



The Gromov–Wasserstein Distance Between Spheres

Shreya Arya¹ · Arnab Auddy² · Ranthony A. Clark³ · Sunhyuk Lim⁴ ·
Facundo Mémoli⁵ · Daniel Packer⁵

Received: 9 August 2023 / Accepted: 25 July 2024
© The Author(s) 2024

Abstract

The Gromov–Wasserstein distance—a generalization of the usual Wasserstein distance—permits comparing probability measures defined on possibly different metric spaces. Recently, this notion of distance has found several applications in Data Science and in Machine Learning. With the goal of aiding both the interpretability of dissimilarity measures computed through the Gromov–Wasserstein distance and the assessment of the approximation quality of computational techniques designed to estimate the Gromov–Wasserstein distance, we determine the precise value of a certain variant of the Gromov–Wasserstein distance between unit spheres of different dimensions. Indeed, we consider a two-parameter family $\{d_{\text{GW},p,q}\}_{p,q=1}^{\infty}$ of Gromov–Wasserstein distances between metric measure spaces. By exploiting a suitable interaction between specific values of the parameters p and q and the metric of the underlying spaces, we are able to determine the exact value of the distance $d_{\text{GW},4,2}$ between all pairs of unit spheres of different dimensions endowed with their Euclidean distance and their uniform measure.

Keywords Gromov–Wasserstein distances · Metric geometry · Metric-measure spaces · Optimal transport · Monge maps

Mathematics Subject Classification Primary 53C23 · 54E35 · 60D05 · 49Q22 · 90C26

Contents

1	Introduction	1
	Ideas Behind the Proof of Theorem 1 and Related Work	1
	Other Related Work	1
	Organization of the Paper	1
1.1	Notation and Terminology	1
2	The (p, q) -Gromov–Wasserstein Distance Between Spheres	2
2.1	The Equatorial Coupling	2
2.2	Exact Determination of $d_{\text{GW},4,2}(\mathbb{S}_E^m, \mathbb{S}_E^n)$	2
2.3	The Proof of Theorem 1	2

Communicated by Lénaïc Chizat.

Extended author information available on the last page of the article

2.3.1 Preliminaries	...
2.3.2 A Change of Coordinates	...
2.3.3 Optimizing $J(\gamma)$ Over Couplings Between Standard Gaussians	...
2.3.4 The Conclusion of the Proof of Theorem 1	...
3 General Lower Bounds	...
3.1 Invariants	...
3.2 Diameter Lower Bound	...
3.3 Second Lower Bound	...
3.4 Third Lower Bound	...
3.5 The Complete Hierarchy of Lower Bounds	...
3.6 Lower Bounds in the case of Spheres	...
3.6.1 Diameter Lower Bound for Spheres	...
3.6.2 Second and Third Lower Bound for Spheres	...
3.6.3 Lower Bounds for $d_{GW4,2}$ Between Spheres	...
4 Experimental Illustration	...
Experiment with Varying Number of Sample Points	...
Observations	...
5 Conclusions and Perspectives	...
Some Questions	...
Appendix A Relegated Proofs	...
A.1 Proof of Theorem 2	...
A.2 Proof of Proposition 3.10 and an Example	...
A.3 Proofs of Lemmas	...
Appendix B Calculations	...
B.1 Lower Bounds for $d_{GW4,2}$ Between Spheres with the Geodesic Distance	...
B.2 Lower Bounds for $d_{GW4,2}$ Between Spheres with the Euclidean Metric	...
B.3 Distortion $\text{dis}_{4,2}$ Under the Equatorial Coupling	...
Appendix C Another Experiment: Varying Dimensions Experiment	...
References	...

1 Introduction

Shape comparison ideas are utilized in a variety of fields with a wide range of application domains ranging from phylogenetics [9, 25, 44], medicine [47], neuroscience [36, 55], oral biology [43], language structure [2], social and biological networks [7, 23], to political science [17, 24] and computer vision [26, 45]. Many context specific tools have been developed to study the diverse set of problems which appear in these domains. Classical approaches such as statistical landmark analysis [5] turn physical shapes into sequences of vectors, allowing for the rotation-dilation based approach of Procrustes Analysis (see [13, 21, 37]). On the other hand, one can also understand a shape from the perspective of metric geometry, where the essence of a shape is captured by its pairwise interpoint distances [29, 33]. Then in order to compare two shapes, i.e., in order to *quantify* their failure to be isometric, one compares their metric information directly. The Gromov–Hausdorff distance (see [6, 22]), d_{GH} , provides a framework to compare distinct (compact) metric spaces X and Y , where

$$d_{GH}(X, Y) := \frac{1}{2} \inf_{R \in \mathcal{R}(X, Y)} \text{dis}(R)$$

and $\mathcal{R}(X, Y)$ denotes the collection of all *correspondences* between X and Y , that is, all subsets $R \subseteq X \times Y$ such that the canonical projections of R onto the first and

second coordinates satisfy $\pi_1(R) = X$ and $\pi_2(R) = Y$, and where

$$\text{dis}(R) := \sup_{(x,y),(x',y') \in R} |d_X(x, x') - d_Y(y, y')|$$

is the *distortion* of R . The Gromov–Hausdorff distance has been considered in the context of shape comparison and shape classification problems [33]. However, it does not account for the distributional properties of a given data sample. The Gromov–Wasserstein distance [30] offers a robust alternative by viewing the shapes as metric measure spaces: triples (X, d_X, μ_X) where d_X is the metric on X and μ_X is a fully supported Borel probability measure on X .

It is natural to consider metric measure spaces in the context of shape and data comparison, since they allow us to associate to each point in our shape a weight that represents its relative importance within the dataset. The Gromov–Wasserstein distance provides a solution to the problem of finding the “best” way to align two shapes equipped with probability measures, where the best alignment is found by making use of the notion of *coupling*, a cognate of the notion of correspondence which is ubiquitous in the Kantorovich formulation of optimal transport [54]. Given measure spaces (X, μ_X) , (Y, μ_Y) , a coupling between X and Y is a measure γ on the product space $X \times Y$ whose marginals over X and Y are μ_X and μ_Y respectively.¹ We denote the space of all such measures by $\mathcal{M}(\mu_X, \mu_Y)$. Intuitively, couplings align points in X to those in Y . The distortion of a coupling provides insight to how well a given coupling interacts with the underlying metric structures of X and Y in order to preserve distances. For $p \in [1, \infty)$, the p -distortion induced by a coupling $\gamma \in \mathcal{M}(\mu_X, \mu_Y)$ is defined as:

$$\text{dis}_p(\gamma) := \left(\int_{X \times Y} \int_{X \times Y} |d_X(x, x') - d_Y(y, y')|^p \gamma(dx \times dy) \gamma(dx' \times dy') \right)^{1/p}.$$

This distortion is then minimized (see [30]) over all possible couplings to define the p -Gromov–Wasserstein distance between X and Y :

$$d_{\text{GW}_p}(X, Y) := \frac{1}{2} \inf_{\gamma \in \mathcal{M}(\mu_X, \mu_Y)} \text{dis}_p(\gamma).$$

In this work, we consider a two parameter family $d_{\text{GW}_{p,q}}$ (for $p, q \in [1, \infty)$) of Gromov–Wasserstein distances. In contrast to the d_{GW_p} distance recalled above, we consider the (p, q) -distortion of a coupling $\gamma \in \mathcal{M}(\mu_X, \mu_Y)$ defined as (see Definition 1.15):

$$\text{dis}_{p,q}(\gamma) := \left(\int_{X \times Y} \int_{X \times Y} |d_X^q(x, x') - d_Y^q(y, y')|^{p/q} \gamma(dx \times dy) \gamma(dx' \times dy') \right)^{1/p}, \quad (1)$$

¹ More precisely, the pushforwards of γ under the canonical projection maps π_1 and π_2 satisfy $(\pi_1)_\# \gamma = \mu_X$ and $(\pi_2)_\# \gamma = \mu_Y$.

which is then minimized over all possible couplings to define the (p, q) -Gromov–Wasserstein distance between X and Y :

$$d_{\text{GW}_{p,q}}(X, Y) := \frac{1}{2} \inf_{\gamma \in \mathcal{M}(\mu_X, \mu_Y)} \text{dis}_{p,q}(\gamma).$$

The (p, q) -Gromov–Wasserstein distance $d_{\text{GW}_{p,q}}$ interpolates between two previously studied versions of the Gromov–Wasserstein distance: for $q = 1$, it reduces to the d_{GW_p} distance of [30], while for $q = \infty$, it coincides with the p -ultrametric Gromov–Wasserstein distance defined in [31].

This formulation exhibiting one additional parameter makes the Gromov–Wasserstein distances more amenable to analysis. Raising the distances to the q -th power allows for more explicit control of the difference in distances by emphasizing structural properties of d_X and d_Y . In this sense, of particular interest is the case of the Euclidean metric, with $q = 2$, where taking squares of distances allows one to move from norms to inner products. A construction related to the case of $q = 2$ and $p = 4$ was considered in [12] to determine the value of the Gromov–Wasserstein distances between arbitrary Gaussian measures. A similar notion of (p, q) -distortion was considered by Sturm in [48] with the distinction that the difference between the q -th powers of the distances is raised to the p -th power (as opposed to the (p/q) -th power). This implies that the $d_{\text{GW}_{p,q}}$ distance we consider has absolute homogeneity, while Sturm’s version does not (see Remark 1.22).

We now connect the (p, q) -Gromov–Wasserstein distance to some existing computational approaches. Note that the computation of $d_{\text{GW}_{p,q}}(X, Y)$ involves optimizing the (p, q) -distortion over the set of all possible couplings $\gamma \in \mathcal{M}(\mu_X, \mu_Y)$. This reduces to a non-convex quadratic optimization problem [30, Section 7], which is in general NP-hard (see, e.g., [41, 46]). Nonetheless, there exist numerous computational approaches to find approximate solutions to the above problem and its variants: see [2, 7, 8, 41, 50–52] and references therein. Perhaps the most standard of these approaches is the use of gradient descent algorithms. In the absence of an algorithm that provably finds the global optima of this problem, practitioners often depend on heuristic initializations and find local optima through these gradient based methods. It is hence essential to assess the (sub)-optimality of each of these local optima.

A particularly popular approach for assessing sub-optimality is to consider lower bounds of d_{GW_p} by using “signatures” or invariants of metric measure spaces (see [30, 32]). We exhibit three invariant based lower bounds for the (p, q) -Gromov–Wasserstein distance between two arbitrary metric measure spaces. All these three bounds are constructed from signatures related to particular invariants of and distributions on metric measure spaces. For example, the (p, q) -Second Lower Bound, denoted $\text{SLB}_{p,q}$, is a counterpart to the Second Lower Bound for d_{GW_p} from [30], and utilizes the global distribution of distances between points of the two metric measure spaces. In Proposition 3.10 we also establish a hierarchy of poly-time computable lower bounds for this distance in the spirit of [7, 30–32].

As discussed above, these lower bounds aspire to be useful for determining whether the output of an algorithm is sufficiently close to the global optimum. Even though the exact values of the (p, q) -Gromov–Wasserstein distances may not always be available,

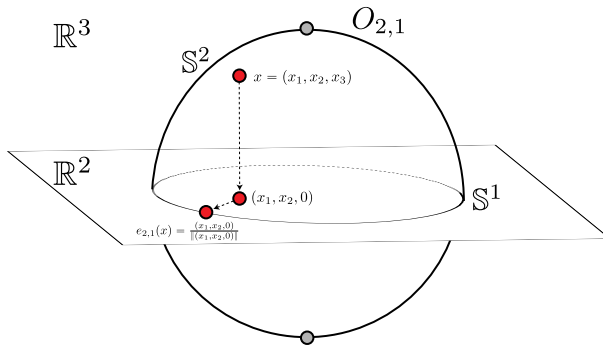


Fig. 1 A depiction of the equatorial map $e_{2,1} : \mathbb{S}^2 \setminus O_{2,1} \rightarrow \mathbb{S}^1$. The measure zero set $O_{2,1}$ consists of the north and south poles of $\mathbb{S}^2$. See Definition 2.1 for the general case

in practice one can compare the objective function, i.e. the (p, q) -distortion, at the computed (local) optima against our lower bounds to evaluate the performance of the given algorithm. In the same spirit, lower bounds are useful in accelerating shape classification, where knowing the relative strengths of the lower bounds allows one to progressively filter out comparisons of the most distinct examples by comparing examples with successively stronger lower bounds from the hierarchy (see [30]).

We further compute the precise value of these lower and upper bounds on the $d_{\text{GW},p,q}(X, Y)$ distance in the case where $p = 4, q = 2$ and the metric measure spaces X, Y are spheres equipped with geodesic or Euclidean distances, and uniform measures. Spheres, being canonical spaces, are a natural starting point for understanding Gromov–Wasserstein distances. These can provide useful *benchmarks* for assessing the quality of a given algorithm for estimating Gromov–Wasserstein and related distances (see, e.g., [27]).

Furthermore, with the goal of providing such benchmarks, in Theorem 1 we determine, for all non-negative integers m and n , the *exact* value of $d_{\text{GW},4,2}(\mathbb{S}_E^m, \mathbb{S}_E^n)$, i.e., the $(4, 2)$ -Gromov–Wasserstein distance between m and n dimensional unit spheres equipped with the Euclidean metric and uniform measures. Our results imply that, in this setting, optimal (i.e. $(4, 2)$ -distortion minimizing) couplings exist and, more importantly, that one distinguished such optimal coupling $\gamma_{m,n}$ is induced by the Monge map that projects the first m coordinates from the n -dimensional unit sphere ($m \leq n$) to the m -dimensional unit sphere. Inspired by the case of $m = 1$ and $n = 2$, we call this the *equatorial map* (see Definition 2.1 and Fig. 1) and call $\gamma_{m,n}$ the *equatorial coupling*.

Theorem 1 (Main Theorem) *The equatorial coupling $\gamma_{m,n}$ is an optimal coupling for $d_{\text{GW},4,2}(\mathbb{S}_E^m, \mathbb{S}_E^n)$. In particular, for $n \geq m \geq 0$,*

$$\begin{aligned} d_{\text{GW}4,2}(\mathbb{S}_E^m, \mathbb{S}_E^n) &= \frac{1}{2} \text{dis}_{4,2}(\gamma_{m,n}, \mathbb{S}_E^m, \mathbb{S}_E^n) \\ &= \frac{1}{\sqrt{2}} \left[\frac{1}{m+1} + \frac{1}{n+1} - \frac{2}{m+1} \left(\frac{\Gamma(\frac{m+2}{2}) \Gamma(\frac{n+1}{2})}{\Gamma(\frac{m+1}{2}) \Gamma(\frac{n+2}{2})} \right)^2 \right]^{1/4}. \end{aligned}$$

Note that when $n = m + 2$ the expression given in the theorem reduces to

$$d_{\text{GW}4,2}(\mathbb{S}_E^m, \mathbb{S}_E^{m+2}) = \frac{1}{2^{1/4}} \left[\frac{1}{(m+1)(m+3)} + \frac{1}{(m+2)^2(m+3)} \right]^{1/4}.$$

In particular, this means that $d_{\text{GW}4,2}(\mathbb{S}_E^m, \mathbb{S}_E^{m+2}) = O(m^{-1/2})$ as $m \rightarrow \infty$. These asymptotics do not depend on the fact that $n = m + 2$. Indeed, as explained in Remark 2.11, $d_{\text{GW}4,2}(\mathbb{S}_E^m, \mathbb{S}_E^n) = O(m^{-1/2})$ when $n = m + k_0$ for a fixed k_0 . A different interesting regime (see Remark 2.12) is the one when $n = m + k$ for fixed m but $k \rightarrow \infty$; in that case we have

$$d_{\text{GW}4,2}(\mathbb{S}_E^m, \mathbb{S}_E^{m+k}) = \frac{1}{\sqrt{2}} \left[\frac{1}{m+1} \right]^{1/4} + O(k^{-1/4})$$

as $k \rightarrow \infty$. See Sect. 2.2 for additional related results.

Ideas Behind the Proof of Theorem 1 and Related Work

The proof proceeds by first observing that minimizing the $(4, 2)$ -distortion over all couplings γ between the respective uniform measures on $\mathbb{S}_E^m$ and $\mathbb{S}_E^n$ is equivalent to *maximizing* the functional

$$J(\gamma) := \left\| \underbrace{\int_{\mathbb{S}^m \times \mathbb{S}^n} xy^\top \gamma(dx \times dy)}_{M_\gamma} \right\|_F^2.$$

Then, via the singular value decomposition of the matrix M_γ , we identify a change of coordinates (see Sect. 2.3.2) which allows us to argue that one can restrict attention to couplings $\tilde{\gamma}$ for which the matrix $M_{\tilde{\gamma}}$ has the form $(\Lambda_\gamma \mathbf{0}_{(m+1) \times (n-m)})$ where Λ_γ is a $m \times m$ diagonal matrix containing the singular values of M_γ —a simplification which, in turn, implies that one can equivalently restrict attention to maximizing the (simpler) functional

$$D(\gamma) := \sum_{k=1}^{m+1} \left[\int_{\mathbb{S}^m \times \mathbb{S}^n} x_k y_k \gamma(dx \times dy) \right]^2.$$

The final step of the proof can be roughly described as an intricate application of the Cauchy–Schwarz inequality to a suitably disintegrated expression of the functional D .

To better describe some of the main ideas behind the proof of Theorem 1 without these intricacies, we state and prove a subsidiary result (Proposition 2.16, which is particular case of [12, Proposition 4.1]) about maximizing the functional J in the case when γ ranges over all couplings between standard Gaussian measures. We then show how to suitably alter the proof of Proposition 2.16 in order to establish Theorem 1.

The functional $J(\gamma)$ appears naturally when considering certain inner product based variants of the Gromov–Wasserstein distance [12, 15, 51]. The change of coordinates step appeared in Vayer’s PhD thesis [49, Chapter 4]. It was also used in [15, Theorem 3.2] when studying the existence of Monge maps for the J functional defined above.

There are results in the literature providing a precise description of optimal Monge maps in the context of Gromov–Wasserstein distances but none of those seems to be applicable in our setting:

- In [48, Section 9.4] Sturm provides such a characterization result for optimal couplings under the assumption that both measures are rotationally invariant and absolutely continuous w.r.t. the Lebesgue measure *in the same ambient Euclidean space*. The setting of Theorem 1 does not fit into the one considered by Sturm: the measures considered therein are not absolutely continuous. One might nevertheless contemplate applying Sturm’s result to suitable smoothings of the uniform measures on the respective spheres. However, it does not seem possible to guarantee the smoothed measure resulting from the lower dimensional sphere to be rotationally invariant.
- In [15, Theorem 3.2] Dumont et al. give a precise description of optimal Monge maps for the J functional under the assumption that one of the measures is absolutely continuous w.r.t. Lebesgue measure.² Therefore their results are also not applicable to our setting.
- In [12, Proposition 4.1] Salmona et al. contend with the case of Gaussian measures and find the precise structure of an optimal coupling that, in the case of standard Gaussians, boils down to a coupling between Gaussians with similar structure to the equatorial coupling. As far as we know, it does not seem possible to apply their results in our setting (i.e. in order to establish Theorem 1); see Question 3 and also Sect. 2.3.3.

