 1]$  and let  $\lambda$  denote Lebesgue measure, restricted to  $I$ . We will consider the product measure space  $(I^2, \lambda^2) = (I \times I, \lambda \otimes \lambda)$ . The strategy of the proof is to transform the general problem into the simpler subproblem on the class of measure networks of the form  $(I^2, \omega, \lambda^2)$ , via a classical isomorphism theorem. Such a strategy was proposed by Gangbo in [23, Proposition A.3] to prove equality of Wasserstein and Monge distances in the classical optimal transport setting. However, Pratelli observes in [36, Section 1.2] that such an approach cannot work in general, due to the fact that the maps appearing in the isomorphism theorem may be discontinuous. We will show that this lack of continuity is no longer a problem in the GW/GM setting, leading to a relatively simpler proof than the classical optimal transport results [5, 36]. We will require a few lemmas.

**Lemma 3.2.** *Let  $\mathcal{X}$  be a measure network with non-atomic probability measure  $\mu_X$ . Then there exists a measure network  $\mathcal{I}_{\mathcal{X}} = (I^2, \omega_{I^2}^X, \lambda^2)$  which is strongly isomorphic to  $\mathcal{X}$ .*

*Proof.* It is a standard fact of measure theory that, since  $(X, \mu_X)$  is a Polish space with non-atomic probability measure, there exists a measure-preserving bijection  $\phi : I \rightarrow X$  (with respect to  $\lambda$  and  $\mu_X$ ) such that its inverse is also measure-preserving (i.e.,  $\phi$  is a *measure space isomorphism*) [37, Ch. 15, Theorem 16]. Likewise, we have a measure space isomorphism  $\psi : I \rightarrow I^2$ , with respect to  $\lambda$  and  $\lambda^2$ . Then  $\tau := \phi \circ \psi^{-1} : I^2 \rightarrow X$  is a measure space isomorphism with respect to  $\lambda^2$  and  $\mu_X$ . We then define

$$\omega_{I^2}^X(s, t) := \omega_X(\tau(s), \tau(t))$$

for  $s, t \in I$ . Then  $\tau$  is a strong isomorphism, by definition.  $\square$

**Remark 3.3.** *Following the proof of Lemma 3.2, one sees that we could have alternatively identified  $\mathcal{X}$  strongly isomorphically with a measure network of the form  $(I, \omega_I^X, \lambda)$ —indeed, this is a common method for parameterizing  $m$ -nets or  $mm$ -spaces (see [15, Section 2.5.1] or [40, Lemma 5.3]). The reason for modeling our spaces over the square  $(I^2, \lambda^2)$  is so that we can apply a technical lemma (Lemma 3.5), as is further explained in Remark 3.6.*

The next result then follows by Theorem 1 and Proposition 2.15 (in particular, the triangle inequalities and the identification of strongly isomorphic spaces under both  $\text{GW}_p$  and  $\text{GM}_p$ ).

**Lemma 3.4.** *Let  $\mathcal{X}$  and  $\mathcal{Y}$  be Polish measure networks with non-atomic probability measures, and let  $\mathcal{I}_{\mathcal{X}}$  and  $\mathcal{I}_{\mathcal{Y}}$  be as in Lemma 3.2. Then  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) = \text{GW}_p(\mathcal{I}_{\mathcal{X}}, \mathcal{I}_{\mathcal{Y}})$  and  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) = \text{GM}_p(\mathcal{I}_{\mathcal{X}}, \mathcal{I}_{\mathcal{Y}})$ .*

The following is a special case of [11, Theorem 1.1 (i)].

**Lemma 3.5** ([11]). *For any  $\pi \in \mathcal{C}(\lambda^2, \lambda^2)$ , there exists a sequence of measure-preserving maps  $\phi_m : I^2 \rightarrow I^2$  such that  $(\text{id}_{I^2} \times \phi_m) \# \lambda^2$  converges weakly to  $\pi$ .*

**Remark 3.6.** *In the original work of Brenier and Gangbo [11], the result was stated for approximation of couplings over the cube  $I^d$  for  $d \geq 2$ , as this is the general setting for other results in that paper. This is the reason that we have chosen to identify an arbitrary  $m$ -net  $\mathcal{X}$  with an  $m$ -net  $\mathcal{I}_{\mathcal{X}}$  over the square  $(I^2, \lambda^2)$  in Lemma 3.2, rather than over the interval  $(I, \lambda)$  (see Remark 3.3). In fact, checking the details of the proof of [11, Theorem 1.1] one sees that it applies also to the  $d = 1$  case (i.e., to  $(I, \lambda)$ ). However, in order to give a precise reference, we have opted to work with the  $d = 2$  case.*

*Proof of Theorem 2.* It suffices to consider the case of measure networks  $\mathcal{I} = (I^2, \omega, \lambda^2)$  and  $\mathcal{I}' = (I^2, \omega', \lambda^2)$  in  $\mathbb{N}_p$ . Indeed, if the result holds for measure networks over  $(I^2, \lambda^2)$ , then for general non-atomic measure networks  $\mathcal{X}$  and  $\mathcal{Y}$ , one has

$$\text{GW}_p(\mathcal{X}, \mathcal{Y}) = \text{GW}_p(\mathcal{I}_{\mathcal{X}}, \mathcal{I}_{\mathcal{Y}}) = \text{GM}_p(\mathcal{I}_{\mathcal{X}}, \mathcal{I}_{\mathcal{Y}}) = \text{GM}_p(\mathcal{X}, \mathcal{Y}),$$

by Lemma 3.4.

Let  $\pi$  be an optimal coupling for  $\mathcal{I}$  and  $\mathcal{I}'$  (see Remark 2.11). By Lemma 3.5, there is a sequence of measure-preserving maps  $\phi_m : I^2 \rightarrow I^2$  such that  $\pi_m := (\text{id}_{I^2} \times \phi_m) \# \xrightarrow{m \rightarrow \infty} \pi$ , in the weak topology. It is proved in [15, Lemma 2.3] that the distortion function  $\text{dis}_p$  is continuous in the weak topology, when it is defined with respect to measure networks in  $\mathbb{N}_\infty$ . The same proof goes through for measure networks in  $\mathbb{N}_p$ , so we have

$$\text{GM}_p(\mathcal{I}, \mathcal{I}') \leq \lim_{m \rightarrow \infty} \text{dis}_p(\pi_m) = \text{dis}_p(\pi) = \text{GW}_p(\mathcal{I}, \mathcal{I}').$$

By the general bound  $\text{GW}_p(\mathcal{I}, \mathcal{I}') \leq \text{GM}_p(\mathcal{I}, \mathcal{I}')$  (Remark 2.10), this shows  $\text{GW}_p(\mathcal{I}, \mathcal{I}') = \text{GM}_p(\mathcal{I}, \mathcal{I}')$  and completes the proof of the theorem.  $\square$

## 3.2 Discrete Spaces

We now consider the scenario which is the “opposite” of the non-atomic setting: finite spaces with discrete measures. For this subsection, let  $\mathcal{C}_{\text{fin}}$  denote the class of measure networks  $\mathcal{X}$  finite and  $\mu_X$  uniform; we put no additional restriction on network functions at this point. We begin with a simple structural result, characterizing infinite values and asymmetry for Gromov-Monge distances.

**Proposition 3.7.** *Let  $\mathcal{X}, \mathcal{Y} \in \mathcal{C}_{\text{fin}}$  with  $|X| = m$  and  $|Y| = n$ . Then  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) < \infty$  if and only if  $n$  divides  $m$ . In the case  $m = n$ ,  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) = \text{GM}_p(\mathcal{Y}, \mathcal{X}) < \infty$ .*

*Proof.* Suppose that  $n$  divides  $m$ , say  $m = kn$ . Then any function  $\phi : X \rightarrow Y$  whose fibers all have cardinality  $k$  is a measure-preserving map; in particular,  $\mathcal{T}(\mu_X, \mu_Y) \neq \emptyset$ , so  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) < \infty$ . Conversely, let  $\phi : X \rightarrow Y$  be a measure-preserving map. For any  $y \in Y$ , we have

$$\frac{1}{n} = \mu_Y(y) = \mu_X(\phi^{-1}(y)) = |\phi^{-1}(y)| \frac{1}{m},$$

hence  $m = |\phi^{-1}(y)|n$ .

If  $m = n$ , then the measure preserving maps are exactly the bijections  $\phi : X \rightarrow Y$ , and

$$\text{dis}_{p,\mathcal{X},\mathcal{Y}}(\phi) = \text{dis}_{p,\mathcal{Y},\mathcal{X}}(\phi^{-1}) < \infty.$$

□

In light of the previous result, it is natural to restrict our attention to the subclass  $\mathcal{C}_n$  consisting of  $\mathcal{X} \in \mathcal{C}_{\text{fin}}$  of fixed cardinality,  $|X| = n$ . As was stated in Section 2.3, Problem 2.17 has been solved by Vayer in the subclass of  $\mathcal{C}_n$  consisting of measure networks  $\mathcal{X}$  with  $X \subset \mathbb{R}^d$  (for some arbitrary dimension  $d$ ) and  $\omega_X(x, x') = \|x - x'\|^2$ : in this subclass, GW 2-distance is realized by a Monge map (i.e., a permutation) [42, Theorem 4.1.2]. The proof uses ideas of [29] and, in particular, relies on the observation that squared Euclidean distance matrices are conditionally negative definite. We now prove a similar result for the subclass of  $\mathcal{C}_n$  consisting of m-nets with symmetric positive definite network functions. That is, for  $\mathcal{X} \in \mathcal{C}_n$ , we pick an ordering  $(x_1, \dots, x_n)$  of  $X$  and consider the network function as a matrix  $\omega_X \in \mathbb{R}^{n \times n}$  by setting  $\omega_X(i, j) = \omega_X(x_i, x_j)$  (see Example 2.16). We say that  $\omega_X$  is **symmetric positive definite** if it is symmetric positive definite as a matrix.

**Proposition 3.8.** *For any  $n \in \mathbb{Z}_{>0}$ , consider  $\mathcal{X}, \mathcal{Y} \in \mathcal{C}_n$  such that their network functions are symmetric positive definite. Then  $\text{GW}_2(\mathcal{X}, \mathcal{Y})$  is realized by a measure-preserving map.*

As an example of a natural symmetric positive definite network function, consider a graph with vertex set  $X$ ,  $\mu_X$  uniform and  $\omega_X$  given by the heat kernel  $\exp(-tL)$  for some parameter  $t > 0$ , where  $L$  is the graph Laplacian matrix. The idea of the proof comes from [19, Theorem 2], which specifically considers graph heat kernels, and [4, Lemma 4.3], which offers a similar computation in the setting of cosine similarity matrices; neither of these results make the connection to Monge maps, although the former derives a bound on sparsity of optimal couplings.

*Proof.* As above, consider  $\omega_X$  and  $\omega_Y$  as matrices in  $\mathbb{R}^{n \times n}$ . Likewise, we consider probability measures as column vectors  $\mu_X, \mu_Y \in \mathbb{R}^n$  (specifically, they are both equal to the column vector with all entries equal to  $\frac{1}{n}$ ). Since it is symmetric positive definite,  $\omega_X$  admits a Cholesky decomposition  $\omega_X = U_X^T U_X$ . Let  $\omega_Y = V_Y^T V_Y$  be defined similarly.

A coupling of  $\mu_X$  and  $\mu_Y$  can be considered as a matrix  $\pi \in \mathbb{R}^{n \times n}$  whose row and column sums agree with  $\mu_X$  and  $\mu_Y$ , respectively. We can express the distortion of  $\pi$  as

$$\begin{aligned} \text{dis}_2(\pi)^2 &= \sum_{i,j,k,\ell} (\omega_X(i, k) - \omega_Y(j, \ell))^2 \pi(i, j) \pi(k, \ell) \\ &= \sum_{i,j,k,\ell} \omega_X(i, k)^2 \pi(i, j) \pi(k, \ell) + \sum_{i,j,k,\ell} \omega_Y(j, \ell)^2 \pi(i, j) \pi(k, \ell) \\ &\quad - 2 \sum_{i,j,k,\ell} \omega_X(i, k) \omega_Y(j, \ell) \pi(i, j) \pi(k, \ell) \\ &= \sum_{i,k} \omega_X(i, k)^2 \mu_X(i) \mu_X(k) + \sum_{j,\ell} \omega_Y(j, \ell)^2 \mu_Y(j) \mu_Y(\ell) \\ &\quad - 2 \sum_{i,j,k,\ell} \omega_X(i, k) \omega_Y(j, \ell) \pi(i, j) \pi(k, \ell). \end{aligned}$$