Other Related Work

Finally, this project is related to a recent effort to compute the precise value of the (closely related) Gromov–Hausdorff distance between spheres [1, 27]. In [27], the authors provide nontrivial upper and lower bounds for the Gromov–Hausdorff distance $d_{\text{GH}}(\mathbb{S}^m, \mathbb{S}^n)$ between spheres $\mathbb{S}^m$ and $\mathbb{S}^n$ (endowed with the geodesic metric) for $0 \leq m < n \leq \infty$. Some of these lower bounds were motivated by topological ideas related to a quantitative version of the Borsuk–Ulam theorem [14]. Through explicit constructions of (optimal) correspondences it was proved that their lower bounds

² Their results generalize [49, Theorem 4.2.3]

were tight in the cases of $d_{\text{GH}}(\mathbb{S}^0, \mathbb{S}^n)$, $d_{\text{GH}}(\mathbb{S}^m, \mathbb{S}^\infty)$, $d_{\text{GH}}(\mathbb{S}^1, \mathbb{S}^2)$, $d_{\text{GH}}(\mathbb{S}^1, \mathbb{S}^3)$, and $d_{\text{GH}}(\mathbb{S}^2, \mathbb{S}^3)$. Interestingly, the optimal correspondences achieving these distances are very different in nature from the optimal coupling achieving the exact value of $d_{\text{GW}4,2}(\mathbb{S}_E^m, \mathbb{S}_E^n)$, in the sense that these optimal correspondences are induced by highly irregular maps, whereas the optimal coupling described above is induced by certain natural projection maps.

Organization of the Paper

The rest of the paper is organized as follows.

Section 1.1 introduces notation and terminology that will be used throughout the paper.

Section 2 is the central section of our paper. There we introduce the requisite background and supporting results used to prove Theorem 1, which we also do therein.

In Sect. 3 we recall several invariants of metric measure spaces and use them to prove lower bounds for the (p, q) -Gromov–Wasserstein distance. In Sect. 3.6 we evaluate those lower bounds for spheres with their Euclidean and geodesic distances. For the former case we compare these lower bounds against the exact value provided by Theorem 1.

Section 4 contains a description of some experiments illustrating the result from Theorem 1. In particular, our experiments provide a computational perspective and an indication of the performance of discrete Gromov–Wasserstein solvers from [18] and [10] when estimating the Gromov–Wasserstein distance between spheres.

Section 5 provides a discussion and contains several questions (Questions 1, 2, 3, and 4) that might suggest further research directions.

To enhance readability, the proofs of several results and other details are relegated to Appendices A, B, and C.

1.1 Notation and Terminology

We now define the main concepts used in the paper.

Given a measurable space (X, Σ_X) , we denote the set of all probability measures on X by $\mathcal{P}(X)$.

Definition 1.1 Let (X, Σ_X, μ_X) and (Y, Σ_Y, μ_Y) be measure spaces such that $\mu_X \in \mathcal{P}(X)$ and $\mu_Y \in \mathcal{P}(Y)$. A *coupling* between μ_X and μ_Y is a (probability) measure γ on the product space $(X \times Y, \Sigma_X \otimes \Sigma_Y)$ such that

$$\gamma(A \times Y) = \mu_X(A) \quad \text{and} \quad \gamma(X \times B) = \mu_Y(B)$$

for all $A \in \Sigma_X$ and $B \in \Sigma_Y$. We denote the set of all couplings between μ_X and μ_Y by $\mathcal{M}(\mu_X, \mu_Y)$. Note that $\mathcal{M}(\mu_X, \mu_Y)$ is never empty as it always contains the product measure $\mu_X \otimes \mu_Y$.

Let (X, Σ_X, μ_X) and (Y, Σ_Y, μ_Y) be measure spaces such that $Y = \{y_0\}$, $\mu_X \in \mathcal{P}(X)$ and $\mu_Y \in \mathcal{P}(Y)$. Then, $\mu_Y = \delta_{y_0}^Y$ and $\mathcal{M}(\mu_X, \mu_Y)$ contains exactly one coupling. That is, $\mathcal{M}(\mu_X, \mu_Y) = \{\mu_X \otimes \delta_{y_0}^Y\}$ where $\mu_Y = \delta_{y_0}^Y$ is a Dirac delta.

Let (X, Σ_X) and (Y, Σ_Y) be measurable spaces and $T : X \rightarrow Y$ a measurable map. The *pushforward* of a measure α on (X, Σ_X) by T , denoted $T_{\#}\alpha$, is the measure on Y given by

$$T_{\#}\alpha(A) = \alpha(T^{-1}(A))$$

for every $A \in \Sigma_Y$. We can then describe the set of all couplings between μ_X and μ_Y as

$$\mathcal{M}(\mu_X, \mu_Y) = \{\gamma \in \mathcal{P}(X \times Y) \mid (\pi_X)_{\#}\gamma = \mu_X, (\pi_Y)_{\#}\gamma = \mu_Y\}$$

where $\pi_X : X \times Y \rightarrow X$ and $\pi_Y : X \times Y \rightarrow Y$ are the canonical projections onto the first and second components respectively.

Given a topological space X , unless indicated otherwise, we will assume all measures on X to be Borel measures, and will denote the Borel sigma algebra of X by Σ_X . Furthermore, in this case, $\mathcal{P}(X)$ will denote the set of all Borel probability measures on X .

Definition 1.2 The *support* of a Borel measure α on a topological space X is the smallest closed subset $X_0 \subset X$ so that $\alpha(X \setminus X_0) = 0$, that is, for any $A \subset X$, if $A \cap X_0 = \emptyset$, then $\alpha(A) = 0$. We denote the support of α by $\text{supp}[\alpha]$.

Definition 1.3 Let α be a Borel probability measure on $\mathbb{R}$ and $r \in [1, \infty)$. Then the *r-moment* of α is

$$m_r(\alpha) := \left(\int_{\mathbb{R}} x^r \alpha(dx) \right)^{1/r}.$$

Now let (X, d_X) be a metric space. We define,

$$\mathcal{P}_r(X) := \{\mu \in \mathcal{P}(X) \mid d_X(x_0, \cdot)_{\#}\mu \text{ has finite } r\text{-moment for some } x_0 \in X\}.$$

In fact, the choice of x_0 is immaterial – if the moment is finite for one reference point, it is finite for any reference point.

Definition 1.4 Let (X, d_X) be a metric space, $r \in [1, \infty]$, and $\alpha, \beta \in \mathcal{P}_r(X)$.³ The *r-Wasserstein* distance on X between α and β is given by

$$d_{W_r}^X(\alpha, \beta) := \inf_{\gamma \in \mathcal{M}(\alpha, \beta)} \left(\int_{X \times X} d_X^r(x, x') \gamma(dx \times dx') \right)^{1/r}$$

³ We define $\mathcal{P}_{\infty}(X)$ as the set of those probability measures on X with bounded support.

for $1 \leq r < \infty$, and

$$d_{W\infty}^X(\alpha, \beta) := \inf_{\gamma \in \mathcal{M}(\alpha, \beta)} \sup_{(x, x') \in \text{supp}[\gamma]} d_X(x, x').$$

Example 1.5 When $\alpha, \beta \in \mathcal{P}_r(\mathbb{R})$, the r -Wasserstein distance on $\mathbb{R}$ (with the usual metric) can be explicitly computed as follows (see, e.g., [54]):

$$d_{W_r}^{\mathbb{R}}(\alpha, \beta) = \left(\int_0^1 |F_\alpha^{-1}(u) - F_\beta^{-1}(u)|^r du \right)^{1/r} \quad (2)$$

where $F_\alpha(t) := \alpha((-\infty, t])$ and $F_\beta(t) := \beta((-\infty, t])$ are the cumulative distributions of α and β , respectively, and their generalized inverses are defined as:

$$F_\alpha^{-1}(u) := \inf\{t \in \mathbb{R} \mid F_\alpha(t) > u\} \quad (3)$$

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

Example 1.6 For $r \in [1, \infty)$, the r -Wasserstein distance on the real line between $\alpha \in \mathcal{P}_r(\mathbb{R})$ and the Dirac delta δ_0 equals the r -moment of α :

$$d_{W_r}^{\mathbb{R}}(\alpha, \delta_0) = \left(\int_{\mathbb{R} \times \mathbb{R}} |t - s|^r (\alpha \otimes \delta_0)(dt \times ds) \right)^{1/r} = \left(\int_{\mathbb{R}_+} t^r \alpha(dt) \right)^{1/r} = m_r(\alpha).$$

Definition 1.7 For each $q \in [1, \infty]$, we define $\Lambda_q : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ in the following way (cf. [35]):

$$\begin{aligned} \Lambda_q(a, b) &:= |a^q - b^q|^{\frac{1}{q}} \text{ if } q < \infty, \text{ and} \\ \Lambda_\infty(a, b) &:= \begin{cases} \max\{a, b\} & \text{if } a \neq b \\ 0 & \text{if } a = b. \end{cases} \end{aligned}$$

Remark 1.8 Note that $\Lambda_1(a, b) = |a - b|$ for all $a, b \geq 0$. One of the claims of the following proposition is that Λ_q is a metric on $\mathbb{R}_+$ for each $q \geq 1$.

Proposition 1.9 ([35, Lemma 2.2, Example 2.7, Proposition 2.11]) Λ_q defines a metric on $\mathbb{R}_+$ for each $q \in [1, \infty]$, i.e., it is symmetric, non-negative, it satisfies $\Lambda_q(a, b) = 0$ if and only if $a = b$, and it satisfies the triangle inequality:

$$\Lambda_q(a, b) \leq \Lambda_q(a, c) + \Lambda_q(c, b) \text{ for all } a, b, c \geq 0.$$

Also, if $1 \leq q \leq q' \leq \infty$, then $\Lambda_q \leq \Lambda_{q'}$.

The fact that $(\mathbb{R}_+, \Lambda_q)$ is a metric space, enables us to consider the p -Wasserstein distance $d_{W_p}^{(\mathbb{R}_+, \Lambda_q)}$ as a generalization of $d_{W_p}^{(\mathbb{R}_+, \Lambda_1)} = d_{W_p}^{\mathbb{R}_+}$.

Remark 1.10 (Closed-form solution for $d_{\mathbb{W}_p}^{(\mathbb{R}_+, \Lambda_q)}$) For $1 \leq q \leq p < \infty$, we have the following equality which generalizes Example 1.5:

$$d_{\mathbb{W}_p}^{(\mathbb{R}_+, \Lambda_q)}(\alpha, \beta) = \left(\int_0^1 (\Lambda_q(F_\alpha^{-1}(u), F_\beta^{-1}(u)))^p du \right)^{1/p}.$$

While the above equality is a special case of [31, Theorem A.4], we include a proof here for pedagogical reasons:

$$\begin{aligned} \left(d_{\mathbb{W}_p}^{(\mathbb{R}_+, \Lambda_q)}(\alpha, \beta) \right)^p &= \inf_{\gamma \in \mathcal{M}(\alpha, \beta)} \int_{\mathbb{R}_+ \times \mathbb{R}_+} (\Lambda_q(a, b))^p \gamma(da \times db) \\ &= \inf_{\gamma \in \mathcal{M}(\alpha, \beta)} \int_{\mathbb{R}_+ \times \mathbb{R}_+} |S_q(a) - S_q(b)|^{p/q} \gamma(da \times db) \\ &= \inf_{\gamma \in \mathcal{M}(\alpha, \beta)} \int_{\mathbb{R}_+ \times \mathbb{R}_+} |s - t|^{p/q} ((S_q)_\# \gamma)(ds \times dt) \\ &= \inf_{\gamma \in \mathcal{M}((S_q)_\# \alpha, (S_q)_\# \beta)} \int_{\mathbb{R}_+ \times \mathbb{R}_+} |s - t|^{p/q} \gamma(ds \times dt) \\ &= \int_0^1 \left| F_{(S_q)_\# \alpha}^{-1}(u) - F_{(S_q)_\# \beta}^{-1}(u) \right|^{p/q} du \\ &= \int_0^1 (\Lambda_q(F_\alpha^{-1}(u), F_\beta^{-1}(u)))^p du \end{aligned}$$

where $S_q : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is the map sending x to x^q . The fourth equality holds by [7, Lemma 3.2], fifth equality holds by Example 1.5, and the last equality holds since $F_{(S_q)_\# \alpha}^{-1} = (F_\alpha^{-1})^q$.

The following remark makes the connection between the generalized Wasserstein distance $d_{\mathbb{W}_p}^{(\mathbb{R}_+, \Lambda_q)}$ and the Wasserstein distance on $\mathbb{R}_+$ with the usual metric.

Remark 1.11 (Relationship between $d_{\mathbb{W}_p}^{(\mathbb{R}_+, \Lambda_q)}$ and $d_{\mathbb{W}_{p/q}}^{\mathbb{R}}$) In the previous remark, we saw that when $1 \leq q \leq p < \infty$,

$$\begin{aligned} d_{\mathbb{W}_p}^{(\mathbb{R}_+, \Lambda_q)}(\alpha, \beta) &= \left(\int_0^1 (\Lambda_q(F_\alpha^{-1}(u), F_\beta^{-1}(u)))^p du \right)^{1/p} \\ &= \left(\int_0^1 |F_{(S_q)_\# \alpha}^{-1}(u) - F_{(S_q)_\# \beta}^{-1}(u)|^{p/q} du \right)^{1/p}. \end{aligned}$$

The right hand side of the above expression coincides with the p/q -Wasserstein distance between the measures $(S_q)_\# \alpha$ and $(S_q)_\# \beta$ on $\mathbb{R}_+$ with the usual metric Λ_1 as follows:

$$d_{\mathbb{W}_p}^{(\mathbb{R}_+, \Lambda_q)}(\alpha, \beta) = \left(d_{\mathbb{W}_{p/q}}^{\mathbb{R}}((S_q)_\# \alpha, (S_q)_\# \beta) \right)^{1/q}.$$

Definition 1.12 A *metric measure space* is a triple (X, d_X, μ_X) where (X, d_X) is a compact metric space, and μ_X is a Borel probability measure on X such that $\text{supp}[\mu_X] = X$. We denote the collection of all metric measure spaces by $\mathcal{G}_w$. We will often abuse notation and write X to represent the triple $(X, d_X, \mu_X) \in \mathcal{G}_w$.

The next example is central to our paper.

Example 1.13 ($\mathbb{S}_E^n$ and $\mathbb{S}_G^n$) For each integer $n \geq 1$, we consider the n -dimensional unit sphere $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ as a metric measure space by equipping it with the uniform measure and the geodesic or Euclidean metric. For example, when endowed with its geodesic distance, the usual n -dimensional unit sphere gives rise to $(\mathbb{S}^n, d_n, \mu_n) \in \mathcal{G}_w$, where $d_n(x, x') := \arccos(\langle x, x' \rangle)$ for $x, x' \in \mathbb{S}^n$. We henceforth write $\mathbb{S}_E^n$ and $\mathbb{S}_G^n$ to denote the spheres equipped with the Euclidean and geodesic metrics, respectively, as metric measure spaces.

We also consider $\mathbb{S}_G^0$, the 0-dimensional sphere consisting of two points at distance π and, similarly, $\mathbb{S}_E^0$ consists of two points at distance 2. In both cases we view these 0-dimensional spheres as metric measure spaces by endowing them with the uniform measure (on two points). Note that $\text{diam}(\mathbb{S}_G^n) = \pi$ and $\text{diam}(\mathbb{S}_E^n) = 2$ for all integers $n \geq 0$.

Definition 1.14 (p -diameter). The *diameter* $\text{diam}(A)$ of bounded subset A of a metric space (X, d_X) is defined as

$$\text{diam}(A) := \sup_{x, x' \in A} d_X(x, x').$$

Let $(X, d_X, \mu_X) \in \mathcal{G}_w$. The p -diameter of X for $p \in [1, \infty]$ is:

$$\text{diam}_p(X) := \left(\int_X \int_X d_X^p(x, x') \mu_X(dx) \mu_X(dx') \right)^{1/p} \text{ for } 1 \leq p < \infty$$

and

$$\text{diam}_\infty(X) := \text{diam}(\text{supp}[\mu_X]).$$

Definition 1.15 ((p, q) -distortion). Let (X, d_X, μ_X) and (Y, d_Y, μ_Y) be metric measure spaces and let $\gamma \in \mathcal{M}(\mu_X, \mu_Y)$. Then, for each $p, q \in [1, \infty]$, the (p, q) -distortion of the coupling γ is defined as:

$$\text{dis}_{p,q}(\gamma) := \left(\int_{X \times Y} \int_{X \times Y} (\Lambda_q(d_X(x, x'), d_Y(y, y')))^p \gamma(dx \times dy) \gamma(dx' \times dy') \right)^{1/p}$$

for $1 \leq p < \infty$, and

$$\text{dis}_{\infty,q}(\gamma) := \sup_{(x,y), (x',y') \in \text{supp}[\gamma]} \Lambda_q(d_X(x, x'), d_Y(y, y')).$$

Example 1.16 Consider the coupling $\gamma = \{\mu_X \otimes \delta_{y_0}^Y\} \in \mathcal{M}(\mu_X, \mu_Y)$ where $Y = \{y_0\}$ is the one point metric measure space. Then, for all $p, q \in [1, \infty]$, one can verify that $\text{dis}_{p,q}(\gamma) = \text{diam}_p(X)$ as follows:

$$\begin{aligned} & \text{dis}_{p,q}(\{\mu_X \otimes \delta_{y_0}^Y\}) \\ &= \left(\int_{X \times Y} \int_{X \times Y} (\Lambda_q(d_X(x, x'), d_Y(y_0, y_0)))^p \gamma(dx \times dy) \gamma(dx' \times dy') \right)^{1/p} \\ &= \left(\int_{X \times Y} \int_{X \times Y} |d_X^q(x, x')|^{p/q} \gamma(dx \times dy) \gamma(dx' \times dy') \right)^{1/p} \\ &= \left(\int_{X \times X} d_X^p(x, x') \mu_X(dx) \mu_X(dx') \right)^{1/p} \\ &= \text{diam}_p(X) \end{aligned}$$

for the $p < \infty$ case. The $p = \infty$ case can be checked in a similar way.