Since the first two terms do not depend on  $\pi$ , minimizing  $\text{dis}_2(\pi)^2$  is equivalent to maximizing the quantity  $\sum_{i,j,k,\ell} \omega_X(i, k) \omega_Y(j, \ell) \pi(i, j) \pi(k, \ell)$  over all couplings  $\pi$ . Let  $\langle \cdot, \cdot \rangle$  denote the Frobenius inner product on  $\mathbb{R}^{n \times n}$  and  $\|\cdot\|$  the associated norm. Then our object is to maximize

$$\begin{aligned} \sum_{i,j,k,\ell} \omega_X(i, k) \omega_Y(j, \ell) \pi(i, j) \pi(k, \ell) &= \langle \omega_X \pi, \pi \omega_Y \rangle = \langle U_X^T U_X \pi, \pi V_Y^T V_Y \rangle \\ &= \langle U_X \pi V_Y^T, U_X \pi V_Y^T \rangle = \|U_X \pi V_Y^T\|^2. \end{aligned}$$

The function  $\pi \mapsto \|U_X \pi V_Y^T\|^2$  is a convex function on  $\mathbb{R}^{n \times n}$ . Moreover, considered as a space of matrices, the set of couplings  $\mathcal{C}(\mu_X, \mu_Y)$  is given by

$$\mathcal{C}(\mu_X, \mu_Y) = \left\{ \pi \in \mathbb{R}^{n \times n} \mid \sum_i \pi(i, j) = \sum_j \pi(i, j) = \frac{1}{n} \text{ and } \pi(i, j) \geq 0 \ \forall i, j \right\};$$

that is, it is a rescaling of the set of bistochastic matrices by a factor of  $\frac{1}{n}$ . By Birkhoff's theorem,  $\mathcal{C}(\mu_X, \mu_Y)$  is the convex hull of the set of scaled (by  $\frac{1}{n}$ ) permutation matrices. These correspond to couplings which are induced by measure-preserving mappings  $X \rightarrow Y$ . By standard optimization theory, the convex function  $\pi \mapsto \|U_X \pi V_X^T\|$  is always maximized at an extremal point of its constraint polytope [7]. This proves the claim.  $\square$

### 3.3 The General Metric Measure Space Setting

We consider a variant of Gromov-Monge distance on the class of mm-spaces which allows for a certain notion of mass splitting.

**Definition 3.9.** Let  $(X, d_X)$  be a metric space, let  $Z$  be a set and let  $\phi : Z \rightarrow X$  be a function. The associated **pullback pseudometric** is the function  $\phi^* d_X : Z \times Z \rightarrow \mathbb{R}$  defined by

$$(\phi^* d_X)(z, z') = d_X(\phi(z), \phi(z')).$$

It is easy to check that  $\phi^* d_X$  indeed satisfies the axioms of a pseudometric.

**Definition 3.10.** Let  $\mathcal{X}$  be an mm-space. A **mass splitting** of  $\mathcal{X}$  is a measure network  $\mathcal{Z}$  such that there exists a measure-preserving map  $\rho : Z \rightarrow X$  with the property that  $\omega_Z = \rho^* d_X$ .

**Theorem 3.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be metric measure spaces. Then

$$\text{GW}_p(\mathcal{X}, \mathcal{Y}) = \inf_{\mathcal{Z}} \text{GM}_p(\mathcal{Z}, \mathcal{Y}),$$

where the infimum is taken over mass-splittings of  $\mathcal{X}$ .

Similar ideas of splitting mass to realize Gromov-Wasserstein distance through Monge maps were employed in [18] to construct explicit geodesics between networks in a shape analysis setting. The algorithm used in [18] is an implementation of the geodesic characterization first obtained in [40, Theorem 3.1]. The result is also reminiscent of [16, Proposition 2.7.5], which considers a similar characterization of a Gromov-Hausdorff-like distance on the space of networks (without measures).

**Example 3.11.** To illustrate the idea of the theorem with a simple example, consider the space  $\mathcal{X}$  consisting of a single point and the space  $\mathcal{Y}$  consisting of two points  $Y = \{y_1, y_2\}$  with  $d_Y(y_1, y_2) = 1$  and with uniform weights. The Gromov-Wasserstein  $p$ -distance between the spaces is realized by the unique coupling  $\pi$  satisfying  $\pi((x, y_j)) = \frac{1}{2}$  so that  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) = \frac{1}{2^{1/p}}$ , while  $\text{GM}_p(\mathcal{X}, \mathcal{Y}) = \infty$  since  $\mathcal{T}(\mu_X, \mu_Y) = \emptyset$ . On the other hand, consider the mass splitting  $\mathcal{Z}$  with  $Z = \{z_1, z_2\}$ ,  $\omega_Z(z_1, z_2) = 0$  and with uniform weights. It is easy to check that  $\mathcal{Z}$  is a mass-splitting of  $\mathcal{X}$  with  $\rho(z_j) = x$ . Moreover,  $\text{GM}_p(\mathcal{Z}, \mathcal{Y}) = \frac{1}{2^{1/p}}$  and this quantity is realized by the measure-preserving map  $\phi(z_j) = y_j$ . This general setup is illustrated by the following diagram:

$$\begin{array}{ccc} \mathcal{Z} & & \\ \downarrow \rho & \searrow \phi & \\ \mathcal{X} & & \mathcal{Y} \end{array}$$

*Proof of Theorem 3.* Given a measure coupling  $\pi$  of  $\mu_X$  and  $\mu_Y$ , define a mass splitting  $\mathcal{Z}$  by setting  $Z = \text{supp}(\pi) \subset X \times Y$ ,  $\mu_Z = \pi$ ,  $\rho = \rho_X|_Z$  (with  $\rho_X : X \times Y \rightarrow X$  denoting projection onto the first coordinate) and  $\omega_Z = \rho^* d_X$ . Then  $\phi = \rho_Y|_Z : Z \rightarrow Y$  (with  $\rho_Y : X \times Y \rightarrow Y$  the projection map) is a measure-preserving map with

$$\begin{aligned} \text{dis}_{p, \mathcal{Z}, \mathcal{Y}}(\phi)^p &= \iint_{Z \times Z} |\omega_Z(z, z') - d_Y(\phi(z), \phi(z'))|^p \mu_Z(dz) \mu_Z(dz') \\ &= \iint_{(X \times Y)^2} |d_X(x, x') - d_Y(y, y')|^p \pi(dx \times dy) \pi(dx' \times dy') = \text{dis}_{p, \mathcal{X}, \mathcal{Y}}(\pi)^p. \end{aligned}$$

We conclude that

$$\text{GW}_p(\mathcal{X}, \mathcal{Y}) \geq \inf_{\mathcal{Z}} \text{GM}_p(\mathcal{Z}, \mathcal{Y}).$$

Conversely, let  $\mathcal{Z} = (Z, \omega_Z, \mu_Z)$  be a mass splitting of  $\mathcal{X}$ , with measure-preserving map  $\rho : Z \rightarrow X$  such that  $\omega_Z = \rho^* d_X$ , and let  $\phi : Z \rightarrow Y$  be a measure-preserving map (we assume that one exists; otherwise, the desired inequality  $\text{GW}_p(\mathcal{X}, \mathcal{Y}) \leq \text{GM}_p(\mathcal{Z}, \mathcal{Y})$  follows trivially). We define a probability measure  $\pi$  on  $X \times Y$  as  $\pi = (\rho \times \phi)_\# \mu_Z$ . Then  $\pi$  defines a measure coupling of  $\mu_X$  and  $\mu_Y$ . Indeed,

$$(\rho_X)_\# \pi = (\rho_X)_\# (\rho \times \phi)_\# \mu_Z = (\rho_X \circ (\rho \times \phi))_\# \mu_Z = \rho_\# \mu_Z = \mu_X,$$

and, by a similar argument, we also have  $(\rho_Y)_\# \pi = \mu_Y$ . Consider the following calculation (we suppress all function arguments to condense notation):

$$\begin{aligned} \iint_{(X \times Y)^2} |d_X - d_Y|^p \pi \otimes \pi &= \iint_{(X \times Y)^2} |d_X - d_Y|^p (\rho \times \phi)_\# \mu_Z \otimes (\rho \times \phi)_\# \mu_Z \\ &= \iint_{(X \times Y)^2} |d_X - d_Y|^p ((\rho \times \phi) \times (\rho \times \phi))_\# \mu_Z \otimes \mu_Z \end{aligned} \quad (1)$$

$$= \iint_{Z \times Z} |d_X \circ (\rho \times \rho) - d_Y \circ (\phi \times \phi)|^p \mu_Z \otimes \mu_Z \quad (2)$$

$$= \iint_{Z \times Z} |\omega_Z - d_Y \circ (\phi \times \phi)|^p \mu_Z \otimes \mu_Z. \quad (3)$$

Equality (1) 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  $(X \times Y)^2$  and (2) and (3) both follow from the change-of-variables formula, with the latter also using  $d_X \circ (\rho \times \rho) = \omega_Z$ . This calculation shows that

$$\text{GW}_p(\mathcal{X}, \mathcal{Y}) \leq \inf_{\mathcal{Z}} \text{GM}_p(\mathcal{Z}, \mathcal{Y})$$

and the proof is therefore complete.  $\square$

## 4 An Embedding Formulation of Gromov-Monge Distance

In this section, we restrict our attention to mm-spaces (rather than general m-nets) and study alternative notions of distance between them.

### 4.1 Gromov-Monge Distances from Joint Embeddings

In [39], Sturm defined an alternative metric for comparing mm-spaces, which is analogous to the embedding formulation of Gromov-Hausdorff distance. For mm-spaces  $\mathcal{X}$  and  $\mathcal{Y}$ , let

$$\text{GW}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) := \inf_{Z, \phi_X, \phi_Y} W_p^Z((\phi_X)_\# \mu_X, (\phi_Y)_\# \mu_Y),$$

where the infimum is taken over all isometric embeddings  $\phi_X : X \rightarrow Z$  and  $\phi_Y : Y \rightarrow Z$  into some metric space  $Z$  and  $W_p^Z$  is the usual Wasserstein  $p$ -distance between probability measures on  $Z$ . We analogously define a new extended quasi-metric  $\text{GM}_p^{\text{em}}$  on the space of mm-spaces by

$$\text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) := \inf_{Z, \phi_X, \phi_Y} M_p^Z((\phi_X)_\# \mu_X, (\phi_Y)_\# \mu_Y),$$

where the infimum is once again taken over isometric embeddings into a common ambient metric space.

**Theorem 4.** *Let  $\mathcal{X}$  and  $\mathcal{Y}$  be mm-spaces with non-atomic probability measures. Then for  $p \in [1, \infty)$ ,  $\text{GW}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) = \text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y})$ .*

*Proof.* Clearly,  $\text{GW}_p^{\text{em}} \leq \text{GM}_p^{\text{em}}$  holds in general; in particular, if  $\text{GW}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) = \infty$  (which is possible without any boundedness assumption on the mm-spaces) then  $\text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) = \infty$  as well. Assume, then, that  $\text{GW}_p^{\text{em}}(\mathcal{X}, \mathcal{Y})$  is finite, and let  $\phi_X : X \rightarrow Z$  and  $\phi_Y : Y \rightarrow Z$  be isometric embeddings into some metric space  $(Z, d_Z)$ . It follows from [39, Lemma 3.3 (ii)] that, in the context of computing  $\text{GW}_p^{\text{em}}$ , we may assume without loss of generality that  $(Z, d_Z)$  is complete and separable—that is, we can restrict the definition to the infimum of embeddings into such spaces without changing the value of the metric. The same proof applies to show that this can be assumed without loss of generality in the context of computing  $\text{GM}_p^{\text{em}}$ . Since the push-forwards  $(\phi_X)_\# \mu_X$  and  $(\phi_Y)_\# \mu_Y$  are nonatomic (this follows by injectivity and the assumption that the original measures were nonatomic) and the function  $Z \times Z \rightarrow \mathbb{R}$  defined by  $(z, z') \mapsto d_Z(z, z')^p$  is continuous, Pratelli's result [36, Theorem B] implies that  $W_p^Z((\phi_X)_\# \mu_X, (\phi_Y)_\# \mu_Y) = M_p^Z((\phi_X)_\# \mu_X, (\phi_Y)_\# \mu_Y)$ . Because this equality holds for every choice of  $Z, \phi_X, \phi_Y$ , passing to the infimum over such embeddings yields  $\text{GW}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) = \text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y})$ .  $\square$

## 4.2 Comparing Gromov-Monge Distances

Here, we show that the two notions of Gromov-Monge distance considered in this paper are comparable for arbitrary mm-spaces, but are not equivalent in general. The following is a Gromov-Monge version of [32, Theorem 5.1(g)].