Example 1.17 Let $p = 4$ and $q = 2$ and $\gamma \in \mathcal{M}(\mu_X, \mu_Y)$, then we have:

$$\begin{aligned} (\text{dis}_{4,2}(\gamma))^4 &= \left(\int_{X \times Y} \int_{X \times Y} (d_X^2(x, x') - d_Y^2(y, y'))^2 \gamma(dx \times dy) \gamma(dx' \times dy') \right) \\ &= \int_{X \times X} d_X^4(x, x') \mu_X(dx) \mu_X(dx') + \int_{Y \times Y} d_Y^4(y, y) \mu_Y(dy) \mu_Y(dy') \\ &\quad - 2 \int_{X \times Y} \int_{X \times Y} d_X^2(x, x') d_Y^2(y, y') \gamma(dx \times dy) \gamma(dx' \times dy'). \end{aligned}$$

Remark 1.18 Note that the marginals μ_X and μ_Y determine the first two terms in Example 1.17 (in fact the sum of the first two terms is $(\text{diam}_4(X))^4 + (\text{diam}_4(Y))^4$) and thus,

$$\begin{aligned} \gamma \text{ minimizes } \text{dis}_{4,2}(\gamma) &\Leftrightarrow \\ \gamma \text{ maximizes } &\int_{X \times Y} \int_{X \times Y} d_X^2(x, x') d_Y^2(y, y') \gamma(dx \times dy) \gamma(dx' \times dy'). \end{aligned}$$

The equivalence above will prove instrumental in Sect. 2 of our paper where we prove the optimality of a coupling for achieving the $(4, 2)$ -Gromov–Wasserstein distance between spheres. Passing to the distance squared allows us to unfold Euclidean distances into expressions that depend solely on inner products. Our proof of the theorem depends on the favorable interplay between these inner products and linear maps. In fact, our introduction of the broader family of (p, q) -Gromov–Wasserstein distances was motivated by this ease of analysis in the case $p = 4, q = 2$.

Definition 1.19 ((p, q) -Gromov–Wasserstein distance). Let (X, d_X, μ_X) and (Y, d_Y, μ_Y) be metric measure spaces. Let $\gamma \in \mathcal{M}(\mu_X, \mu_Y)$. The (p, q) -Gromov–Wasserstein dis-

tance between μ_X and μ_Y is given by one-half of the infimum of the (p, q) -distortion:

$$d_{\text{GW},p,q}(X, Y) := \frac{1}{2} \inf_{\gamma \in \mathcal{M}(\mu_X, \mu_Y)} \text{dis}_{p,q}(\gamma).$$

where $p, q \in [1, \infty]$.

Remark 1.20 In the case $q = 1$ we recover the p -Gromov–Wasserstein distance $d_{\text{GW},p}$ from [30].

$$d_{\text{GW},p,1}(X, Y) = \frac{1}{2} \inf_{\gamma \in \mathcal{M}(\mu_X, \mu_Y)} \text{dis}_{p,1}(\gamma) = d_{\text{GW},p}(X, Y).$$

In the case $q = \infty$ we recover the p -ultrametric Gromov–Wasserstein distance $u_{\text{GW},p}$ from [31].

$$d_{\text{GW},p,\infty}(X, Y) = \frac{1}{2} \inf_{\gamma \in \mathcal{M}(\mu_X, \mu_Y)} \text{dis}_{p,\infty}(\gamma) = \frac{1}{2} u_{\text{GW},p}(X, Y).$$

Example 1.21 Let $X, Y \in \mathcal{G}_w$ where $Y = \{y_0\}$. It follows from Example 1.16,

$$d_{\text{GW},p,q}(X, Y) = \frac{1}{2} \text{dis}_{p,q}(\{\mu_X \otimes \delta_{y_0}^Y\}) = \frac{1}{2} \text{diam}_p(X)$$

for all $p, q \in [1, \infty]$.

The following theorem shows that the (p, q) -Gromov–Wasserstein distance is a well defined metric on $\mathcal{G}_w$. This is a generalization of both Theorem 5.1 in [30], which shows that the original p -Gromov–Wasserstein distance $d_{\text{GW},p}$ is a metric on $\mathcal{G}_w$, and of Theorem 3.10 in [31], which shows that the ultrametric p -Gromov–Wasserstein distance $u_{\text{GW},p}$ is a p -metric on the collection of compact ultrametric spaces.⁴

Theorem 2 The (p, q) -Gromov–Wasserstein distance, $d_{\text{GW},p,q}$, is a metric on the collection of isomorphism classes of $\mathcal{G}_w$ for all $p, q \in [1, \infty]$. Furthermore, $d_{\text{GW},p,q} \leq d_{\text{GW},p',q'}$ whenever $p \leq p'$ and $q \leq q'$.

We defer the proof of this theorem to Sect. A.1.

Remark 1.22 In [48, Section 9] Sturm considers a two parameter family of distances, $\mathbb{A}_{p,q}$, which is closely related to but differs from $d_{\text{GW},p,q}$. A precise relationship is,

$$\frac{1}{2} \mathbb{A}_{p/q,q}(X, Y) = (d_{\text{GW},p,q}(X, Y))^q,$$

for $X, Y \in \mathcal{G}_w$ and $p, q \in [1, \infty)$. See Remark A.2.

Also, in contrast with $d_{\text{GW},p,q}$, $\mathbb{A}_{p,q}$ is not homogeneous: if for $\lambda \geq 0$, λX denotes the metric measure space $(X, \lambda d_X, \mu_X)$ (resp. for λY), $\mathbb{A}_{p,q}(\lambda X, \lambda Y) = \lambda^q \mathbb{A}_{p,q}(X, Y)$ whereas $d_{\text{GW},p,q}(\lambda X, \lambda Y) = \lambda d_{\text{GW},p,q}(X, Y)$.

⁴ For $p \geq 1$, a p -metric $d : X \times X \rightarrow \mathbb{R}_+$ on a set X is any metric on X satisfying the following strengthened triangle inequality: $d^p(x, x') \leq d^p(x, x'') + d^p(x'', x')$ for all $x, x', x'' \in X$.

2 The (p, q) -Gromov–Wasserstein Distance Between Spheres

Despite the increasing number of applications, the precise value of the Gromov–Wasserstein distance is only known for a few cases [12, 30]. In this section, we compute the exact value of $d_{\text{GW},4,2}(\mathbb{S}_E^m, \mathbb{S}_E^n)$ for arbitrary m and n .

2.1 The Equatorial Coupling

In the next two sections, we will consider the equatorial coupling (defined below) and show that when $p = 4$ and $q = 2$, the equatorial coupling is optimal for the case of spheres with Euclidean distance.

Assuming $n > m$ we will implicitly use the (isometric) embedding $\mathbb{R}^{m+1} \hookrightarrow \mathbb{R}^{n+1}$ given by

$$(x_1, \dots, x_{m+1}) \mapsto (x_1, \dots, x_{m+1}, 0, \dots, 0).$$

Definition 2.1 (Projection and Equatorial map). For all $n > m$, we define the *projection map* $\pi_{n+1,m+1} : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{m+1}$ in the following way:

$$\begin{aligned} \pi_{n+1,m+1} : \mathbb{R}^{n+1} &\rightarrow \mathbb{R}^{m+1} \\ (x_1, \dots, x_{n+1}) &\mapsto (x_1, \dots, x_{m+1}). \end{aligned}$$

Note that $\pi_{n+1,m+1}$ is a measurable map from $\mathbb{R}^{n+1}$ to $\mathbb{R}^{m+1}$.

The *equatorial map* $e_{n,m} : \mathbb{S}^n \setminus O_{n,m} \rightarrow \mathbb{S}^m$ is defined in the following way:

$$\begin{aligned} e_{n,m} : \mathbb{S}^n \setminus O_{n,m} &\rightarrow \mathbb{S}^m \\ (x_1, \dots, x_{n+1}) &\mapsto \frac{\pi_{n+1,m+1}(x_1, \dots, x_{n+1})}{\|\pi_{n+1,m+1}(x_1, \dots, x_{n+1})\|} = \frac{(x_1, \dots, x_{m+1})}{\|(x_1, \dots, x_{m+1})\|}, \end{aligned}$$

where $O_{n,m} := \{x \in \mathbb{S}^n \subset \mathbb{R}^{n+1} \mid x_1 = \dots = x_{m+1} = 0\}$. See Fig. 1.

Remark 2.2 Note that, since $\mu_n(O_{n,m}) = 0$, one can construct a measurable extension $\tilde{e}_{n,m} : \mathbb{S}^n \rightarrow \mathbb{S}^m$ of $e_{n,m}$ by setting $\tilde{e}_{n,m}|_{O_{n,m}} \equiv \tilde{z}$, where $\tilde{z}$ is an arbitrary point in $\mathbb{S}^m$. We will incur a slight abuse of notation and use the symbol $e_{n,m}$ to denote one such extension.

We then have the following claim whose proof we omit for brevity.

Claim 1 For all $n > m$, the equatorial map $e_{n,m} : \mathbb{S}^n \rightarrow \mathbb{S}^m$ induces a coupling $\gamma_{m,n} \in \mathcal{M}(\mu_m, \mu_n)$, where μ_m and μ_n are the uniform measures on $\mathbb{S}^m$ and $\mathbb{S}^n$ respectively, and $\gamma_{m,n}$ is given by:

$$\gamma_{m,n} := (e_{n,m}, \text{id}_{\mathbb{S}^n})_{\#} \mu_n. \quad (4)$$

We call $\gamma_{m,n} \in \mathcal{M}(\mu_m, \mu_n)$ from Claim 1 *the equatorial coupling*.

Remark 2.3 Since $\gamma_{m,n} \in \mathcal{M}(\mu_m, \mu_n)$, it follows trivially that

$$d_{\text{GW},p,q}(\mathbb{S}_{\bullet}^m, \mathbb{S}_{\bullet}^n) \leq \frac{1}{2} \text{dis}_{p,q}(\gamma_{m,n}, \mathbb{S}_{\bullet}^m, \mathbb{S}_{\bullet}^n).$$

Example 2.4 ($\text{dis}_{4,2}(\gamma_{0,1}, \mathbb{S}_G^0, \mathbb{S}_G^1)$) By Remark 1.18 and Example B.2, we have

$$\begin{aligned} & (\text{dis}_{4,2}(\gamma_{0,1}, \mathbb{S}_G^0, \mathbb{S}_G^1))^4 \\ &= (\text{diam}_4(\mathbb{S}_G^0))^4 + (\text{diam}_4(\mathbb{S}_G^1))^4 \\ &\quad - 2 \int (d_0(e_{1,0}(y), e_{1,0}(y')))^2 (d_1(y, y'))^2 \mu_1(dy) \mu_1(dy') \\ &= \frac{\pi^4}{2} + \frac{\pi^4}{5} - 2 \int_{\mathbb{S}^1} \int_{\mathbb{S}^1} (d_0(e_{1,0}(y), e_{1,0}(y')))^2 (d_1(y, y'))^2 \mu_1(dy) \mu_1(dy') \\ &= \frac{\pi^4}{2} + \frac{\pi^4}{5} - 2 \times \frac{\pi^4}{4} = \frac{\pi^4}{5}. \end{aligned}$$

where the value of the integral in the last line follows from the calculation in Example B.7. Hence

$$\text{dis}_{4,2}(\gamma_{0,1}, \mathbb{S}_G^0, \mathbb{S}_G^1) = \left(\frac{1}{5}\right)^{1/4} \pi \approx 2.101.$$

This implies that

$$d_{\text{GW},4,2}(\mathbb{S}_G^0, \mathbb{S}_G^1) \leq \frac{1}{2} \text{dis}_{4,2}(\gamma_{0,1}, \mathbb{S}_G^0, \mathbb{S}_G^1) \approx 1.050.$$

Example 2.5 One can carry out calculations analogous to those in the previous in the case of $\mathbb{S}_G^1$ and $\mathbb{S}_G^2$ to obtain

$$\begin{aligned} & (\text{dis}_{4,2}(\gamma_{1,2}, \mathbb{S}_G^1, \mathbb{S}_G^2))^4 \\ &= (\text{diam}_4(\mathbb{S}_G^1))^4 + (\text{diam}_4(\mathbb{S}_G^2))^4 \\ &\quad - 2 \int (d_1(e_{2,1}(y), e_{2,1}(y')))^2 (d_2(y, y'))^2 \mu_2(dy) \mu_2(dy') \\ &= \frac{\pi^4}{5} + 24 - 6\pi^2 + \frac{\pi^4}{2} \\ &\quad - 2 \int_{\mathbb{S}^1} \int_{\mathbb{S}^1} (d_1(e_{2,1}(y), e_{2,1}(y')))^2 (d_2(y, y'))^2 \mu_1(dy) \mu_1(dy') \\ &\approx \frac{\pi^4}{5} + 24 - 6\pi^2 + \frac{\pi^4}{2} - 2 \times 14.159 \approx 4.651 \end{aligned}$$

where we compute the integral in the last step using numerical integration along with the values of 4-diameters computed in Example B.2. This immediately implies that

$$d_{\text{GW}4,2}(\mathbb{S}_G^1, \mathbb{S}_G^2) \leq \frac{1}{2} \text{dis}_{4,2}(\gamma_{1,2}, \mathbb{S}_G^1, \mathbb{S}_G^2) \approx 0.734.$$

2.2 Exact Determination of $d_{\text{GW}4,2}(\mathbb{S}_E^m, \mathbb{S}_E^n)$

In this section, we establish that the equatorial coupling $\gamma_{m,n}$ is a minimizer of the $(4, 2)$ -distortion functional

$$\text{dis}_{4,2}(\cdot, \mathbb{S}_E^m, \mathbb{S}_E^n) : \mathcal{M}(\mu_m, \mu_n) \rightarrow \mathbb{R}_+$$

amongst all couplings between μ_m and μ_n . Our first result is the following lemma, which exactly computes the $(4, 2)$ -distortion of the equatorial coupling $\gamma_{m,n}$ for all m and n .

Lemma 2.6 *The $(4, 2)$ -distortion of the equatorial coupling between spheres $\mathbb{S}_E^m$ and $\mathbb{S}_E^n$ respectively equipped with their Euclidean distance and uniform measure with $n \geq m$ is*

$$\text{dis}_{4,2}(\gamma_{m,n}, \mathbb{S}_E^m, \mathbb{S}_E^n) = \left[\frac{4}{m+1} + \frac{4}{n+1} - \frac{8}{m+1} \left(\frac{\Gamma(\frac{m+2}{2}) \Gamma(\frac{n+1}{2})}{\Gamma(\frac{m+1}{2}) \Gamma(\frac{n+2}{2})} \right)^2 \right]^{1/4}.$$

We defer the proof of this lemma to Sect. A.3. The main result of this section establishes the optimality of the equatorial map:

Theorem 1 (Main Theorem). *The equatorial coupling $\gamma_{m,n}$ is an optimal coupling for $d_{\text{GW}4,2}(\mathbb{S}_E^m, \mathbb{S}_E^n)$. In particular, for $n \geq m \geq 0$,*

$$\begin{aligned} d_{\text{GW}4,2}(\mathbb{S}_E^m, \mathbb{S}_E^n) &= \frac{1}{2} \text{dis}_{4,2}(\gamma_{m,n}, \mathbb{S}_E^m, \mathbb{S}_E^n) \\ &= \frac{1}{\sqrt{2}} \left[\frac{1}{m+1} + \frac{1}{n+1} - \frac{2}{m+1} \left(\frac{\Gamma(\frac{m+2}{2}) \Gamma(\frac{n+1}{2})}{\Gamma(\frac{m+1}{2}) \Gamma(\frac{n+2}{2})} \right)^2 \right]^{1/4}. \end{aligned}$$

Remark 2.7 The fact that the equatorial coupling is optimal for $d_{\text{GW}4,2}$ is in sharp contrast to what takes place at the level of the closely related Gromov–Hausdorff distance, where cognates of the equatorial coupling are far from being optimal [27].

Remark 2.8 ($m = 0$) In the special case when $m = 0$, Theorem 1 implies:

$$d_{\text{GW}4,2}(\mathbb{S}_E^0, \mathbb{S}_E^n) = \frac{1}{\sqrt{2}} \left[\frac{n+2}{n+1} - \frac{2}{\pi} \left(\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n+2}{2})} \right)^2 \right]^{1/4}.$$

In particular,

$$d_{\text{GW4,2}}(\mathbb{S}_E^0, \mathbb{S}_E^1) = \frac{1}{\sqrt{2}} \left(\frac{3}{2} - \frac{8}{\pi^2} \right)^{1/4} \approx 0.644 \quad \text{and}$$

$$d_{\text{GW4,2}}(\mathbb{S}_E^0, \mathbb{S}_E^2) = \frac{1}{\sqrt{2}} \left(\frac{5}{6} \right)^{1/4} \approx 0.676.$$

Remark 2.9 ($n = m + 1$) In the interesting case when $n = m + 1$, Theorem 1 implies:

$$d_{\text{GW4,2}}(\mathbb{S}_E^m, \mathbb{S}_E^{m+1}) = \frac{1}{\sqrt{2}} \left[\frac{2m+3}{(m+1)(m+2)} - \left(\frac{2}{m+1} \right)^3 \left(\frac{\Gamma(\frac{m+2}{2})}{\Gamma(\frac{m+1}{2})} \right)^4 \right]^{1/4}.$$

In particular,

$$d_{\text{GW4,2}}(\mathbb{S}_E^0, \mathbb{S}_E^1) = \frac{1}{\sqrt{2}} \left(\frac{3}{2} - \frac{8}{\pi^2} \right)^{1/4} \approx 0.644,$$

$$d_{\text{GW4,2}}(\mathbb{S}_E^1, \mathbb{S}_E^2) = \frac{1}{\sqrt{2}} \left(\frac{5}{6} - \frac{\pi^2}{16} \right)^{1/4} \approx 0.482.$$

and

$$d_{\text{GW4,2}}(\mathbb{S}_E^2, \mathbb{S}_E^3) = \frac{1}{\sqrt{2}} \left(\frac{7}{12} - \frac{8}{27} \left(\frac{16}{\pi^2} \right) \right)^{1/4} \approx 0.400.$$

Remark 2.10 ($n = m + 2$). In some cases, it is possible to simplify the formula given by Theorem 1 to an explicit one. For example, when $n = m + 2$, the quantity in Theorem 1 simplifies to the following *explicit* formula:

$$d_{\text{GW4,2}}(\mathbb{S}_E^m, \mathbb{S}_E^{m+2}) = \frac{1}{2^{1/4}} \left[\frac{1}{(m+1)(m+3)} + \frac{1}{(m+2)^2(m+3)} \right]^{1/4}.$$

This implies that $d_{\text{GW4,2}}(\mathbb{S}_E^m, \mathbb{S}_E^{m+2}) = O(m^{-1/2})$ as $m \rightarrow \infty$. We compute some exact values below:

$$d_{\text{GW4,2}}(\mathbb{S}_E^0, \mathbb{S}_E^2) = \left(\frac{5}{24} \right)^{1/4} \approx 0.676; \quad d_{\text{GW4,2}}(\mathbb{S}_E^1, \mathbb{S}_E^3) = \left(\frac{11}{144} \right)^{1/4} \approx 0.526.$$

Remark 2.11 (Asymptotics for large m and n). It is clear from Theorem 1 that $d_{\text{GW4,2}}(\mathbb{S}_E^m, \mathbb{S}_E^n) \rightarrow 0$ as $m, n \rightarrow \infty$. More precisely, note that by resorting to the Stirling approximation, we have

$$\frac{\Gamma(\frac{m+2}{2})}{\Gamma(\frac{m+1}{2})} = \sqrt{\frac{m+1}{2}} (1 + O(m^{-1})); \quad \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n+2}{2})} = \sqrt{\frac{2}{n+1}} (1 + O(n^{-1})),$$

which implies that

$$\begin{aligned} d_{\text{GW4},2}(\mathbb{S}_E^m, \mathbb{S}_E^n) &= \frac{1}{\sqrt{2}} \left[\frac{1}{m+1} - \frac{1}{n+1} + O(m^{-2}) \right]^{1/4} \\ &= \frac{1}{\sqrt{2}} \left[\frac{n-m}{(m+1)(n+1)} + O(m^{-2}) \right]^{1/4}, \end{aligned}$$

as $m \rightarrow \infty$. Thus, if $n - m = O(1)$, we have $d_{\text{GW4},2}(\mathbb{S}_E^m, \mathbb{S}_E^n) = O(m^{-1/2})$ as $m \rightarrow \infty$.