**Theorem 5.** *For mm-spaces  $\mathcal{X}$  and  $\mathcal{Y}$  and for any  $p \in [1, \infty]$ ,*

$$\frac{1}{2} \text{GM}_p(\mathcal{X}, \mathcal{Y}) \leq \text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}). \quad (4)$$

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

$$\frac{1}{2} \text{GM}_\infty(\mathcal{X}, \mathcal{Y}) = \text{GM}_\infty^{\text{em}}(\mathcal{X}, \mathcal{Y}). \quad (5)$$

*Proof.* If  $\text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) = \infty$  then (4) is trivially satisfied, so let us assume that  $\text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) < \infty$ . To prove that (4) holds under this assumption, we will show that whenever  $\text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) < r$ , there exists  $\phi \in \mathcal{T}(\mu_X, \mu_Y)$  such that  $\|d_X - d_Y\|_{L^p(\pi_\phi \otimes \pi_\phi)} \leq 2r$ . If  $\text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) < r$  then we can find isometries  $\phi_X : X \rightarrow Z$  and  $\phi_Y : Y \rightarrow Z$  into a metric space  $Z$  such that  $\text{M}_p^Z((\phi_X)_\# \mu_X, (\phi_Y)_\# \mu_Y) < r$ . 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  $\text{M}_p^Z$ , there exists  $\phi \in \mathcal{T}(\mu_X, \mu_Y)$  such that  $\|d_Z\|_{L^p(\pi_\phi)} < r$ . 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

$$\|d_X - d_Y\|_{L^p(\pi_\phi \otimes \pi_\phi)} \leq 2\|d_Z\|_{L^p(\pi_\phi)} < 2r.$$

This establishes (4).

Now we wish to show that  $\frac{1}{2} \text{GM}_\infty(\mathcal{X}, \mathcal{Y}) \geq \text{GM}_\infty^{\text{em}}(\mathcal{X}, \mathcal{Y})$ . If  $\text{GM}_\infty(\mathcal{X}, \mathcal{Y}) = \infty$  then we are done, so assume not. Let  $\phi_0 : X \rightarrow Y$  be any measure preserving map with

$$\|d_X - d_Y\|_{L^\infty(\text{supp}(\pi_{\phi_0} \otimes \pi_{\phi_0}))} = \sup_{x, x' \in X} |d_X(x, x') - d_Y(\phi_0(x), \phi_0(x'))| = 2r.$$

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  $\text{M}_\infty^Z((\phi_X)_\# \mu_X, (\phi_Y)_\# \mu_Y) \leq r$ . 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

$$\begin{aligned} \text{M}_\infty^Z(\mu_X, \mu_Y) &= \inf_{\phi \in \mathcal{T}(\mu_X, \mu_Y)} \sup_{x \in X} d_Z(x, \phi(x)) \leq \sup_{x \in X} d_Z(x, \phi_0(x)) \\ &= \sup_{x \in X} \inf_{x' \in X} \{d_X(x, x') + r + d_Y(\phi_0(x'), \phi_0(x))\} = r. \end{aligned}$$

This completes the proof.  $\square$

**Example 4.1** ( $\text{GM}_p$  and  $\text{GM}_p^{\text{em}}$  are Not BiLipschitz Equivalent). *Consider the family of mm-spaces  $\Delta_n$ . Each  $\Delta_n$  consists of the space  $X_n = \{1, \dots, n\}$  with metric  $d_n(i, j) = 1 - \delta_{ij}$  and measure  $\nu_n$  defined by  $\nu_n(i) = 1/n$ . For  $p < \infty$ ,*

$$\text{GM}_p^{\text{em}}(\Delta_{2n}, \Delta_n) \geq \text{GW}_p^{\text{em}}(\Delta_{2n}, \Delta_n) \geq \frac{1}{4},$$

where the former bound holds generally and the latter is shown in [32, Claim 5.3]. On the other hand, for  $p < \infty$ ,  $\text{GM}_p(\Delta_{2n}, \Delta_n) = 1/(2n)^{1/p}$ . This follows from the fact that  $\mathcal{T}(\nu_{2n}, \nu_n)$  is simply the set of 2-to-1 maps from  $\{1, \dots, 2n\}$  to  $\{1, \dots, n\}$  and all such maps have the same Monge cost, so that the quantity is obtained by a direct calculation. It follows that  $\text{GM}_p$  and  $\text{GM}_p^{\text{em}}$  are not bi-Lipschitz equivalent for any  $p \in [1, \infty)$ .

**Example 4.2** (Tightness of the Inequality). *We now show, by example, that the factor of  $\frac{1}{2}$  in (4) cannot be improved. The analysis provided here is similar to that of [32, Remark 5.14].*

Using the notation of Example 4.1, consider the spaces  $\Delta_n$  and  $\Delta_1 \approx *$  (the latter is the one point mm-space). The distance  $\text{GM}_p(\Delta_n, *)$  is the  $p$ -size of  $\Delta_n$  (see Remark 2.13), which is given explicitly by

$$\text{GM}_p(\Delta_n, *) = \left( \sum_{i,j=1}^n d_n(i,j)^p \cdot \frac{1}{n^2} \right)^{1/p} = \left( \sum_{i \neq j} 1^p \cdot \frac{1}{n^2} \right)^{1/p} = \left( n(n-1) \cdot \frac{1}{n^2} \right)^{1/p} = \left( 1 - \frac{1}{n} \right)^{1/p}$$

(this agrees with the more general formula given in [32, Remark 5.17]). On the other hand, we claim that  $\text{GM}_p^{\text{em}}(\Delta_n, *) = \frac{1}{2}$  for all  $n$ . We then have that, for any  $c < 2$ , one can take  $n$  sufficiently large so that  $\text{GM}_p(\Delta_n, *) > c \cdot \text{GM}_p^{\text{em}}(\Delta_n, *)$ , which shows that the inequality (4) is tight.

It remains to derive the value of  $\text{GM}_p^{\text{em}}(\Delta_n, *)$ . In this setting, any isometric embeddings of  $\Delta_n$  and  $*$  into a common metric space amount to choosing distances  $\alpha_i$  between  $i \in \Delta_n$  and the single point  $x$ . Given such a choice  $\vec{\alpha} = (\alpha_1, \dots, \alpha_n)$ , the Monge distance between the pushforward measures is given by

$$\left( \sum_{i=1}^n \alpha_i^p \cdot \frac{1}{n} \right)^{1/p} = \frac{1}{n^{1/p}} \|\vec{\alpha}\|_p,$$

where  $\|\cdot\|_p$  is the standard  $\ell_p$ -norm on  $\mathbb{R}^n$ . It follows that

$$\text{GM}_p^{\text{em}}(\Delta_n, *) = \inf_{\vec{\alpha}} \frac{1}{n^{1/p}} \|\vec{\alpha}\|_p,$$

where the entries of  $\vec{\alpha}$  are subject to constraints which guarantee a valid metric. Namely, we must have  $\alpha_i > 0$  for all  $i$ , as well as  $\alpha_i - \alpha_j \leq 1 = d_n(i, j) \leq \alpha_i + \alpha_j$  for all  $i \neq j$ . In particular, the constraints of the form  $1 \leq \alpha_i + \alpha_j$  with  $i < j$  imply that

$$\binom{n}{2} = \sum_{i < j} 1 \leq \sum_{i < j} (\alpha_i + \alpha_j) = (n-1) \sum_i \alpha_i = (n-1) \|\vec{\alpha}\|_1,$$

whence we obtain, for all  $p \geq 1$ ,

$$\frac{1}{2} \leq \frac{1}{n} \|\vec{\alpha}\|_1 \leq \frac{1}{n^{1/p}} \|\vec{\alpha}\|_p,$$

where the first inequality follows from the discussion above and the second is the generalized means inequality. Therefore,  $\text{GM}_p^{\text{em}}(\Delta_n, *) \geq \frac{1}{2}$  holds for all  $p \geq 1$ . On the other hand, the constant vector  $\vec{\alpha} = (1/2, \dots, 1/2)$  gives a valid metric and realizes this lower bound, so that  $\text{GM}_p^{\text{em}}(\Delta_n, *) = \frac{1}{2}$ .