Remark 2.12 (Asymptotics for fixed m , large n) As above, for large k , Stirling approximation yields

$$\frac{\Gamma\left(\frac{m+k+1}{2}\right)}{\Gamma\left(\frac{m+k+2}{2}\right)} = \sqrt{\frac{2}{m+k+1}} \left(1 + O(k^{-1})\right).$$

Theorem 1 then implies that

$$\begin{aligned} &(d_{\text{GW4},2}(\mathbb{S}_E^m, \mathbb{S}_E^{m+k}))^4 - \frac{1}{4(m+1)} \\ &= \frac{1}{4(m+k+1)} \left[1 - \frac{4}{m+1} \left(\frac{\Gamma\left(\frac{m+2}{2}\right)}{\Gamma\left(\frac{m+1}{2}\right)} \right)^2 \right] + O(k^{-2}). \end{aligned}$$

Thus, for a fixed m we have

$$d_{\text{GW4},2}(\mathbb{S}_E^m, \mathbb{S}_E^{m+k}) = \frac{1}{\sqrt{2}} \left[\frac{1}{m+1} \right]^{1/4} + O(k^{-1/4})$$

as $k \rightarrow \infty$.

2.3 The Proof of Theorem 1

We divide the proof into several steps.

2.3.1 Preliminaries

For $x, x' \in \mathbb{S}_E^m$ and $y, y' \in \mathbb{S}_E^n$ one has

$$\|x - x'\|^2 = 2(1 - \langle x, x' \rangle) \quad \text{and} \quad \|y - y'\|^2 = 2(1 - \langle y, y' \rangle).$$

Consider any coupling $\gamma \in \mathcal{M}(\mu_m, \mu_n)$. By the definition of $(4, 2)$ -distortion from Eq. (1.15), when d_X and d_Y are both the Euclidean distances, one has

$$\text{dis}_{4,2}^4(\gamma) = \int_{\mathbb{S}^m \times \mathbb{S}^n} \int_{\mathbb{S}^m \times \mathbb{S}^n} (\|x - x'\|^2 - \|y - y'\|^2)^2 \gamma(dx \times dy) \gamma(dx' \times dy')$$

$$\begin{aligned}
 &= \int_{\mathbb{S}^m \times \mathbb{S}^n} \int_{\mathbb{S}^m \times \mathbb{S}^n} |2\langle x, x' \rangle - 2\langle y, y' \rangle|^2 \gamma(dx \times dy) \gamma(dx' \times dy') \\
 &= 4 \iint_{\mathbb{S}^m \times \mathbb{S}^m} \langle x, x' \rangle^2 \mu_m(dx) \mu_m(dx') + 4 \iint_{\mathbb{S}^n \times \mathbb{S}^n} \langle y, y' \rangle^2 \mu_n(dy) \mu_n(dy') \\
 &\quad - 8 \int_{\mathbb{S}^m \times \mathbb{S}^n} \int_{\mathbb{S}^m \times \mathbb{S}^n} \langle x, x' \rangle \langle y, y' \rangle \gamma(dx \times dy) \gamma(dx' \times dy'). \quad (5)
 \end{aligned}$$

Thus for any coupling $\gamma \in \mathcal{M}(\mu_m, \mu_n)$ we have

$$\text{dis}_{4,2}^4(\gamma) = 4 \iint_{\mathbb{S}^m \times \mathbb{S}^m} \langle x, x' \rangle^2 \mu_m(dx) \mu_m(dx') + 4 \iint_{\mathbb{S}^n \times \mathbb{S}^n} \langle y, y' \rangle^2 \mu_n(dy) \mu_n(dy') - 8J(\gamma) \quad (6)$$

where we define

$$J(\gamma) := \int \int \langle x, x' \rangle \langle y, y' \rangle \gamma(dx \times dy) \gamma(dx' \times dy'). \quad (7)$$

Since the first two terms on the right hand side of Eq. (6) do not depend on the coupling γ , we have the following equivalent optimization problem:

$$\gamma \text{ minimizes } \text{dis}_{4,2}^4(\gamma) \Leftrightarrow \gamma \text{ maximizes } J(\gamma)$$

where both optimizations are over the space of couplings $\gamma \in \mathcal{M}(\mu_m, \mu_n)$. In the rest of the proof we therefore focus on *maximizing* $J(\gamma)$.

2.3.2 A Change of Coordinates

In this section we prove a lemma which permits simplifying the functional J defined above. See the discussion on page 7.

Lemma 2.13 *Let $\alpha \in \mathcal{P}(\mathbb{R}^{m+1})$ and $\beta \in \mathcal{P}(\mathbb{R}^{n+1})$ where $n \geq m$ be two rotationally invariant measures with barycenters coinciding with the respective origins. Consider the functional $J : \mathcal{M}(\alpha, \beta) \rightarrow \mathbb{R}$ defined above. Then,*

$$\max_{\gamma \in \mathcal{M}(\alpha, \beta)} J(\gamma) = \max_{\gamma \in \mathcal{M}(\alpha, \beta)} D(\gamma)$$

where

$$D(\gamma) := \sum_{k=1}^{m+1} \left[\int x_k y_k \gamma(dx \times dy) \right]^2.$$

Proof Applying the linearity of the integral and the identity $\text{trace}(AB) = \text{trace}(BA)$ for conformable matrices⁵ A, B , we compute,

$$\begin{aligned}
 J(\gamma) &= \int \int \langle x, x' \rangle \langle y, y' \rangle \gamma(dx \times dy) \gamma(dx' \times dy') \\
 &= \int \int \text{tr} \left[(x')^\top x y^\top y' \right] \gamma(dx \times dy) \gamma(dx' \times dy') \\
 &= \int \int \text{tr} \left[x y^\top y' (x')^\top \right] \gamma(dx \times dy) \gamma(dx' \times dy') \\
 &= \text{tr} \left[\int \int x y^\top (x' (y')^\top)^\top \gamma(dx \times dy) \gamma(dx' \times dy') \right] \\
 &= \text{tr} \left[\int x y^\top \gamma(dx \times dy) \left(\int x' (y')^\top \gamma(dx' \times dy') \right)^\top \right] \\
 &= \left\| \underbrace{\int x y^\top \gamma(dx \times dy)}_{M_\gamma} \right\|_F^2
 \end{aligned}$$

We now manipulate the matrix $M_\gamma \in \mathbb{R}^{(m+1) \times (n+1)}$ in order to simplify the optimization problem; see the discussion about related work on page 6. Hereon, we write $\mathbb{I}_d$ to mean the identity matrix of size d .

Consider the (possibly non-unique) *singular value decomposition*

$$M_\gamma = P_\gamma \Delta_\gamma Q_\gamma^\top$$

where

- $\Delta_\gamma \in \mathbb{R}^{(m+1) \times (m+1)}$ is a diagonal matrix containing the singular values of M_γ ,
- $P_\gamma \in \mathbb{R}^{(m+1) \times (m+1)}$ and $Q_\gamma \in \mathbb{R}^{(n+1) \times (m+1)}$ satisfy

$$P_\gamma^\top P_\gamma = P_\gamma P_\gamma^\top = \mathbb{I}_{m+1} \text{ and } Q_\gamma^\top Q_\gamma = \mathbb{I}_{m+1},$$

i.e. P_γ is orthonormal and Q_γ is semi-orthonormal.

We now define

$$U_\gamma := P_\gamma^\top \in \mathbb{R}^{(m+1) \times (m+1)} \text{ and } V_\gamma := (Q_\gamma \ Q_\gamma^\perp)^\top \in \mathbb{R}^{(n+1) \times (n+1)}$$

where $Q_\gamma^\perp \in \mathbb{R}^{(n+1) \times (n-m)}$ is any semi-orthonormal matrix, i.e., $(Q_\gamma^\perp)^\top Q_\gamma^\perp = \mathbb{I}_{n-m}$, which also satisfies $Q_\gamma^\top Q_\gamma^\perp = \mathbf{0}_{(m+1) \times (n-m)}$. Note that by construction U_γ and V_γ are orthonormal, i.e., $U_\gamma^\top U_\gamma = U_\gamma U_\gamma^\top = \mathbb{I}_{m+1}$ and $V_\gamma^\top V_\gamma = V_\gamma V_\gamma^\top = \mathbb{I}_{n+1}$.

Recall that the marginals of γ are μ_m and μ_n . Let $U_\gamma \in \mathbb{R}^{(m+1) \times (m+1)}$ and $V_\gamma \in \mathbb{R}^{(n+1) \times (n+1)}$ be the two orthonormal matrices defined above and T_{U_γ} and T_{V_γ} be the

⁵ I.e. the matrices can be multiplied.

linear maps they induce by left multiplication (i.e. $T_{U_\gamma} : \mathbb{R}^{m+1} \rightarrow \mathbb{R}^{m+1}$ is defined as $v \mapsto U_\gamma v$, with T_{V_γ} defined similarly). By the assumed symmetry of the measures α and β , we may pushforward γ through the associated maps and still obtain a coupling between α and β , that is:

$$(T_{U_\gamma}, T_{V_\gamma})_\# \gamma \in \mathcal{M}(\alpha, \beta).$$

Now, we define $\tilde{\gamma} := (T_{U_\gamma}, T_{V_\gamma})_\# \gamma$, and see that

$$\begin{aligned} M_{\tilde{\gamma}} &= \int xy^\top \tilde{\gamma}(dx \times dy) \\ &= U_\gamma \left(\int xy^\top \gamma(dx \times dy) \right) V_\gamma^\top \\ &= U_\gamma M_\gamma V_\gamma^\top \\ &= (\Delta_\gamma \mathbf{0}_{(m+1) \times (n-m)}). \end{aligned}$$

Since the Frobenius norm of a matrix is simply the Euclidean norm of its singular values, we have

$$J(\gamma) = \|M_\gamma\|_F^2 = \|\Delta_\gamma\|_F^2 = \|M_{\tilde{\gamma}}\|_F^2 = J(\tilde{\gamma}).$$

That is, for any optimal coupling γ , there exists another optimal coupling $\tilde{\gamma}$ for which $M_{\tilde{\gamma}}$ is of the form $(\Delta_\gamma \mathbf{0}_{(m+1) \times (n-m)})$ for a diagonal matrix $\Delta_\gamma \in \mathbb{R}^{(m+1) \times (m+1)}$.

We can then write:

$$\begin{aligned} J(\gamma) = J(\tilde{\gamma}) &= \|M_{\tilde{\gamma}}\|_F^2 = \sum_{k,l} \left[\int x_k y_l \tilde{\gamma}(dx \times dy) \right]^2 \\ &= \sum_{k=1}^{m+1} \left[\int x_k y_k \tilde{\gamma}(dx \times dy) \right]^2 \end{aligned}$$

where the last equality follows since we know that $M_{\tilde{\gamma}} = (\Delta_\gamma \mathbf{0}_{(m+1) \times (n-m)})$ for a diagonal matrix $\Delta_\gamma \in \mathbb{R}^{(m+1) \times (m+1)}$. Thus, maximizing $J(\gamma)$ is equivalent to:

$$\text{maximize } D(\gamma) = \sum_{k=1}^{m+1} \left[\int x_k y_k \gamma(dx \times dy) \right]^2 \text{ over } \gamma \in \mathcal{M}(\alpha, \beta). \quad (8)$$

That is, any optimizer of D is an optimizer of J and any optimizer of J can be pushed forward via a rotation to an optimizer of D . $\square$

2.3.3 Optimizing $J(\gamma)$ Over Couplings Between Standard Gaussians

In this section, we focus on optimizing $J(\gamma)$ for standard Gaussian marginals. Despite the close connections between the standard Gaussian measure and the uniform measure

on the sphere, there are fundamental differences among the two in terms of optimizing $J(\gamma)$ over all possible couplings. In particular, the techniques of this section will be, strictly speaking, applicable only to Gaussian measures and not to the uniform measures. The general method of proof will nevertheless pave the way for our proof of Theorem 1; see Remark 2.18.

We will use the notation η_d to denote the standard Gaussian measure on $\mathbb{R}^d$ (so that η_1 will denote the standard Gaussian measure on $\mathbb{R}$). In the usual notation for Normal distributions, η_d corresponds to $N(\mathbf{0}, \mathbb{I}_d)$.

Using the projection map $\pi_{n+1,m+1}$ from Definition 2.1, we define the following coupling between standard Gaussians:

Definition 2.14 For all $n \geq m$, the projection map $\pi_{n+1,m+1} : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{m+1}$ induces a coupling $\gamma_{m+1,n+1}^{\text{gauss}} \in \mathcal{M}(\eta_{m+1}, \eta_{n+1})$ given by:

$$\gamma_{m+1,n+1}^{\text{gauss}} := (\pi_{n+1,m+1}, \text{id}_{\mathbb{R}^{n+1}})_{\#} \eta_{n+1}.$$

Remark 2.15 Note that we can recover the equatorial coupling from Eq. (4) as follows $\gamma_{m,n} = (f_{m+1}, f_{n+1})_{\#} \gamma_{m+1,n+1}^{\text{gauss}}$ where $f_{m+1} : \mathbb{R}^{m+1} \setminus \{\mathbf{0}\} \rightarrow \mathbb{S}^m$ is the central projection map: $x \mapsto \frac{x}{\|x\|}$.

When the marginals of γ are standard Gaussian measures η_{m+1} and η_{n+1} , the optimization problem in Eq. (8) can be solved by relaxing the optimization into an optimization over the coordinate wise pushforwards of γ . This leads to the following proposition.

Proposition 2.16 Suppose $n \geq m$. Then,

$$\max_{\gamma \in \mathcal{M}(\eta_{m+1}, \eta_{n+1})} J(\gamma) = J(\gamma_{m+1,n+1}^{\text{gauss}}) = m + 1.$$

Remark 2.17 Note that [12, Proposition 4.1] gives a more general claim than Proposition 2.16 and consequently requires a much more sophisticated method of proof (cf. [12, Lemma 3.2]).

Proof of Proposition 2.16 By Lemma 2.13, we can equivalently maximize the functional $D(\gamma) = \sum_{k=1}^{m+1} \left[\int x_k y_k \gamma(dx \times dy) \right]^2$ over all couplings. To proceed with this, write

$$\sup_{\gamma \in \mathcal{M}(\eta_{m+1}, \eta_{n+1})} D(\gamma) \leq \sum_{k=1}^{m+1} \sup_{\gamma \in \mathcal{M}(\eta_{m+1}, \eta_{n+1})} \left[\int x_k y_k \gamma(dx \times dy) \right]^2 \quad (9)$$

$$= (m+1) \sup_{\gamma \in \mathcal{M}(\eta_{m+1}, \eta_{n+1})} \left[\int x_1 y_1 \gamma(dx \times dy) \right]^2. \quad (10)$$

Since the optimization on the right hand side depends only on x_1, y_1 , one can then optimize over the first coordinate pushforwards of γ . To be more precise, if

$\varphi_1^{m+1} : \mathbb{R}^{m+1} \rightarrow \mathbb{R}$ and $\varphi_1^{n+1} : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ denote the respective projections onto the first coordinate, consider $\gamma_1 := (\varphi_1^{m+1}, \varphi_1^{n+1})_{\#} \gamma$. It follows by Lemma 3.2 of [7] that

$$(\varphi_1^{m+1}, \varphi_1^{n+1})_{\#}(\mathcal{M}(\eta_{m+1}, \eta_{n+1})) = \mathcal{M}((\varphi_1^{m+1})_{\#}(\eta_{m+1}), (\varphi_1^{n+1})_{\#}(\eta_{n+1})) = \mathcal{M}(\eta_1, \eta_1),$$

where the last step follows from the fact that the pushforward through a (one dimensional) coordinate projection of the standard Gaussian in $\mathbb{R}^d$ is a standard Gaussian in $\mathbb{R}$. Thus, the optimization on the right hand side above can be equivalently written as

$$\begin{aligned} & \sup_{\gamma \in \mathcal{M}(\eta_{m+1}, \eta_{n+1})} \left[\int_{\mathbb{R}^{m+1} \times \mathbb{R}^{n+1}} x_1 y_1 \gamma(dx \times dy) \right]^2 \\ &= \sup_{\gamma_1 \in \mathcal{M}(\eta_1, \eta_1)} \left[\int_{\mathbb{R} \times \mathbb{R}} x_1 y_1 \gamma_1(dx_1 \times dy_1) \right]^2 \\ &\leq \left[\int_{\mathbb{R}} x_1^2 \eta_1(dx_1) \right] \left[\int_{\mathbb{R}} y_1^2 \eta_1(dy_1) \right] \\ &= \left[\int_{\mathbb{R}} z^2 \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) dz \right]^2 = 1. \end{aligned} \quad (11)$$

The first inequality follows by applying the Cauchy–Schwarz inequality. The last equality uses the well-known computation of the second moment of a one dimensional standard Gaussian measure. Plugging Eq. (11) into Eq. (10) shows that

$$\sup_{\gamma \in \mathcal{M}(\eta_{m+1}, \eta_{n+1})} D(\gamma) \leq (m+1). \quad (12)$$

To finish the proof of Proposition 2.16 we now note that

$$\begin{aligned} & \sup_{\gamma \in \mathcal{M}(\eta_{m+1}, \eta_{n+1})} D(\gamma) \geq D(\gamma_{m+1, n+1}^{\text{gauss}}) \\ &= \sum_{k=1}^{m+1} \left[\int_{\mathbb{R}^{m+1} \times \mathbb{R}^{n+1}} x_k y_k \gamma_{m+1, n+1}^{\text{gauss}}(dx \times dy) \right] \\ &= \sum_{k=1}^{m+1} \left[\int_{\mathbb{R}^{m+1}} x_k^2 \eta_m(dx) \right] \\ &= \sum_{k=1}^{m+1} \left[\int_{\mathbb{R}^{m+1}} x_k^2 \frac{1}{(2\pi)^{(m+1)/2}} \exp\left(-\sum_{k=1}^{m+1} x_k^2/2\right) dx_1 \dots dx_{m+1} \right] \\ &= \sum_{k=1}^{m+1} \left[\int_{\mathbb{R}} x_k^2 \frac{1}{\sqrt{2\pi}} \exp(-x_k^2/2) dx_k \right] = \sum_{k=1}^{m+1} 1 = (m+1). \end{aligned}$$

Here the first equality follows by definition of $D(\gamma)$, the second follows by the definition of $\gamma_{m+1, n+1}^{\text{gauss}}$, the third one is due to the definition $\eta_{m+1} := N(\mathbf{0}, \mathbb{I}_{m+1})$, the fourth

and the fifth inequalities follow by standard Gaussian integral computations. Together with Eq. (12) this finishes the proof of Proposition 2.16. $\square$

Remark 2.18 The strategy used in Proposition 2.16 for finding the optimal coupling for standard Gaussian marginals, does not, however, work for uniform measures on the spheres because the inequality in Eq. (9) is *not tight for uniform measures on $\mathbb{S}^m$* whenever $m > 0$. This is because whenever $m > 0$, the uniform measure on $\mathbb{S}^m$ is not a product measure over its coordinates—and thus the problem of maximizing $D(\gamma)$ cannot be solved by a coordinate wise approach: Optimizing in the first coordinate leads to constraints on the feasible set of the optimization in the second coordinate, and so on. Note that this is in contrast with the case where γ has standard Gaussian measures as marginals (considered in Sect. 2.3.3), which are indeed product distributions. See Question 3.

We now describe a variant of the approach used in the proof of Proposition 2.16 that applies to uniform measures on spheres and by taking into consideration the required dependence between the coordinates.

2.3.4 The Conclusion of the Proof of Theorem 1

Recall from the calculations in Sect. 2.3.1 that minimizing $\text{dis}_{4,2}$ over all couplings between the uniform measures μ_m and μ_n leads to maximizing the functional J defined in Eq. (7). Note that the uniform measures on spheres $\mu_m \in \mathcal{P}(\mathbb{S}^m)$ and $\mu_n \in \mathcal{P}(\mathbb{S}^n)$ are elements of $\mathcal{P}(\mathbb{R}^{m+1})$ and $\mathcal{P}(\mathbb{R}^{n+1})$, respectively. Hence, we can invoke Lemma 2.13 and equivalently maximize the functional D over all such couplings. Before tackling this, we need some preparations.

To simplify subsequent computations, for $y \in \mathbb{R}^{n+1}$ we define the projections $y_A \in \mathbb{R}^{m+1}$ and $y_B \in \mathbb{R}^{n-m}$ such that $y = (y_A, y_B)^\top$. Fixing γ , we use the projection just defined to introduce a decomposition of $\mathbb{S}^n$ as a *disjoint* union of products of spheres of smaller dimension. Indeed, let

$$A_t := \{y \in \mathbb{S}^n : \|y_A\| = t\} = (t \cdot \mathbb{S}^m) \times \left(\sqrt{1-t^2} \cdot \mathbb{S}^{n-m-1} \right), \quad t \in [0, 1]$$

so that

$$\mathbb{S}^n = \bigcup_{t=0}^1 A_t.$$

Let the measure $\nu \in \mathcal{P}([0, 1])$ be the pushforward of γ by $(x, y) \mapsto \|y_A\|$. Then, by the Disintegration Theorem [3, Theorem 5.3.1], there is a measure-valued map $t \mapsto \gamma_t$ from $[0, 1]$ to $\mathcal{P}(\mathbb{S}^m \times \mathbb{S}^n)$ such that:

- (1) $t \mapsto \gamma_t(B)$ is measurable for all Borel set $B \subseteq \mathbb{S}^m \times \mathbb{S}^n$,
- (2) $\gamma = \int_0^1 \gamma_t \nu(dt)$, and
- (3) $\text{supp}[\gamma_t] \subseteq A_t \times \mathbb{S}^m$ (so we will view γ_t as a probability measure on $A_t \times \mathbb{S}^m$ for each $t \in [0, 1]$).

Marginalizing this disintegration over its first factor, $\mathbb{S}^m$, we derive a disintegration of the uniform measure μ_n according to the map $y \mapsto \|y_A\|$, which we denote by $\overline{\gamma}_t$. This new marginal disintegration is, in particular, defined such that for all Borel subsets $B \subseteq \mathbb{S}^n$, $\overline{\gamma}_t(B) := \gamma_t(B \times \mathbb{S}^m)$. To check that this is indeed a disintegration, let $\varphi : \mathbb{S}^n \rightarrow \mathbb{R}$ be a measurable function, and then, since γ marginalizes to μ_n , we have

$$\begin{aligned} \int_{\mathbb{S}^n} \varphi(y) \mu_n(dy) &= \int_{\mathbb{S}^m \times \mathbb{S}^n} \varphi(y) \gamma(dx \times dy) \\ &= \int_0^1 \int_{\mathbb{S}^m \times A_t} \varphi(y) \gamma_t(dx \times dy) v(dt) \\ &= \int_0^1 \int_{A_t} \varphi(y) \overline{\gamma}_t(dy) v(dt). \end{aligned} \quad (13)$$

Since $\overline{\gamma}_t$ is a disintegration of μ_n , it has a symmetry informed by the symmetry of μ_n . In particular, for any $\varphi : \mathbb{S}^n \rightarrow \mathbb{R}$ and any $U \in O(m+1)$ and $V \in O(n-m)$,

$$\begin{aligned} &\int_0^1 \int_{A_t} \varphi(y_A, y_B) ((T_U, T_V)_\# \overline{\gamma}_t)(dy_A \times dy_B) v(dt) \\ &= \int_0^1 \int_{A_t} \varphi(Uy_A, Vy_B) \overline{\gamma}_t(dy_A \times dy_B) v(dt) \\ &= \int_{\mathbb{S}^n} \varphi(Uy_A, Vy_B) \mu_n(dy_A \times dy_B) \\ &= \int_{\mathbb{S}^n} \varphi(y_A, y_B) (T_U, T_V)_\# \mu_n(dy_A \times dy_B) \\ &= \int_{\mathbb{S}^n} \varphi(y_A, y_B) \mu_n(dy_A \times dy_B) \\ &= \int_0^1 \int_{A_t} \varphi(y_A, y_B) \overline{\gamma}_t(dy_A \times dy_B) v(dt). \end{aligned}$$

So, for any $U \in O(m+1)$ and $V \in O(n-m)$, $(T_U, T_V)_\# \overline{\gamma}_t = \overline{\gamma}_t$ for almost every t . The only probability measure on A_t that satisfies these conditions is the product of uniform measures on both factors y_A and y_B (a.e.). Marginalizing over y_B , we denote the induced measure on y_A as $\mu_{t, \mathbb{S}^m}$. By the above argument $\mu_{t, \mathbb{S}^m}$ is the uniform measure over $t \cdot \mathbb{S}^m$.

The disintegration described above allows for the computation:

$$\begin{aligned} D(\gamma) &= \sum_{k=1}^{m+1} \left[\int_{\mathbb{S}^m \times \mathbb{S}^n} x_k y_k \gamma(dx \times dy) \right]^2 \\ &= \sum_{k=1}^{m+1} \left[\int_0^1 t \cdot \int_{\mathbb{S}^m \times A_t} \frac{x_k y_k}{t} \gamma_t(dx \times dy) v(dt) \right]^2 \end{aligned}$$

$$\begin{aligned}
 &\leq \sum_{k=1}^{m+1} \left[\int_0^1 t \cdot \left(\int_{\mathbb{S}^m \times A_t} \frac{y_k^2}{t^2} \gamma_t(dx \times dy) \right)^{1/2} \cdot \left(\int_{\mathbb{S}^m \times A_t} x_k^2 \gamma_t(dx \times dy) \right)^{1/2} v(dt) \right]^2 \\
 &= \sum_{k=1}^{m+1} \left[\int_0^1 t \cdot \left(\int_{\mathbb{S}^m} y_k^2 \mu_m(dx) \right)^{1/2} \cdot \left(\int_{\mathbb{S}^m \times A_t} x_k^2 \gamma_t(dx \times dy) \right)^{1/2} v(dt) \right]^2 \\
 &= \frac{1}{m+1} \sum_{k=1}^{m+1} \left[\int_0^1 t \cdot \left(\int_{\mathbb{S}^m \times A_t} x_k^2 \gamma_t(dx \times dy) \right)^{1/2} v(dt) \right]^2.
 \end{aligned}
 \tag{*}$$

The inequality above is an application of the Cauchy–Schwarz inequality, and the subsequent equality is justified by the following computation:

$$\int_{\mathbb{S}^m \times A_t} \frac{y_k^2}{t^2} \mu_t(dx \times dy) = \int_{A_t} \frac{y_k^2}{t^2} \bar{\gamma}_t(dy) = \int_{t \cdot \mathbb{S}^m} \frac{y_k^2}{t^2} \mu_{t \cdot \mathbb{S}^m}(dy_A) = \int_{\mathbb{S}^m} \tilde{y}_k^2 \mu_m(d\tilde{y}),
 \tag{14}$$

where the first two equalities are given by integrating out x and y_B respectively. The final equality follows from a change of variables by the map $y \mapsto t^{-1} \cdot y$, using the fact that $\mu_{t \cdot \mathbb{S}^m}$ is a uniform measure on $t \cdot \mathbb{S}^m$.

In the last expression above, we would like to pass the summation inside the square in order to apply the Cauchy–Schwarz inequality, so we will write out the squared integral as a product of integrals over independent variables:

$$\begin{aligned}
 D(\gamma) &\leq \frac{1}{m+1} \sum_{k=1}^{m+1} \left[\int_0^1 t \cdot \left(\int_{\mathbb{S}^m \times A_t} x_k^2 \gamma_t(dx \times dy) \right)^{1/2} v(dt) \right]^2 \\
 &= \frac{1}{m+1} \sum_{k=1}^{m+1} \left[\int_0^1 t \cdot \left(\int_{\mathbb{S}^m \times A_t} x_k^2 \gamma_t(dx \times dy) \right)^{1/2} v(dt) \cdot \int_0^1 t' \cdot \left(\int_{\mathbb{S}^m \times A_{t'}} x_k'^2 \gamma_{t'}(dx' \times dy') \right)^{1/2} v(dt') \right] \\
 &= \frac{1}{m+1} \int_0^1 \int_0^1 t t' \cdot \sum_{k=1}^{m+1} \left(\int_{\mathbb{S}^m \times A_t} x_k^2 \gamma_t(dx \times dy) \right)^{1/2} \left(\int_{\mathbb{S}^m \times A_{t'}} x_k'^2 \gamma_{t'}(dx' \times dy') \right)^{1/2} v(dt') v(dt) \\
 &\leq \frac{1}{m+1} \int_0^1 \int_0^1 t t' \cdot \left(\sum_{k=1}^{m+1} \int_{\mathbb{S}^m \times A_t} x_k^2 \gamma_t(dx \times dy) \right)^{1/2} \left(\sum_{k=1}^{m+1} \int_{\mathbb{S}^m \times A_{t'}} x_k'^2 \gamma_{t'}(dx' \times dy') \right)^{1/2} v(dt') v(dt) \\
 &= \frac{1}{m+1} \int_0^1 \int_0^1 t t' \cdot \left(\int_{\mathbb{S}^m \times A_t} \gamma_t(dx \times dy) \right)^{1/2} \left(\int_{\mathbb{S}^m \times A_{t'}} \gamma_{t'}(dx' \times dy') \right)^{1/2} v(dt') v(dt) \\
 &= \frac{1}{m+1} \left(\int_{\mathbb{S}^m \times \mathbb{S}^n} \|y_A\| \gamma(dx \times dy) \right)^2 \\
 &= \frac{1}{m+1} \left(\int_{\mathbb{S}^n} \|y_A\| \mu_n(dy) \right)^2.
 \end{aligned}
 \tag{**}$$

We appeal to a well known characterization of μ_n to compute this integral. More precisely, if $Z_1, \dots, Z_{n+1}$ are independent $N(0, 1)$ distributed random variables, the law of

$$y = \frac{(Z_1, Z_2, \dots, Z_{n+1})}{(Z_1^2 + \dots + Z_{n+1}^2)^{1/2}}$$

is given by μ_n , which follows by the spherical symmetry of the $(n + 1)$ -dimensional standard Gaussian measure. Then by definition of y_A , we have

$$\|y_A\|^2 \stackrel{d}{=} \frac{Z_1^2 + Z_2^2 + \dots + Z_{m+1}^2}{Z_1^2 + Z_2^2 + \dots + Z_{n+1}^2} \sim \text{Beta}\left(\frac{m+1}{2}, \frac{n-m}{2}\right),$$

i.e., the Beta distribution with parameters $\frac{m+1}{2}$ and $\frac{n-m}{2}$. See, e.g., Theorem 5.8.4 and Section 8.2 of [11]. The “ $\stackrel{d}{=}$ ” symbol denotes an equality in distribution. Note that if $X \sim \text{Beta}(a, b)$, then

$$\begin{aligned} \mathbb{E}(\sqrt{X}) &= \frac{1}{\beta(a, b)} \int_0^1 \sqrt{x} x^{a-1} (1-x)^{b-1} dx = \frac{1}{\beta(a, b)} \int_0^1 x^{a-1/2} (1-x)^{b-1} dx \\ &= \frac{\beta(a+1/2, b)}{\beta(a, b)} \end{aligned}$$

where $\beta(a, b) := \int_0^1 x^{a-1} (1-x)^{b-1} dx$. Thus,

$$\int_{\mathbb{S}^n} \|y_A\| \mu_n(dx) = \frac{\beta((m+2)/2, (n-m)/2)}{\beta((m+1)/2, (n-m)/2)} = \frac{\Gamma\left(\frac{m+2}{2}\right) \Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{m+1}{2}\right) \Gamma\left(\frac{n+2}{2}\right)}.$$

We now return to compute the three terms on the right hand side of Eq. (5) one by one. First, by the spherical symmetry and the fact that μ_m is the uniform measure on $\mathbb{S}^m$, we observe that $x' \mapsto \int_{\mathbb{S}^m} \langle x, x' \rangle^2 \mu_m(dx)$ is constant for all $x' \in \mathbb{S}^m$. In particular,

$$\begin{aligned} \iint_{\mathbb{S}^m \times \mathbb{S}^m} \langle x, x' \rangle^2 \mu_m(dx) \mu_m(dx') &= \int_{\mathbb{S}^m} \langle x, e_1 \rangle^2 \mu_m(dx) \\ &= \int_{\mathbb{S}^m} \langle x, e_2 \rangle^2 \mu_m(dx) \\ &= \dots \\ &= \int_{\mathbb{S}^m} \langle x, e_{m+1} \rangle^2 \mu_m(dx). \end{aligned}$$

Hence

$$(m+1) \iint_{\mathbb{S}^m \times \mathbb{S}^m} \langle x, x' \rangle^2 \mu_m(dx) \mu_m(dx') = \sum_{i=1}^{m+1} \int_{\mathbb{S}^m} \langle x, e_i \rangle^2 \mu_m(dx) = \int_{\mathbb{S}^m} \sum_{i=1}^{m+1} \langle x, e_i \rangle^2 \mu_m(dx) = 1$$

Therefore,

$$\begin{aligned} \iint_{\mathbb{S}^m \times \mathbb{S}^m} \langle x, x' \rangle^2 \mu_m(dx) \mu_m(dx') &= \frac{1}{m+1} \text{ and} \\ \iint_{\mathbb{S}^n \times \mathbb{S}^n} \langle y, y' \rangle^2 \mu_n(dy) \mu_n(dy') &= \frac{1}{n+1}. \end{aligned}$$

It follows that any coupling $\gamma \in \mathcal{M}(\mu_m, \mu_n)$ satisfies

$$\begin{aligned} (\text{dis}_{4,2}(\gamma))^4 &= 4 \left(\frac{1}{m+1} + \frac{1}{n+1} \right) - 8J(\gamma) \\ &\leq 4 \left(\frac{1}{m+1} + \frac{1}{n+1} \right) - \frac{8}{m+1} \left(\int_{\mathbb{S}^n} \|y_A\| \mu_n(dx) \right)^2 \\ &= 4 \left(\frac{1}{m+1} + \frac{1}{n+1} \right) - \frac{8}{m+1} \left(\frac{\Gamma(\frac{m+2}{2}) \Gamma(\frac{n+1}{2})}{\Gamma(\frac{m+1}{2}) \Gamma(\frac{n+2}{2})} \right)^2 \\ &= (\text{dis}_{4,2}(\gamma_{m,n}, \mathbb{S}_E^m, \mathbb{S}_E^n))^4, \end{aligned}$$

via Lemma 2.6, thus showing that the equatorial map is optimal. $\square$

Remark 2.19 By analyzing the equality conditions for the Cauchy–Schwarz inequality in the above proof, one obtains a proof of Theorem 1 without relying on the explicit computation of Lemma 2.6. The equatorial coupling achieves equality in each inequality, so the computation of the bound of the distortion of γ is also a computation of the distortion of the equatorial coupling and a proof that it is optimal. The two inequalities occurring in the proof are:

- (1) (Cauchy–Schwarz in Eq. (*)): holds with equality if, conditional on $\|y_A\| = t$,

$$x_k = C_1 \frac{y_k}{\|y_A\|}$$

for a constant C_1 possibly dependent on $\|y_A\|$ but not on k .

- (2) (Cauchy–Schwarz in Eq. (**)): holds with equality if, for almost every t, t' , there exists a constant C_2 such that for all $k \leq n+1$:

$$\left(\int_{\mathbb{S}^m \times A_t} x_k^2 \gamma_t(dx \times dy) \right)^{1/2} = C_2 \left(\int_{\mathbb{S}^m \times A_{t'}} x_k^2 \gamma_{t'}(dx \times dy) \right)^{1/2}.$$

The equatorial map satisfies both of these conditions, so it is necessarily optimal.

3 General Lower Bounds

In this section, we will describe a number of different functions $\text{LB}_{p,q} : \mathcal{G}_w \times \mathcal{G}_w \rightarrow \mathbb{R}_+$ which will become lower bounds for the (p, q) -Gromov–Wasserstein distance.

Lower bounds for the p -Gromov–Wasserstein distance have been previously discussed [7, 30, 32]. In [30] three lower bounds for the p -Gromov–Wasserstein distance called the First, Second, and Third Lower Bounds (denoted FLB_p , SLB_p and TLB_p respectively) were constructed from certain invariants of metric measure spaces. Two of these lower bounds were based on global and local distributions of distances. SLB_p was constructed using the Wasserstein distance on the real line between global distributions of distances, and TLB_p was constructed using local distribution of distances.

In Sect. 3.2 we consider a generalization of the lower bound based on the p -diameter of a metric measure space introduced in [30]. Note that we do not consider a generalization of FLB_p introduced in [30], which is based on the p -eccentricity function associated to a metric measure space $X \in \mathcal{G}_w$ that assigns to each point in X a value reflecting a notion of average distance to all other points in the space. In Sects. 3.3 and 3.4, we construct $\text{TLB}_{p,q}$ and $\text{SLB}_{p,q}$ using the local distributions of distances and global distributions of distances respectively that depend on the parameter q . For the choice $q = 1$, our bounds $\text{TLB}_{p,q}$ and $\text{SLB}_{p,q}$ agree with TLB_p and SLB_p . Finally, in Proposition 3.10 we give a hierarchy of our lower bounds for the setting of the (p, q) -Gromov–Wasserstein distance.