**Remark 4.3.** Example 4.2 also shows that the inequality  $\frac{1}{2} \text{GW}_p(\mathcal{X}, \mathcal{Y}) \leq \text{GW}_p^{\text{em}}(\mathcal{X}, \mathcal{Y})$  from [32, Theorem 5.1(g)] (note the slightly different normalization convention used there) is tight, since it is not hard to see that  $\text{GW}_p(\Delta_n, *) = \text{GM}_p(\Delta_n, *)$  and  $\text{GW}_p^{\text{em}}(\Delta_n, *) = \text{GM}_p^{\text{em}}(\Delta_n, *)$ . It was previously only shown in [32, Remark 5.14] that this is not an equality, in general.

### 4.3 Gromov-Monge Distances for Euclidean Spaces

A **Euclidean mm-space** is an mm-space  $\mathcal{X}$  such that  $X \subset \mathbb{R}^n$  for some  $n$  and  $d_X$  is the restriction of Euclidean distance (and  $\mu_X$  is a Borel measure with respect to the subspace topology). For Euclidean mm-spaces  $\mathcal{X}$  and  $\mathcal{Y}$  with the same ambient space  $\mathbb{R}^n$ , we can consider the **isometry-invariant Monge  $p$ -distance**

$$\text{M}_p^{\mathbb{R}^n, \text{iso}}(\mathcal{X}, \mathcal{Y}) = \inf_{T \in E(n)} \left( \inf_{\phi \in \mathcal{T}(\mu_X, \mu_Y)} \int_X \|T(x) - \phi(x)\|^p \mu_X(dx) \right)^{1/p},$$

where we use  $E(n)$  to denote the group of Euclidean isometries of  $\mathbb{R}^n$  and  $\|\cdot\|$  for the Euclidean norm. Fixing  $\mathbb{R}^n$  as the ambient Euclidean space, unless specified otherwise, we use the cleaner notation  $\text{M}_p^{\text{iso}} = \text{M}_p^{\mathbb{R}^n, \text{iso}}$ .

**Theorem 6.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  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

$$\text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) \leq \text{M}_p^{\text{iso}}(\mathcal{X}, \mathcal{Y}) \leq M^{3/4} \cdot c_n \cdot (\text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}))^{1/4},$$

where  $c_n$  is a constant depending only on  $n$ .

The theorem is analogous to [31, Theorem 4], which compares  $\text{GW}_p^{\text{em}}$  and isometry-invariant Wasserstein distance for Euclidean spaces. The proof of the theorem is essentially the same, with the main difference being the use of the following technical lemma, which is a variation of [31, Lemma 2], specialized to the Monge setting. Recall that a **correspondence** between sets  $X$  and  $Y$  is a subset  $R \subset X \times Y$  such that the coordinate projections satisfy  $\rho_X(R) = X$  and  $\rho_Y(R) = Y$ . We use  $\mathcal{R}(X, Y)$  to denote the set of correspondences between  $X$  and  $Y$ . Observe that the theorem is only interesting if the quantity  $M$  is finite, so we state the following result with this assumption.

**Lemma 4.4.** *Let  $\mathcal{X}$  and  $\mathcal{Y}$  be bounded mm-spaces and let  $M = \max\{\text{diam}(X), \text{diam}(Y)\}$ . If*

$$\text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) \leq \epsilon \cdot M$$

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  $\phi : X \rightarrow Y$  satisfying  $\phi(X_\epsilon) = Y_\epsilon$  such that

$$\min\{\mu_X(X_\epsilon), \mu_Y(Y_\epsilon), \pi_\phi(R_\epsilon)\} \geq 1 - \epsilon^{p/2} \quad (6)$$

and

$$\sup_{(x,y),(x',y') \in R_\epsilon} |d_X(x, x') - d_Y(y, y')| \leq \epsilon^{1/2} M. \quad (7)$$

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

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

We then define  $X_\epsilon = \rho_X(R_\epsilon)$  and  $Y_\epsilon = \rho_Y(R_\epsilon)$ , 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 (7). 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

$$\pi_\phi(R_\epsilon) = (\text{id}_X \times \phi)_\# \mu_X(R_\epsilon) = \mu_X((\text{id}_X \times \phi)^{-1}(R_\epsilon)) = \mu_X(X_\epsilon),$$

and (6) is satisfied.  $\square$

*Proof of Theorem 6.* The inequality on the left is obvious, so let us consider the inequality on the right, under the assumption that  $\mathcal{X}$  and  $\mathcal{Y}$  are bounded (as this is the only interesting case). If either quantity is infinite, then both are, so assume not. It is easy to show that

$$\text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) \leq M = \max\{\text{diam}(X), \text{diam}(Y)\}.$$

We therefore suppose, without loss of generality, that  $\text{GM}_p^{\text{em}}(\mathcal{X}, \mathcal{Y}) = \epsilon M$  for some  $\epsilon \leq 1$ . Let  $X_\epsilon, Y_\epsilon, \phi$  and  $R_\epsilon$  be as in Lemma 4.4 and let  $R_\epsilon^c = (X \times Y) \setminus R_\epsilon$ . The condition (7) on  $R_\epsilon$  implies that the Gromov-Hausdorff distance between  $X_\epsilon$  and  $Y_\epsilon$  is bounded above by  $\epsilon^{1/2} M/2$  (see [13, 7.3.25]). In turn, [13, Corollary 7.3.28] says that this implies that there exists a  $(\epsilon^{1/2} M)$ -isometry  $\psi : X_\epsilon \rightarrow Y_\epsilon$ ; that is,  $\psi$  satisfies