3.1 Invariants

We first recall some invariants of metric measure spaces which we will utilize in our construction of lower bounds for the (p, q) -Gromov–Wasserstein distance.

Definition 3.1 (Global distribution of distances of a metric measure space). Let $(X, d_X, \mu_X) \in \mathcal{G}_w$. The *global distribution of distances* associated to X is the function,

$$H_X : [0, \text{diam}(X)] \rightarrow [0, 1] \text{ given by } t \mapsto \mu_X \otimes \mu_X(\{(x, x') \in X \times X \mid d_X(x, x') \leq t\}).$$

Definition 3.2 (Local distribution of distances of a metric measure space). Let $(X, d_X, \mu_X) \in \mathcal{G}_w$. The *local distribution of distances* associated to X is the function,

$$h_X : X \times [0, \text{diam}(X)] \rightarrow [0, 1] \text{ given by } (x, t) \mapsto \mu_X(\{x' \in X \mid d_X(x, x') \leq t\}).$$

Remark 3.3 It is described in [30, Remark 5.4] that all p -diameters of $(X, d_X, \mu_X) \in \mathcal{G}_w$ can be recovered from its global distribution of distances as follows:

$$\text{diam}_p(X) = m_p(dH_X) = \left(\int_0^\infty t^p dH_X(dt) \right)^{1/p}.$$

The local distribution of distances generalizes the global one and we can relate the global and local distributions of distances by noting that

$$h_X(x, t) = \mu_X(\overline{B_X(x, t)})$$

where $\overline{B_X(x, t)}$ is the closed ball centered at x with radius t . Then we have from [30, Remark 5.8], that

$$H_X(t) = \int_X \int_{\overline{B_X(x, t)}} \mu_X(dx') \mu_X(dx) = \int_X h_X(x, t) \mu_X(dx) \quad \text{for } t \in [0, \text{diam}(X)]. \quad (15)$$

Example 3.4 The global distance distribution function for $\mathbb{S}_G^n$ (for $n \geq 1$) is:

$$H_{\mathbb{S}_G^n}(t) = h_{\mathbb{S}_G^n}(t) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \int_0^t \sin^{n-1}(s) ds$$

where $t \in [0, \pi]$. This follows by the fact that μ_n is the uniform measure on $\mathbb{S}^n$ and basic spherical geometry; see e.g. [4, Chapter 1].

Example 3.5 From the previous example, and the fact that $\|x - x'\| = 2 \sin\left(\frac{d_n(x, x')}{2}\right)$ for all $x, x' \in \mathbb{S}^n \subset \mathbb{R}^{n+1}$, we obtain that the global distance distribution function for $\mathbb{S}_E^n$ is:

$$\begin{aligned} H_{\mathbb{S}_E^n}(t) &= \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \int_0^{2 \arcsin(t/2)} \sin^{n-1}(s) ds \\ &= \frac{2^n}{\sqrt{\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \int_0^{t/2} s^{n-1} (\sqrt{1-s^2})^{n-2} ds \end{aligned}$$

where $t \in [0, 2]$.

3.2 Diameter Lower Bound

Definition 3.6 ((p, q) -Diameter Lower Bound). The (p, q) -Diameter Lower Bound for $X, Y \in \mathcal{G}_w$, denoted $\text{DLB}_{p,q}$, for $p, q \in [1, \infty]$ is:

$$\text{DLB}_{p,q}(X, Y) := \Lambda_q(\text{diam}_p(X), \text{diam}_p(Y)) \stackrel{(*)}{=} \left| (\text{diam}_p(X))^q - (\text{diam}_p(Y))^q \right|^{1/q}$$

where $(*)$ holds when $q \in [1, \infty)$.

Remark 3.7 In general, by the triangle inequality for $d_{\text{GW},p,q}$ (cf. Theorem 2) and by Example 1.21, we always have $d_{\text{GW},p,q}(X, Y) \geq \frac{1}{2} |\text{diam}_p(X) - \text{diam}_p(Y)|$ for all $p, q \in [1, \infty]$. The lower bound $\text{DLB}_{p,q}$ depends on both p and q , and provides a better lower bound for $d_{\text{GW},p,q}$ since for all $q \geq 1$, it can be shown that $\text{DLB}_{p,q}(X, Y) \geq |\text{diam}_p(X) - \text{diam}_p(Y)|$.

3.3 Second Lower Bound

Here we consider a general lower bound for the (p, q) -Gromov–Wasserstein distance between two metric measure spaces X and Y based on the distribution of distances. For the case $q = 1$, it is known (see [30, Proposition 6.2]) that the p -Gromov–Wasserstein distance between X and Y is bounded below by the Wasserstein distance between the global distribution of distances of X and Y on the real line. We describe a function which we call (p, q) -Second Lower Bound, denoted $\text{SLB}_{p,q}$, which yields an analogue of this result for $q \geq 1$.

Definition 3.8 ((p, q) -Second Lower Bound). The (p, q) -Second Lower Bound for $X, Y \in \mathcal{G}_w$, denoted $\text{SLB}_{p,q}(X, Y)$, for $p, q \in [1, \infty]$, is:

$$\text{SLB}_{p,q}(X, Y) := d_{\mathbb{W}^p}^{(\mathbb{R}_+, \Lambda_q)}(dH_X, dH_Y).$$

For $X \in \mathcal{G}_w$, dH_X is the unique measure on $\mathbb{R}_+$ defined by $dH_X([a, b]) := H_X(b) - H_X(a)$ for all $a \leq b$. It can be checked that $dH_X = (d_X)_\#(\mu_X \otimes \mu_X)$.

Note that Remark 1.11 relates the $d_{\mathbb{W}^p}^{(\mathbb{R}_+, \Lambda_q)}$ distance to the usual Wasserstein distance between suitably transformed measures. The closed form solution of $d_{\mathbb{W}^p}^{(\mathbb{R}_+, \Lambda_q)}$ ensures that $\text{SLB}_{p,q}$ can be computed very efficiently.

3.4 Third Lower Bound

In analogy with the third lower bound from [30], we consider the local distribution of distances and construct what we call the (p, q) -Third Lower Bound, denoted $\text{TLB}_{p,q}$. For $X \in \mathcal{G}_w$, recall that to the local distribution of distances of X , $h_X(x, \cdot)$, we associate the unique measure on $\mathbb{R}_+$, $dh_X(x)$, where $dh_X(x) = (d_X(x, \cdot))_\# \mu_X$.

Definition 3.9 ((p, q) -Third Lower Bound). The (p, q) -Third Lower Bound, denoted $\text{TLB}_{p,q}$, for $p, q \in [1, \infty]$ and $X, Y \in \mathcal{G}_w$ is:

$$\text{TLB}_{p,q}(X, Y) := \inf_{\gamma \in \mathcal{M}(\mu_X, \mu_Y)} \left(\int_{X \times Y} (d_{\mathbb{W}^p}^{(\mathbb{R}_+, \Lambda_q)}(dh_X(x), dh_Y(y)))^p \gamma(dx \times dy) \right)^{1/p}$$

for $1 \leq p < \infty$, and

$$\text{TLB}_{\infty,q}(X, Y) := \inf_{\gamma \in \mathcal{M}(\mu_X, \mu_Y)} \sup_{(x,y) \in \text{supp}[\gamma]} d_{\mathbb{W}^\infty}^{(\mathbb{R}_+, \Lambda_q)}(dh_X(x), dh_Y(y)).$$

Note that the closed form solution of $d_{\mathbb{W}^p}^{(\mathbb{R}_+, \Lambda_q)}$ from Remark 1.11 allows one to efficiently compute the TLB.

3.5 The Complete Hierarchy of Lower Bounds

Hierarchies of lower bounds have been considered in [7, 29, 32]. A key aspect of [30] was providing a hierarchy between the aforementioned lower bounds, FLB_p , SLB_p and TLB_p , that showed $d_{\text{GW}p} \geq \text{TLB}_p \geq \text{FLB}_p$ and $d_{\text{GW}p} \geq \text{SLB}_p$. In [7] they considered lower bounds in the setting of Gromov–Wasserstein between networks. In particular, they considered the associated pushforwards of the First, Second, and Third Lower Bounds from [30] into the real line denoted, $\mathbb{R}\text{-FLB}_p$, $\mathbb{R}\text{-SLB}_p$ and $\mathbb{R}\text{-TLB}_p$ and showed that $\text{FLB}_p \geq \mathbb{R}\text{-FLB}_p$, $\text{SLB}_p \geq \mathbb{R}\text{-SLB}_p$, and $\text{TLB}_p \geq \mathbb{R}\text{-TLB}_p$.

We note here that [7] and [30] did not provide a complete hierarchy between their lower bounds, where by incomplete we mean only partial relationships between some of the bounds were given. Proposition 2.8 of [32] bridged this gap by giving a hierarchy of lower bounds that related the Third and Second Lower Bounds of [30] to one another and thus strengthened the original hierarchy results from [30] by showing $d_{\text{GW}p} \geq \text{TLB}_p \geq \text{SLB}_p \geq \text{FLB}_p$. Proposition 3.10 below generalizes [32, Proposition 2.8] to the setting of the (p, q) -Gromov–Wasserstein distance.

Proposition 3.10 *For all $X, Y \in \mathcal{G}_w$ and all $p, q \in [1, \infty]$ we have*

$$2d_{\text{GW}p,q}(X, Y) \geq \text{TLB}_{p,q}(X, Y) \geq \text{SLB}_{p,q}(X, Y) \geq \text{DLB}_{p,p \wedge q}(X, Y)$$

where $p \wedge q$ denotes $\min\{p, q\}$.

We defer the proof of this proposition to Sect. A.2. See also Example A.3.

3.6 Lower Bounds in the case of Spheres

In this section we consider the hierarchy of lower bounds for the Gromov–Wasserstein distance between spheres equipped with the geodesic distance and Euclidean distance. Let $\mathbb{S}_{\bullet}^m$ represent the m -sphere equipped with the geodesic or Euclidean metric (see Example 1.13).

3.6.1 Diameter Lower Bound for Spheres

Recall from Remark 3.3 that the p -diameter is related to the p -moment of the global distance distribution as follows:

$$\text{diam}_p(\mathbb{S}_{\bullet}^m) = m_p(d_{H_{\mathbb{S}_{\bullet}^m}}) = \left(\int_0^1 (H_{\mathbb{S}_{\bullet}^m}^{-1}(u))^p du \right)^{1/p},$$

where $H_{\mathbb{S}_{\bullet}^m}^{-1}$ is the generalized inverse of $H_{\mathbb{S}_{\bullet}^m}$ (see Example 1.5). The diameter lower bound then boils down to

$$\text{DLB}_{p,q}(\mathbb{S}_{\bullet}^m, \mathbb{S}_{\bullet}^n) = \Lambda_q(\text{diam}_p(\mathbb{S}_{\bullet}^m), \text{diam}_p(\mathbb{S}_{\bullet}^n)).$$

3.6.2 Second and Third Lower Bound for Spheres

The local distribution of distances is equal to the global distribution of distances in the case of spheres. This implies that the third lower bound is equal to the second lower bound, that is:

$$\begin{aligned} \text{TLB}_{p,q}(\mathbb{S}_{\bullet}^m, \mathbb{S}_{\bullet}^n) &= \inf_{\gamma \in \mathcal{M}(\mu_m, \mu_n)} \left(\int_{\mathbb{S}_{\bullet}^m \times \mathbb{S}_{\bullet}^n} \left(d_{\mathbb{W}^p}^{(\mathbb{R}_+, \Lambda_q)}(dh_{\mathbb{S}_{\bullet}^m}(x), dh_{\mathbb{S}_{\bullet}^n}(y)) \right)^p \gamma(dx \times dy) \right)^{1/p} \\ &= \left(d_{\mathbb{W}^p}^{(\mathbb{R}_+, \Lambda_q)}(dH_{\mathbb{S}_{\bullet}^m}, dH_{\mathbb{S}_{\bullet}^n}) \right)^{1/p} \\ &= \text{SLB}_{p,q}(\mathbb{S}_{\bullet}^m, \mathbb{S}_{\bullet}^n). \end{aligned}$$

It follows that we have the following hierarchy of lower bounds of $d_{\text{GW},p,q}$ for spheres when $p \geq q$:

$$2d_{\text{GW},p,q}(\mathbb{S}_{\bullet}^m, \mathbb{S}_{\bullet}^n) \geq \text{TLB}_{p,q}(\mathbb{S}_{\bullet}^m, \mathbb{S}_{\bullet}^n) = \text{SLB}_{p,q}(\mathbb{S}_{\bullet}^m, \mathbb{S}_{\bullet}^n) \geq \text{DLB}_{p,q}(\mathbb{S}_{\bullet}^m, \mathbb{S}_{\bullet}^n).$$

We can compute the second lower bound (equivalently the third lower bound) between $\mathbb{S}_{\bullet}^m$ and $\mathbb{S}_{\bullet}^n$ as follows:

$$\text{SLB}_{p,q}(\mathbb{S}_{\bullet}^m, \mathbb{S}_{\bullet}^n) = d_{\mathbb{W}^p}^{(\mathbb{R}_+, \Lambda_q)}(dH_{\mathbb{S}_{\bullet}^m}, dH_{\mathbb{S}_{\bullet}^n}) = \left(\int_0^1 |(H_{\mathbb{S}_{\bullet}^m}^{-1}(u))^q - (H_{\mathbb{S}_{\bullet}^n}^{-1}(u))^q|^{p/q} du \right)^{1/p}$$

where the second equality holds by Remark 1.10.

3.6.3 Lower Bounds for $d_{\text{GW},4,2}$ Between Spheres

We now provide example computations for lower bounds of the $(4, 2)$ -Gromov–Wasserstein distance between spheres of dimensions 0, 1 and 2, when equipped with the geodesic distance. We make use of the formulas from Sects. 3.6.1 and 3.6.2. To streamline the presentation, we defer the detailed calculations to Appendix B and only present the final values in Tables 1 and 2.

Remark 3.11 (Hierarchy of lower bounds and $d_{\text{GW},4,2}$ for $\mathbb{S}_G^0, \mathbb{S}_G^1, \mathbb{S}_G^2$) Following B.2 and B.3, the hierarchy of lower bounds for the $(4, 2)$ -Gromov–Wasserstein distance is given in Table 1.

Remark 3.12 (Hierarchy of lower bounds and $d_{\text{GW},4,2}$ for $\mathbb{S}_E^0, \mathbb{S}_E^1, \mathbb{S}_E^2$) Similarly to the case of $\mathbb{S}_G$, calculations for bounding $d_{\text{GW},4,2}$ can be carried out for spheres equipped with the Euclidean distance using the global distribution of distances for $\mathbb{S}_E^m$. These are given in Example 3.5.

In Sect. 2, we determined the exact value of the $(4, 2)$ -Gromov–Wasserstein distance between spheres equipped with the Euclidean distance. So, the values of the

Table 1 This table demonstrates the lower bound hierarchy from Proposition 3.10 in the case of the (4,2)-Gromov–Wasserstein distance between spheres equipped with the geodesic distance

Spheres	$\frac{1}{2}$ DLB _{4,2}	$\frac{1}{2}$ SLB _{4,2} = $\frac{1}{2}$ TLB _{4,2}	$d_{\text{GW } 4,2}$
$\mathbb{S}_G^0$ versus $\mathbb{S}_G^1$	0.801	0.918	≤ 1.050
$\mathbb{S}_G^1$ versus $\mathbb{S}_G^2$	0.431	0.461	≤ 0.734

The values of the lower bounds are computed in Sect. B.1, while the upper bounds of the (4,2)-Gromov–Wasserstein distances are computed using the equatorial coupling (see Claim 1 and Examples 2.4 and 2.5)

Table 2 This table demonstrates the lower bound hierarchy from Proposition 3.10 in the case of the (4,2)-Gromov–Wasserstein distance between spheres with the Euclidean metric

Spheres	$\frac{1}{2}$ DLB _{4,2}	$\frac{1}{2}$ SLB _{4,2} = $\frac{1}{2}$ TLB _{4,2}	$d_{\text{GW } 4,2}$
$\mathbb{S}_E^0$ versus $\mathbb{S}_E^1$	0.308	0.488	0.644
$\mathbb{S}_E^1$ versus $\mathbb{S}_E^2$	0.187	0.276	0.482

The values of the lower bounds are computed using the elements in Sect. B.2, while the exact values of the (4,2)-Gromov–Wasserstein distances are computed via Theorem 1 (see Remark 2.9)

lower bounds can be compared against the exact value of (4, 2)-Gromov–Wasserstein distance between Euclidean spheres as shown in Table 2. Note how the exact values significantly exceed those provided by the lower bounds.

4 Experimental Illustration

The explicit computations of the (4, 2)-Gromov–Wasserstein distance between Euclidean spheres provide a helpful tool for benchmarking common optimal transport solvers and packages. The goal of this section is to benchmark various sampling methods and the number of samples required to obtain accurate estimates of the Gromov–Wasserstein distance while also ascertaining the accuracy of the various solvers in relation to the exact values provided by Theorem 1.

The authors are aware of two Python implementations of optimal transport GW solvers: Python Optimal Transport (POT) [18] and Optimal Transport Tools (OTT) [10]. These packages implement two of the most common methods for computing the Gromov–Wasserstein distance: Conditional Gradient Descent (implemented by POT) [51] and Sinkhorn Projections with entropic regularization (implemented by both OTT and POT) [40, 41].

In our experiments we used the Conditional Gradient Descent (CGD) solver from POT and the Sinkhorn solver from OTT (with regularization parameter 0.01).⁶ All experiments and their results are available in the Github repository [39].

⁶ The reason we used the Sinkhorn solver from OTT instead of the one from POT is that the former appears to be faster for “smaller scale problems” according to their documentation: https://ott-jax.readthedocs.io/en/latest/tutorials/OTT_%26_POT.html.

We ran two types of experiments. First, we examine how the number of samples relates to the choice of: the solver, the subsampling method, and the weights. Second, we fix the number of samples and vary the dimension of the spheres. The former are described next whereas the latter results are presented in Appendix C.

Experiment with Varying Number of Sample Points

In this experiment, we fix the dimensions of both spheres to the values $(m, n) = (1, 2), (1, 3)$ and $(2, 3)$ and vary the number of samples we draw from each of them. For each number of sample points between 10 and 200 (in increments of 10), we run 20 trials of each combination of sampling method, weight procedure, and GW solver (see below). The maximal size 200 was chosen so as to maintain a reasonable computational burden.