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

and that  $\psi(X_\epsilon)$  is a  $(\epsilon^{1/2} M)$ -net for  $Y_\epsilon$  (i.e., for every  $y \in Y_\epsilon$  there is an  $x \in X_\epsilon$  such that  $d_Y(\psi(x), y) \leq \epsilon^{1/2} M$ ). We apply [2, Theorem 2.2] to conclude that there is an isometry  $T \in E(n)$  such that  $\sup_{x \in X_\epsilon} \|T(x) - \psi(x)\| \leq \epsilon^{1/4} \cdot a_n \cdot M$ , 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  $a, b \geq 0$  and  $p \geq 1$ ), we have

$$M_p^{\text{iso}}(\mathcal{X}, \mathcal{Y})^p \leq \int_{R_\epsilon^c} \|T(x) - y\|^p \pi_\phi(dx \times dy) \quad (8)$$

$$+ 2^{p-1} \int_{R_\epsilon} \|T(x) - \psi(x)\|^p \pi_\phi(dx \times dy) \quad (9)$$

$$+ 2^{p-1} \int_{R_\epsilon} \|\psi(x) - y\|^p \pi_\phi(dx \times dy). \quad (10)$$

We bound each term separately. First note that

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

where we have used isometry invariance of  $M_p^{\text{iso}}$  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

$$(8) \leq (2M)^p \pi_\phi(R_\epsilon^c) \leq 2^p \cdot M^p \cdot \epsilon^{p/2}.$$

The bounds on (9) and (10) follow from our assumptions on  $\psi$ :

$$(9) \leq 2^{p-1} \cdot \sup_{x \in X_\epsilon} \|T(x) - \psi(x)\|^p \cdot \pi_\phi(R_\epsilon) \leq 2^{p-1} \cdot \epsilon^{p/4} \cdot a_n^p \cdot M^p$$

and

$$(10) \leq 2^{p-1} \cdot \max_{(x,y) \in R_\epsilon} \|\psi(x) - y\|^p \pi_\phi(R_\epsilon) \leq 2^{p-1} \cdot \epsilon^{p/2} \cdot M^p,$$

where we have used that  $\psi(X_\epsilon)$  is a  $(\epsilon^{1/2}M)$ -net. Combining these estimates, we conclude

$$M_p^{\text{iso}}(\mathcal{X}, \mathcal{Y})^p \leq 2^p M^p (\epsilon^{p/2} + a_n^p \epsilon^{p/4} / 2 + \epsilon^{p/2} / 2) \leq M^p \epsilon^{p/4} \cdot 2^p (3/2 + a_n^p / 2).$$

Taking  $c_n = 3 + 2a_n \geq 2 \cdot (3/2 + a_n^p / 2)^{1/p}$  (to get a constant which only depends on  $n$ ), we have

$$M_p^{\text{iso}}(\mathcal{X}, \mathcal{Y}) \leq c_n \cdot M \cdot \epsilon^{1/4} = c_n \cdot M^{3/4} \cdot GM_p^{\text{em}}(\mathcal{X}, \mathcal{Y})^{1/4}.$$

□

**Remark 4.5.** We now explain connections between embedded GM distances and some metrics which have previously appeared in the literature.

The **continuous Procrustes distance** between embedded surfaces in  $\mathbb{R}^3$  was introduced in [9], where it was used to classify anatomical surfaces (in particular, shapes of primate teeth). The effectiveness at classification of this metric was shown to be roughly on par with that of a trained morphologist. The idea of the continuous Procrustes distance is to compare surfaces by simultaneously registering over rigid motions while looking for optimal measure-preserving maps between them. Theoretical aspects of this metric are studied in [1], where it is shown that optimal mappings are close to being conformal. The continuous Procrustes distance can be viewed as  $M_2^{\mathbb{R}^3, \text{iso}}$ , under the additional constraint that measure-preserving maps must also be continuous. A similar constraint on admissible maps can be made in the  $GM_p^{\text{em}}$ -distance. Theorem 6 (or the corresponding version under an additional constraint) gives an equivalence between embedded GM distance and continuous Procrustes distance.

A metric similar to the continuous Procrustes metric is studied in [24] for applications to 2D image registration. In our terminology, the metric of [24] is  $M_2^{\mathbb{R}^2, \text{iso}}$ , under the additional constraint that measure-preserving maps are smooth diffeomorphisms. Once again, (a slight variant) of Theorem 6 shows the equivalence of this metric with a restricted version of embedded GM distance.

**Remark 4.6.** A natural question is whether the inequalities in Theorem 6 are tight; in particular, can the exponent on the right hand side be improved? There has been some progress on the corresponding problem comparing isometry-invariant Hausdorff distance with Gromov-Hausdorff distance between Euclidean sets—it is shown in [31, Theorem 2] that the isometry-invariant Hausdorff distance between compact subsets  $X, Y \subset \mathbb{R}^n$  is bounded above by  $c'_n \cdot \max\{\text{diam}(X), \text{diam}(Y)\}^{1/2} \cdot \text{GH}(X, Y)^{1/2}$ , where  $\text{GH}$  denotes Gromov-Hausdorff distance and  $c'_n$  is positive a constant depending only on  $n$ . An example constructed in [31] shows that the exponent  $\frac{1}{2}$  on the Gromov-Hausdorff distance cannot be improved (namely, it cannot be made larger). More recently, it has been shown that for finite subsets of the real line  $X, Y \subset \mathbb{R}$ , isometry-invariant Hausdorff distance is bounded above by  $\frac{5}{4}\text{GH}(X, Y)$  [28, Theorem 3.2], and that this bound is tight [28, Theorem 3.10]. Optimality of the bounds in the Gromov-Wasserstein or Gromov-Monge settings has seen less progress, and we leave these as open questions.

## Acknowledgements

We acknowledge funding from these NSF projects: DMS 2107808, DMS 2324962, RI 1901360, CCF 1740761, CCF 1526513, and DMS 1723003 and also from BSF project 2020124.