See Figs. 2, 3, and 4 where the label POT is used to indicate the CGD solver and OTT is used to indicate the Sinkhorn solver. The plotted lines are the mean values estimated from the 20 trials, while the shaded areas represent the central 80% of the samples. Dotted lines correspond to the “true” values established by Theorem 1:

$$(1,2): d_{\text{GW}4,2}(\mathbb{S}_E^1, \mathbb{S}_E^2) = \frac{1}{\sqrt{2}} \left(\frac{5}{6} - \frac{\pi^2}{16} \right)^{1/4} \approx \mathbf{0.482}; \text{ see Remark 2.9.}$$

$$(1,3): d_{\text{GW}4,2}(\mathbb{S}_E^1, \mathbb{S}_E^3) = \left(\frac{11}{144} \right)^{1/4} \approx \mathbf{0.526}; \text{ see Remark 2.10.}$$

$$(2,3): d_{\text{GW}4,2}(\mathbb{S}_E^2, \mathbb{S}_E^3) = \frac{1}{\sqrt{2}} \left(\frac{7}{12} - \frac{8}{27} \left(\frac{16}{\pi^2} \right) \right)^{1/4} \approx \mathbf{0.400}; \text{ see Remark 2.9.}$$

We implemented two sampling strategies:

- (1) **Random:** We draw the desired number of uniform samples via the well known method from [38], which consists of normalizing standard Gaussian samples.
- (2) **Farthest Point Sampling (FPS):** We first sample 10^6 points from the sphere uniformly at random via [38] and we then select the desired number of subsamples via the FPS method [16, 20].⁷

We implement two different procedures for assigning weights to the samples. Given a finite sample $P \subset \mathbb{S}^m$:

- (1) **Voronoi:** This consists on assigning to each point $p \in P$ an estimate of the total mass of the Voronoi cell on the sphere corresponding to p . To estimate this, we construct a set S consisting of 10^6 uniformly sampled points on the sphere and assign to p the proportion of points from S that are closer to it than to any other point in P .
- (2) **Uniform:** We simply give uniform weights $|P|^{-1}$ to all points $p \in P$.

⁷ In a nutshell, given a finite metric space (X, d_X) and a positive integer $N \geq 2$, the FPS method selects the first as a random point from X . The second point will be any point at maximal distance from the first selected point. The third point will be any point at maximal distance from the first two points and so on.

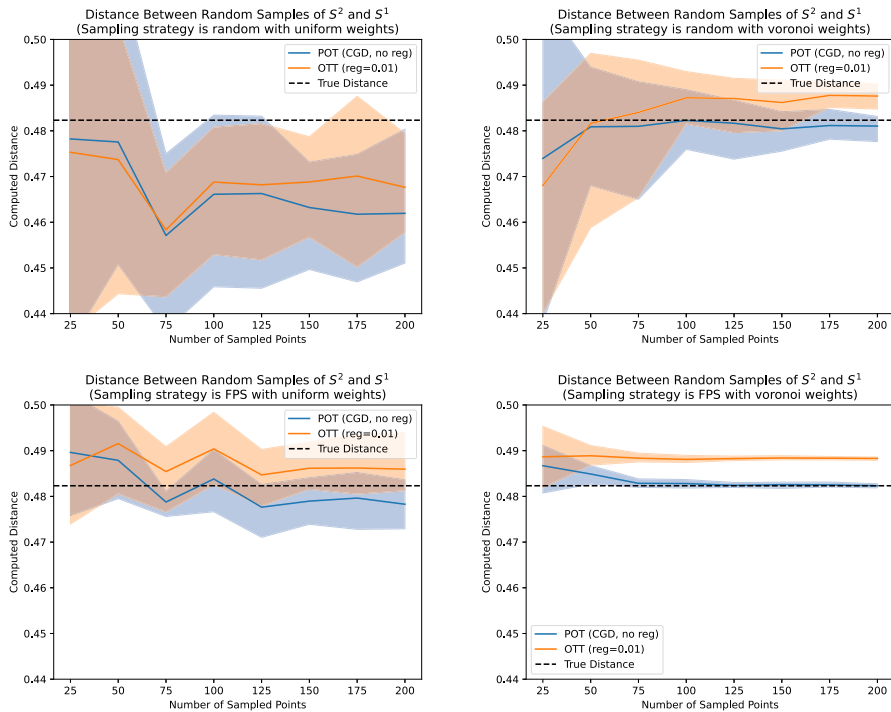


Fig. 2 Estimating the Gromov–Wasserstein distance between S^2_E and S^1_E

Observations

Below it will be convenient to refer to the different combinations of procedures and solvers via the specification of the triple (Sampling, Weights, Solver) where Sampling $\in \{\text{Random, FPS}\}$, Weights $\in \{\text{Uniform, Voronoi}\}$ and Solver $\in \{\text{POT, OTT}\}$.

Figures 2, 3, and 4 suggest the following observations:

- In all figures, the general trend is that FPS sampling outperforms Random sampling and that Voronoi weights outperform Uniform weights. This is expected as Voronoi weights are known to be optimal in the sense of quantization of measures [28, Lemma D.6], and FPS sampling is expected to provide a quasi-optimal sampling of a metric space [20, 34].
- The combination (FPS, Voronoi, POT) produced the best results in all cases. In the case S^1 versus S^2 and in the case S^1 versus S^3 it provided excellent results equal to 200 points. In those cases, the Sinkhorn solver (OTT) exhibited some bias, as is expected from the fact that it uses entropic regularization [42]. The case of S^2 versus S^3 suggests that a dense sampling might be necessary to approach the true value of the distance.
- For the case S^1 vs S^2 it is remarkable that with few samples the plots of (Random, Voronoi, POT), (FPS, Uniform, POT), (FPS, Voronoi, POT) are already quite close to the true distance value. This especially is the case for (FPS, Voronoi, POT).

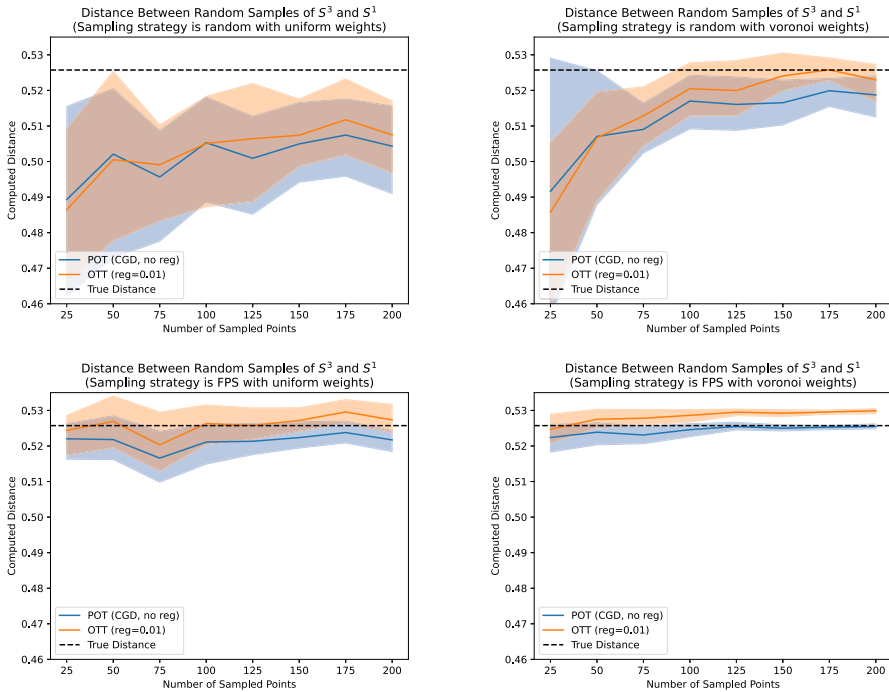


Fig. 3 Estimating the Gromov–Wasserstein distance between S_E^3 and S_E^1

- In the case S^2 versus S^3 , in all likelihood due to the fact that a sample size of 200 is expected to be insufficient to effectively represent S^3 , most combinations exhibited some degree of error.

5 Conclusions and Perspectives

Our results provide one additional⁸ infinite class of metric measure spaces for which we know the exact value of the Gromov–Wasserstein distance. Besides their intrinsic theoretical interest, our results also provide a benchmark against which the standard solvers for the Gromov–Wasserstein distance can be compared.

We now collect a number of questions.

Some Questions

The fact that we have considered an extra parameter q in our construction of the (p, q) -Gromov–Wasserstein distance together with the fact that it is known that for $p = q = \infty$, the resulting distance admits a polynomial time algorithm [31] suggest posing the following question.

⁸ Besides the class induced by Gaussian measures studied in [12].

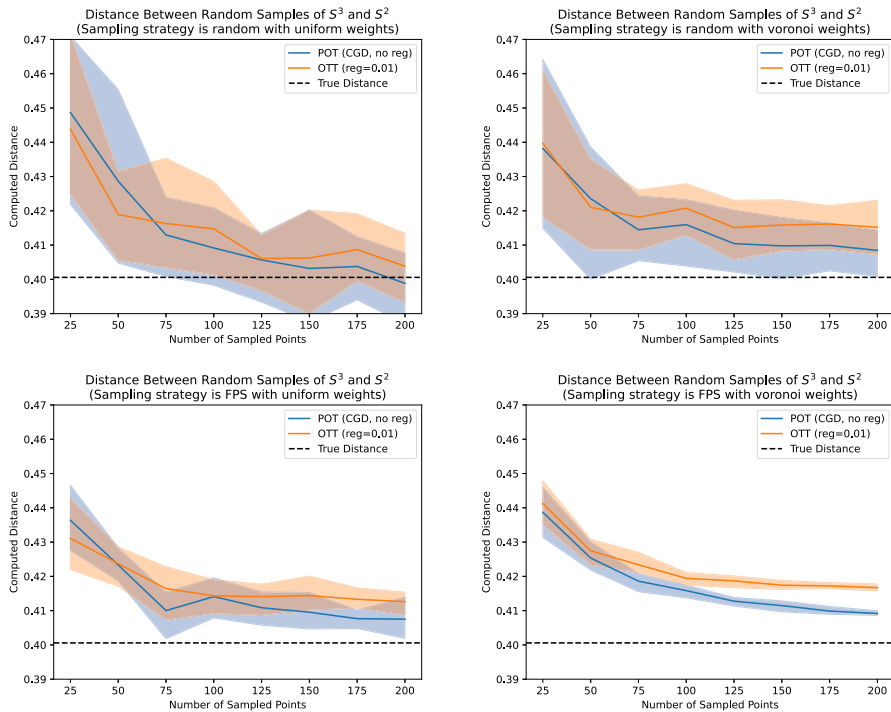


Fig. 4 Estimating the Gromov–Wasserstein distance between $\mathbb{S}_E^3$ and $\mathbb{S}_E^2$

Question 1 Are there classes $\mathcal{C} \subset \mathcal{G}_w$ of metric measures spaces (or networks, as in [7]) and particular choices of p, q , such that there exists a polynomial time algorithm for computing $d_{\text{GW}, p, q}(X, Y)$ for $X, Y \in \mathcal{C}$?

In light of Theorem 1 and Remark 2.7, it is natural to ask at what level of generality the equatorial coupling is optimal.

Question 2 In particular, we would like to know:

- Is the equatorial coupling optimal for $d_{\text{GW}, p, q}(\mathbb{S}_E^m, \mathbb{S}_E^n)$ for values of p and q other than $(p, q) = (4, 2)$?
- Are there values of p, q so that the equatorial coupling is optimal for the (p, q) -Gromov–Wasserstein distance between spheres with their geodesic distances (as opposed to their Euclidean distances) $d_{\text{GW}, p, q}(\mathbb{S}_G^m, \mathbb{S}_G^n)$?

Since the uniform measure μ_m (resp. μ_n) on $\mathbb{S}^m$ (resp. $\mathbb{S}^n$) can be obtained as the pushforward of the standard Gaussian measure on $\mathbb{R}^{m+1}$ (resp. $\mathbb{R}^{n+1}$) under the central projection map and since, by Remark 2.15, the equatorial coupling can be analogously recovered from $\gamma_{m+1, n+1}^{\text{gauss}}$, one may ask:

Question 3 Can one directly invoke [12, Proposition 4.1] or Proposition 2.16, establishing the optimality of the coupling $\gamma_{m+1, n+1}^{\text{gauss}}$ for (a certain variant of) the

Gromov–Wasserstein distance between Gaussian measures, to obtain a different proof of Theorem 1?

Finally we note that in [15, Theorems 3.2 and 3.6], the authors show the existence of a Monge map that induces an optimal coupling for both the “inner product” Gromov–Wasserstein distance and the quadratic (i.e. $p = 2$) Gromov–Wasserstein distance between two measures $\nu_m \in \mathcal{P}(\mathbb{R}^{m+1})$ and $\nu_n \in \mathcal{P}(\mathbb{R}^{n+1})$ with $n \geq m$ and ν_n absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{n+1}$. For example, in the setting of [15, Theorem 3.2], which is the closest to ours, they find that there exists an optimal coupling arising through a Monge map which is the gradient of a convex function (in a manner similar to the celebrated Brenier’s theorem in optimal transport). As we pointed out on page 7, their results do not apply in our setting, since the uniform distribution on a sphere $\mathbb{S}_E^d$ is singular with respect to the Lebesgue measure in $\mathbb{R}^{d+1}$. However, in Theorem 1, we found that an optimal coupling for the uniform distribution between spheres is in fact generated by the Monge map $e_{n,m}$ (see Definition 2.1 and Claim 1). Interestingly, $e_{n,m} : \mathbb{S}^n \rightarrow \mathbb{S}^m$ can be written as $e_{n,m} = T_0(\pi_{n,m}(x))$, where $T_0 : \mathbb{R}^{m+1} \rightarrow \mathbb{S}^m$ is the gradient of the convex function $g : \mathbb{R}^{m+1} \rightarrow \mathbb{R}$ defined as $g(y) := (y_1^2 + \cdots + y_{m+1}^2)^{1/2}$.

This leads us to the following question:

Question 4 Do the conclusions of Theorem 3.2 and 3.6 from [15], i.e., the existence of Monge maps that minimize the Gromov–Wasserstein distance between two metric measure spaces, hold for more general classes of measures?

Acknowledgements This project was part of the 2022 AMS-MRC (Math Research Communities) on “Data Science at the Crossroads of Analysis, Geometry, and Topology”. S.A. and A.A. acknowledge support from NSF through grant DMS-1916439. R.A.C. acknowledges support from NSF MPS Ascending Fellowship Award #2138110. F.M. was supported by the NSF through Grants DMS-1723003, IIS-1901360, CCF-1740761, DMS-1547357, CCF-2310412 and by the BSF under grant 2020124.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A Relegated Proofs

A.1 Proof of Theorem 2

Lemma A.1 Let $(X, d_X, \mu_X), (Y, d_Y, \mu_Y) \in \mathcal{G}_w$ be fixed and let $\gamma \in \mathcal{M}(\mu_X, \mu_Y)$. Then

$$\text{dis}_{p,q}(\gamma) \leq \text{dis}_{p',q'}(\gamma)$$

for all $1 \leq p \leq p' \leq \infty$ and $1 \leq q \leq q' \leq \infty$.

Proof First of all, observe that from the fact that $p' \geq p$ it follows that

$$\text{dis}_{p,q}(\gamma) = \|\Lambda_q(d_X, d_Y)\|_{L^p(\gamma \otimes \gamma)} \leq \|\Lambda_q(d_X, d_Y)\|_{L^{p'}(\gamma \otimes \gamma)} = \text{dis}_{p',q}(\gamma).$$

Also, note that

$$\begin{aligned} (\text{dis}_{p',q}(\gamma))^{p'} &= \int_{X \times Y} \int_{X \times Y} \left(\Lambda_q(d_X(x, x'), d_Y(y, y')) \right)^{p'} \gamma(dx \times dy) \gamma(dx' \times dy') \\ &\leq \int_{X \times Y} \int_{X \times Y} \left(\Lambda_{q'}(d_X(x, x'), d_Y(y, y')) \right)^{p'} \gamma(dx \times dy) \gamma(dx' \times dy') \\ &= (\text{dis}_{p',q'}(\gamma))^{p'} \end{aligned}$$

where the inequality in the second line holds since $\Lambda_q \leq \Lambda_{q'}$ by Proposition 1.9. Combining the above, it follows that $\text{dis}_{p,q}(\gamma) \leq \text{dis}_{p',q}(\gamma) \leq \text{dis}_{p',q'}(\gamma)$ which proves the lemma. $\square$

Proof of Theorem 2 The claim that $d_{\text{GW}p,q} \leq d_{\text{GW}p',q'}$ for all $1 \leq p \leq p' \leq \infty$ and $1 \leq q \leq q' \leq \infty$ follows immediately from Lemma A.1 above.

We now prove that $d_{\text{GW}p,q}(X, Y) = 0$ implies that $X \cong Y$. Suppose $d_{\text{GW}p,q}(X, Y) = 0$. Then,

$$0 = d_{\text{GW}p,q}(X, Y) \geq d_{\text{GW}p,1}(X, Y) = d_{\text{GW}p}(X, Y) \geq 0$$

where the first inequality follows from monotonicity of $d_{\text{GW}p,q}$. By [30, Theorem 5.1 (a)] $d_{\text{GW}p,q}(X, Y) = 0$ implies that $X \cong Y$.

Finally, we establish the triangle inequality for $d_{\text{GW}p,q}$ as follows. Fix arbitrary (X, d_X, μ_X) , (Y, d_Y, μ_Y) , and (Z, d_Z, μ_Z) in $\mathcal{G}_w$. Let $\varepsilon > 0$ be an arbitrary real number. Then, one can choose couplings $\mu_{XZ} \in \mathcal{M}(\mu_X, \mu_Z)$ and $\mu_{ZY} \in \mathcal{M}(\mu_Z, \mu_Y)$ such that

$$\frac{1}{2} \|\Lambda_q(d_X, d_Z)\|_{L^p(\mu_{XZ} \otimes \mu_{XZ})} < d_{\text{GW}p,q}(X, Z) + \varepsilon$$

and

$$\frac{1}{2} \|\Lambda_q(d_Z, d_Y)\|_{L^p(\mu_{ZY} \otimes \mu_{ZY})} < d_{\text{GW}p,q}(Z, Y) + \varepsilon.$$

Next, by the gluing lemma [54, Lemma 7.6], there exists a probability measure ω on $X \times Z \times Y$ such that $(\pi_{XZ})_{\#}\omega = \mu_{XZ}$ and $(\pi_{ZY})_{\#}\omega = \mu_{ZY}$ where $\pi_{XZ} : X \times Z \times Y \rightarrow X \times Z$ and $\pi_{ZY} : X \times Z \times Y \rightarrow Z \times Y$ are the canonical projections. Now, let $\mu_{XY} := (\pi_{XY})_{\#}\omega$. Then,

$$d_{\text{GW}p,q}(X, Y) \leq \frac{1}{2} \|\Lambda_q(d_X, d_Y)\|_{L^p(\mu_{XY} \otimes \mu_{XY})} = \frac{1}{2} \|\Lambda_q(d_X, d_Y)\|_{L^p(\omega \otimes \omega)}$$

$$\begin{aligned}
 &\leq \frac{1}{2} (||\Lambda_q(d_X, d_Z)||_{L^p(\omega \otimes \omega)} + ||\Lambda_q(d_Z, d_Y)||_{L^p(\omega \otimes \omega)}) \\
 &= \frac{1}{2} ||\Lambda_q(d_X, d_Z)||_{L^p(\mu_{XZ} \otimes \mu_{XZ})} + \frac{1}{2} ||\Lambda_q(d_Z, d_Y)||_{L^p(\mu_{ZY} \otimes \mu_{ZY})} \\
 &= d_{\text{GW}_{p,q}}(X, Z) + d_{\text{GW}_{p,q}}(Z, Y) + 2\varepsilon.
 \end{aligned}$$

The second inequality follows from an application of the triangle inequality for Λ_q :

$$\Lambda_q(d_X(x, x'), d_Y(y, y')) \leq \Lambda_q(d_X(x, x'), d_Z(z, z')) + \Lambda_q(d_Z(z, z'), d_Y(y, y'))$$

for $(\omega \otimes \omega)$ -a.e. (x, x', y, y', z, z') . This is possible since Λ_q is a metric on $\mathbb{R}_+$ by Proposition 1.9. Since the choice of ε is arbitrary, one can establish the required triangle inequality. $\square$

Remark A.2 Since $\Delta_{p/q,q}$ is a metric on the collection of isomorphism classes of $\mathcal{G}_w$ whenever $p \geq q$ (see [48, Corollary 9.3]) and by Remark 1.22 $d_{\text{GW}_{p,q}}$ is the q -snowflake transform of $\Delta_{p/q,q}$ multiplied by the constant $2^{-1/q}$, one can conclude that $d_{\text{GW}_{p,q}}$ is also a metric on the collection of isomorphism classes of $\mathcal{G}_w$ for $p \geq q$. This provides an alternative proof of a claim in Theorem 2 for the case when $p \geq q$. We note that our statement in Theorem 2 and the proof above do not have this restriction.