## References

- [1] 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.
- [2] P Alestalo, DA Trotsenko, and J Väisälä. Isometric approximation. *Israel Journal of Mathematics*, 125(1):61–82, 2001.
- [3] David Alvarez-Melis and Tommi Jaakkola. Gromov-Wasserstein alignment of word embedding spaces. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1881–1890, 2018.
- [4] David Alvarez-Melis, Stefanie Jegelka, and Tommi S Jaakkola. Towards optimal transport with global invariances. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1870–1879. PMLR, 2019.
- [5] Luigi Ambrosio. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces*, pages 1–52. Springer, 2003.
- [6] Robert Beinert, Cosmas Heiss, and Gabriele Steidl. On assignment problems related to Gromov-Wasserstein distances on the real line. *arXiv preprint arXiv:2205.09006*, 2022.
- [7] Harold P. Benson. Concave programming. In *Encyclopedia of Optimization*, pages 315–319. Springer, 2001.
- [8] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and Radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.
- [9] Doug M Boyer, Yaron Lipman, Elizabeth St Clair, Jesus Puente, Biren A Patel, Thomas Funkhouser, Jukka Jernvall, and Ingrid Daubechies. Algorithms to automatically quantify the geometric similarity of anatomical surfaces. *Proceedings of the National Academy of Sciences*, 2011.
- [10] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.
- [11] Yann Brenier and Wilfrid Gangbo.  $l^p$  approximation of maps by diffeomorphisms. *Calculus of Variations and Partial Differential Equations*, 16(2):147–164, 2003.
- [12] Charlotte Bunne, David Alvarez-Melis, Andreas Krause, and Stefanie Jegelka. Learning generative models across incomparable spaces. In *International Conference on Machine Learning*, pages 851–861, 2019.
- [13] Dmitri Burago, Yuri Dmitrievich Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33. American Mathematical Soc., 2001.
- [14] Samantha Chen, Sunhyuk Lim, Facundo Mémoli, Zhengchao Wan, and Yusu Wang. Weisfeiler-Lehman meets Gromov-Wasserstein. In *International Conference on Machine Learning*, pages 3371–3416. PMLR, 2022.
- [15] Samir Chowdhury and Facundo Mémoli. The Gromov–Wasserstein distance between networks and stable network invariants. *Information and Inference: A Journal of the IMA*, 8(4):757–787, 2019.
- [16] Samir Chowdhury and Facundo Mémoli. Distances and isomorphism between networks: stability and convergence of network invariants. *Journal of Applied and Computational Topology*, pages 1–119, 2022.
- [17] Samir Chowdhury, David Miller, and Tom Needham. Quantized Gromov-Wasserstein. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 811–827. Springer, 2021.
- [18] Samir Chowdhury and Tom Needham. Gromov-Wasserstein averaging in a Riemannian framework. *arXiv preprint arXiv:1910.04308*, 2019.

- [19] Samir Chowdhury and Tom Needham. Generalized spectral clustering via Gromov-Wasserstein learning. In *International Conference on Artificial Intelligence and Statistics*, pages 712–720. PMLR, 2021.
- [20] Samir Chowdhury, Tom Needham, Ethan Semrad, Bei Wang, and Youjia Zhou. Hypergraph co-optimal transport: Metric and categorical properties. *Journal of Applied and Computational Topology*, pages 1–60, 2023.
- [21] Pinar Demetci, Rebecca Santorella, Björn Sandstede, William Stafford Noble, and Ritambhara Singh. Scot: Single-cell multi-omics alignment with optimal transport. *Journal of Computational Biology*, 29(1):3–18, 2022.
- [22] Théo Dumont, Théo Lacombe, and François-Xavier Vialard. On the existence of Monge maps for the Gromov–Wasserstein problem. *Foundations of Computational Mathematics*, pages 1–48, 2024.
- [23] Wilfrid Gangbo. The Monge mass transfer problem and its applications. *Contemporary Mathematics*, 226:79–104, 1999.
- [24] 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.
- [25] YoonHaeng Hur, Wenzuan Guo, and Tengyuan Liang. Reversible Gromov-Monge sampler for simulation-based inference. *arXiv preprint arXiv:2109.14090v3*, 2022.
- [26] YoonHaeng Hur, Wenzuan Guo, and Tengyuan Liang. Reversible Gromov–Monge sampler for simulation-based inference. *SIAM Journal on Mathematics of Data Science*, 6(2):283–310, 2024.
- [27] F William Lawvere. Metric spaces, generalized logic, and closed categories. *Rendiconti del seminario matematico e fisico di Milano*, 43(1):135–166, 1973.
- [28] Sushovan Majhi, Jeffrey Vitter, and Carola Wenk. Approximating Gromov-Hausdorff distance in euclidean space. *Computational Geometry*, 116:102034, 2024.
- [29] Haggai Maron and Yaron Lipman. (Probably) concave graph matching. *Advances in Neural Information Processing Systems*, 31, 2018.
- [30] Facundo Mémoli. On the use of Gromov-Hausdorff distances for shape comparison. In *Proceedings of Point Based Graphics*, 2007.
- [31] 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.
- [32] Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011.
- [33] Facundo Mémoli and Tom Needham. Distance distributions and inverse problems for metric measure spaces. *Studies in Applied Mathematics*, 149(4):943–1001, 2022.
- [34] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
- [35] Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In *International Conference on Machine Learning*, pages 2664–2672, 2016.
- [36] Aldo Pratelli. On the equality between Monge’s infimum and Kantorovich’s minimum in optimal mass transportation. In *Annales de l’Institut Henri Poincaré (B) Probability and Statistics*, volume 43, pages 1–13. Elsevier, 2007.
- [37] Halsey Lawrence Royden. *Real analysis*, volume 3. Macmillan New York, 1968.
- [38] Antoine Salmona, Julie Delon, and Agnès Desolneux. Gromov-Wasserstein distances between gaussian distributions. *Journal of Applied Probability*, 59(4), 2022.
- [39] Karl-Theodor Sturm. On the geometry of metric measure spaces. *Acta mathematica*, 196(1):65–131, 2006.
- [40] Karl-Theodor Sturm. *The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces*, volume 290. Memoirs of the American Mathematical Society, 2023.

- [41] Vayer Titouan, Ievgen Redko, Rémi Flamary, and Nicolas Courty. Co-optimal transport. *Advances in Neural Information Processing Systems*, 33:17559–17570, 2020.
- [42] Titouan Vayer. A contribution to optimal transport on incomparable spaces. *arXiv preprint arXiv:2011.04447*, 2020.
- [43] Titouan Vayer, Rémi Flamary, Romain Tavenard, Laetitia Chapel, and Nicolas Courty. Sliced Gromov-Wasserstein. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.
- [44] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
- [45] Hongteng Xu, Dixin Luo, and Lawrence Carin. Scalable Gromov-Wasserstein learning for graph partitioning and matching. *Advances in neural information processing systems*, 32, 2019.
- [46] Hongteng Xu, Dixin Luo, Hongyuan Zha, and Lawrence Carin Duke. Gromov-Wasserstein learning for graph matching and node embedding. In *International Conference on Machine Learning*, pages 6932–6941, 2019.
- [47] Zhengxin Zhang, Youssef Mroueh, Ziv Goldfeld, and Bharath Sriperumbudur. Cycle consistent probability divergences across different spaces. In *International Conference on Artificial Intelligence and Statistics*, pages 7257–7285. PMLR, 2022.