A.2 Proof of Proposition 3.10 and an Example

Example A.3 We now provide an example showing that $\text{DLB}_{p,q}$ is not always a lower bound for $d_{\text{GW}_{p,q}}$ in the case where $p < q$. We set $p = 1$, and, for some $\alpha \in [1/2, 1]$,

$$\begin{aligned}
 X &= \{x_1, x_2\}, \text{ with } d_X(x_1, x_2) = 1 \text{ and } \mu_X(\{x_1\}) = \alpha, \mu_X(\{x_2\}) = 1 - \alpha, \\
 Y &= \{y_1, y_2\}, \text{ with } d_Y(y_1, y_2) = 1 \text{ and } \mu_Y(\{y_1\}) = 1/2, \mu_Y(\{y_2\}) = 1/2.
 \end{aligned}$$

Then,

$$\text{DLB}_{1,q}(X, Y) = |(\text{diam}_1(X))^q - (\text{diam}_1(Y))^q|^{1/q} = |(2\alpha(1 - \alpha))^q - (1/2)^q|^{1/q}.$$

On the other hand,

$$\begin{aligned}
 &2d_{\text{GW}_{1,q}} \\
 &\leq \text{dis}_{1,q}(\gamma) \\
 &= 2(\gamma(x_1, y_1) \cdot \gamma(x_1, y_2) + \gamma(x_2, y_1) \cdot \gamma(x_2, y_2) + \gamma(x_1, y_1) \cdot \gamma(x_2, y_1) \\
 &\quad + \gamma(x_1, y_2) \cdot \gamma(x_2, y_2)) \\
 &= 2(\alpha - 1/2)(1 - \alpha) + (\alpha - 1/2) = 4\alpha - 2\alpha^2 - 3/2
 \end{aligned}$$

where γ is the coupling between μ_X and μ_Y described in Fig. 5. Selecting $q = 4$, and $\alpha = 3/4$ (in fact this is a counterexample for any $q > 2.5$ and any $\alpha \in (1/2, 1)$), these

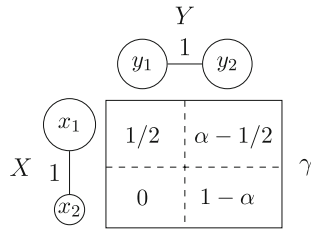


Fig. 5 We have $\mu_X(\{x_1\}) = \alpha$, $\mu_X(\{x_2\}) = 1 - \alpha$, $\mu_Y(\{y_1\}) = \mu_Y(\{y_2\}) = 1/2$ for some $\alpha \in [1/2, 1]$. In this scenario, we construct an example where $\text{DLB}_{p,q}$ is not a lower bound for $d_{\text{GW}_{p,q}}$ when $p < q$. The (p, q) distortion under the coupling γ , illustrated in the square, is used to derive an upper bound on $2d_{\text{GW}_{p,q}}(X, Y)$. See Example A.3 for more details

evaluate to:

$$\begin{aligned} \text{DLB}_{1,4}(X, Y) &= \left| (3/8)^4 - (1/2)^4 \right|^{1/4} \approx 0.45 \\ 2d_{\text{GW}_{1,4}}(X, Y) &\leq 3 - 18/16 - 3/2 = 3/8 = 0.375. \end{aligned}$$

Thus, the diameter lower bound does not hold in general when $p < q$.

We will need the following lemmas to prove Proposition 3.10.

Lemma A.4 ([30, Remark 5.8]) Suppose $X \in \mathcal{G}_w$ is given. Let $dh_X(x)$ be the unique probability measure on $\mathbb{R}$ associated to the local distributions of distances $h_X(x, \cdot)$ and dH_X be the unique probability measure associated to the global distribution of distances H_X . Then, we have the following:

$$\int_X dh_X(x) \mu_X(dx) = dH_X.$$

Lemma A.5 Suppose $X, Y \in \mathcal{G}_w$ and $p, q \in [1, \infty)$ are given. Then, there is a measure-valued map $(x, y) \mapsto \nu_{x,y}$ from $X \times Y$ to $\mathcal{P}(\mathbb{R}_+ \times \mathbb{R}_+)$ such that

- (1) $(x, y) \mapsto \nu_{x,y}(B)$ is measurable for every Borel set $B \subseteq \mathbb{R}_+ \times \mathbb{R}_+$,
- (2) $\nu_{x,y}$ belongs to $\mathcal{M}(dh_X(x), dh_Y(y))$ for each $(x, y) \in X \times Y$, and
- (3) $d_{\text{W}_p}^{(\mathbb{R}, \Lambda_q)}(dh_X(x), dh_Y(y)) = \left(\int_{\mathbb{R}_+ \times \mathbb{R}_+} (\Lambda_q(a, b))^p \nu_{x,y}(da \times db) \right)^{1/p}$.

While the proof of the above lemma is similar to that of Claim 1 in [31, pg.69], we provide it in Sect. A.3 for completeness.

Proof of Proposition 3.10 First, consider the $p < \infty$ case. We divide the proof into the proofs of each inequality.

Proof of $2d_{\text{GW}_{p,q}} \geq \text{TLB}_{p,q}$. Fix an arbitrary coupling $\gamma \in \mathcal{M}(\mu_X, \mu_Y)$. Recall that,

$$(\text{dis}_{p,q}(\gamma))^p = \int_{X \times Y} \int_{X \times Y} (\Lambda_q(d_X(x, x'), d_Y(y, y')))^p \gamma(dx' \times dy') \gamma(dx \times dy).$$

Furthermore, observe that for each $x \in X$ and $y \in Y$, we have

$$\begin{aligned} & \int_{X \times Y} (\Lambda_q(d_X(x, x'), d_Y(y, y')))^p \gamma(dx' \times dy') \\ &= \int_{\mathbb{R}_+ \times \mathbb{R}_+} (\Lambda_q(a, b))^p (d_X(x, \cdot) \times d_Y(y, \cdot))_{\#} \gamma(da \times db) \\ &\geq (d_{\mathbb{W}^p}^{(\mathbb{R}, \Lambda_q)}(dh_X(x), dh_Y(y)))^p \end{aligned}$$

where the inequality holds since $dh_X(x) = (d_X(x, \cdot))_{\#} \mu_X$ and $dh_Y(y) = (d_Y(y, \cdot))_{\#} \mu_Y$, so $(d_X(x, \cdot), d_Y(y, \cdot))_{\#} \gamma$ is a coupling between $dh_X(x)$ and $dh_Y(y)$. This implies that

$$\text{dis}_{p,q}(\gamma) \geq \left(\int_{X \times Y} (d_{\mathbb{W}^p}^{(\mathbb{R}, \Lambda_q)}(dh_X(x), dh_Y(y)))^p \gamma(dx \times dy) \right)^{\frac{1}{p}}.$$

Since the choice of γ is arbitrary, infimizing over $\gamma \in \mathcal{M}(\mu_X, \mu_Y)$ establishes the required inequality. $\square$

Proof of $\text{TLB}_{p,q} \geq \text{SLB}_{p,q}$. First, consider the case $q < \infty$.

Fix an arbitrary coupling $\gamma \in \mathcal{M}(\mu_X, \mu_Y)$. By Lemma A.5, there is a measurable choice $(x, y) \mapsto v_{x,y}$ such that for each $(x, y) \in X \times Y$, $v_{x,y}$ belongs to $\mathcal{M}(dh_X(x), dh_Y(y))$ and

$$d_{\mathbb{W}^p}^{(\mathbb{R}, \Lambda_q)}(dh_X(x), dh_Y(y)) = \left(\int_{\mathbb{R}_+ \times \mathbb{R}_+} (\Lambda_q(a, b))^p v_{x,y}(da \times db) \right)^{\frac{1}{p}}.$$

Next, define a measure ν on $\mathbb{R}_+ \times \mathbb{R}_+$ by:

$$\nu := \int_{X \times Y} v_{x,y} \gamma(dx \times dy).$$

Inspection of the marginals of ν shows that it is a coupling between dH_X and dH_Y . Indeed, for each $S \in \Sigma_{\mathbb{R}_+}$,

$$\begin{aligned} \nu(S \times \mathbb{R}_+) &= \int_{X \times Y} v_{x,y}(S \times \mathbb{R}_+) \gamma(dx \times dy) \\ &= \int_{X \times Y} dh_X(x)(S) \gamma(dx \times dy) \\ &= \int_X dh_X(x)(S) \mu_X(dx) = dH_X(S) \end{aligned}$$

where the last equality holds by Lemma A.4. A similar argument proves that $\nu(\mathbb{R}_+ \times S) = dH_Y(S)$ so indeed $\nu \in \mathcal{M}(dH_X, dH_Y)$. Therefore,

$$\int_{X \times Y} (d_{\mathbb{W}^p}^{(\mathbb{R}, \Lambda_q)}(dh_X(x), dh_Y(y)))^p \gamma(dx \times dy)$$

$$\begin{aligned}
 &= \int_{X \times Y} \int_{\mathbb{R}_+ \times \mathbb{R}_+} (\Lambda_q(a, b))^p v_{X,Y}(da \times db) \gamma(dx \times dy) \\
 &= \int_{\mathbb{R}_+ \times \mathbb{R}_+} (\Lambda_q(a, b))^p v(da \times db) \geq (d_{\mathbb{W}^p}^{(\mathbb{R}, \Lambda_q)}(dH_X, dH_Y))^p.
 \end{aligned}$$

The required inequality follows since the choice of γ is arbitrary. In order to establish the claim when $q = \infty$ we employ the case when $q < \infty$ and the fact that the following equalities hold:

$$\text{TLB}_{p,\infty} = \lim_{q \rightarrow \infty} \text{TLB}_{p,q} \text{ and } \text{SLB}_{p,\infty} = \lim_{q \rightarrow \infty} \text{SLB}_{p,q}.$$

These can be verified by observing that Λ_q uniformly converges to Λ_∞ on the compact set $\{d_X(x, x') \mid x, x' \in X\} \cup \{d_Y(y, y') \mid y, y' \in Y\} \subset \mathbb{R}_+$ as q goes to infinity. $\square$

Proof of $\text{SLB}_{p,q} \geq \text{DLB}_{p,p \wedge q}$. We divide the proof into two cases.

Case 1. ($p \geq q$): Observe first that, since $p < \infty$, we have

$$\begin{aligned}
 (\text{SLB}_{p,q}(X, Y))^q &= \left(\int_0^1 (\Lambda_q(H_X^{-1}(u), H_Y^{-1}(u)))^p du \right)^{q/p} \\
 &= \left(\int_0^1 ((H_X^{-1}(u))^q - (H_Y^{-1}(u))^q)^{p/q} du \right)^{q/p} \\
 &= \left(\int_0^1 (F_{(S_q)\#dH_X}^{-1}(u) - F_{(S_q)\#dH_Y}^{-1}(u))^{p/q} du \right)^{q/p} \\
 &= d_{\mathbb{W}^{p/q}}^{\mathbb{R}}((S_q)\#dH_X, (S_q)\#dH_Y) \\
 &\geq \left| d_{\mathbb{W}^{p/q}}^{\mathbb{R}}((S_q)\#dH_X, \delta_0) - d_{\mathbb{W}^{p/q}}^{\mathbb{R}}((S_q)\#dH_Y, \delta_0) \right|
 \end{aligned}$$

where δ_0 is the Dirac measure at zero. Note that the first equality follows from Remark 1.10 and the last inequality follows from the triangle inequality of $d_{\mathbb{W}^{p/q}}^{\mathbb{R}}$. Next, we obtain via Example 1.6 and Remark 3.3 that

$$d_{\mathbb{W}^{p/q}}^{\mathbb{R}}((S_q)\#dH_X, \delta_0) = \left(\int_X d_X^p(x, x') \mu_X(dx) \mu_X(dx') \right)^{q/p} = (\text{diam}_p(X))^q.$$

This establishes the inequality $\text{SLB}_{p,q} \geq \text{DLB}_{p,q}$ when $p \geq q$ and $p < \infty$.

Case 2. ($p \leq q$): Note that $\text{DLB}_{p,p \wedge q} = \text{DLB}_{p,p}$ in this case. Also, it is easy to verify that $\text{SLB}_{p,q} \geq \text{SLB}_{p,p}$ since $\Lambda_p \leq \Lambda_q$. Moreover, $\text{SLB}_{p,p} \geq \text{DLB}_{p,p}$ by the previous Case 1. Hence, we achieve the inequality $\text{SLB}_{p,q} \geq \text{DLB}_{p,p} = \text{DLB}_{p,p \wedge q}$. $\square$

This completes the proof of the hierarchy in the case where $p < \infty$. More precisely, thus far we have proved that

$$2 d_{\text{GW},p,q} \geq \text{TLB}_{p,q} \geq \text{SLB}_{p,q} \geq \text{DLB}_{p,p \wedge q} \quad \text{for } p \in [1, \infty) \text{ and } q \in [1, \infty]. \quad (16)$$

Lastly, in order to prove the $p = \infty$ case, we proceed as follows. Unless otherwise specified, we consider $q \in [1, \infty]$.

Proof of $2d_{\text{GW},\infty,q} \geq \text{TLB}_{\infty,q}$. Note that by Theorem 2

$$2d_{\text{GW},\infty,q} \geq 2 \limsup_{p \rightarrow \infty} d_{\text{GW},p,q}.$$

Also, it is easy to verify that $\text{TLB}_{\infty,q} \geq \limsup_{p \rightarrow \infty} \text{TLB}_{p,q}$ since $d_{\text{W},p}^{(\mathbb{R}, \Lambda_q)} \leq d_{\text{W},p'}^{(\mathbb{R}, \Lambda_q)}$ whenever $1 \leq p \leq p' \leq \infty$. Therefore, $2d_{\text{GW},\infty,q} \geq \text{TLB}_{\infty,q}$ is obtained from the $p < \infty$ case. $\square$

Proof of $\text{TLB}_{\infty,q} \geq \text{SLB}_{\infty,q}$. This is straightforward since $\text{TLB}_{\infty,q} \geq \limsup_{p \rightarrow \infty} \text{TLB}_{p,q}$, $\text{SLB}_{\infty,q} = \lim_{p \rightarrow \infty} \text{SLB}_{p,q}$ and we proved that $\text{TLB}_{p,q} \geq \text{SLB}_{p,q}$ in the $p < \infty$ case. $\square$

Proof of $\text{SLB}_{\infty,q} \geq \text{DLB}_{\infty,\infty \wedge q}$. If $q < \infty$, then it is easy to verify the claim from the facts that $\text{SLB}_{\infty,q} = \lim_{p \rightarrow \infty} \text{SLB}_{p,q}$, $\text{DLB}_{\infty,q} = \lim_{p \rightarrow \infty} \text{DLB}_{p,q}$ and the $p < \infty$ case. Finally, if $q = \infty$, one can use the facts $\text{SLB}_{\infty,\infty} = \lim_{q \rightarrow \infty} \text{SLB}_{\infty,q}$, $\text{DLB}_{\infty,\infty} = \lim_{q \rightarrow \infty} \text{DLB}_{\infty,q}$, and the previous case when $p = \infty$, $q < \infty$. This completes the proof. $\square$

By the last three cases, we have thus proved that

$$2d_{\text{GW},\infty,q} \geq \text{TLB}_{\infty,q} \geq \text{SLB}_{\infty,q} \geq \text{DLB}_{\infty,\infty \wedge q} \quad \text{for } q \in [1, \infty]. \quad (17)$$

Equations (16) and (17) establish the theorem. $\square$

A.3 Proofs of Lemmas

Proof of Lemma 2.6 By Equations (5), (6) and (7) we have

$$\begin{aligned} \text{dis}_{4,2}^4(\gamma_{m,n}) &= \frac{4}{n+1} + \frac{4}{m+1} \\ &\quad - 8 \int_{\mathbb{S}^n \times \mathbb{S}^m} \int_{\mathbb{S}^n \times \mathbb{S}^m} \langle x, x' \rangle \langle y, y' \rangle \gamma_{m,n}(dx \times dy) \gamma_{m,n}(dx' \times dy'). \end{aligned}$$

We then compute

$$\begin{aligned} &\int_{\mathbb{S}^n \times \mathbb{S}^m} \int_{\mathbb{S}^n \times \mathbb{S}^m} \langle x, x' \rangle \langle y, y' \rangle \gamma_{m,n}(dx \times dy) \gamma_{m,n}(dx' \times dy') \\ &= \iint_{\mathbb{S}^n \times \mathbb{S}^n} \langle e_{n,m}(y), e_{n,m}(y') \rangle \langle y, y' \rangle \mu_n(dy) \mu_n(dy') \\ &= \iint_{\mathbb{S}^n \times \mathbb{S}^n} \sum_{i=1}^{n+1} y_i y'_i \sum_{j=1}^{m+1} \frac{y_j}{\|y_A\|} \frac{y'_j}{\|y'_A\|} \mu_n(dy) \mu_n(dy') \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^{n+1} \sum_{j=1}^{m+1} \left(\int_{\mathbb{S}^n} \frac{y_i y_j}{\|y_A\|} \mu_n(dy) \right)^2 \\
 &= \sum_{j=1}^{m+1} \left(\int_{\mathbb{S}^n} \frac{y_j^2}{\|y_A\|} \mu_n(dy) \right)^2 \\
 &= (m+1) \left(\int_{\mathbb{S}^n} \frac{y_1^2}{\|y_A\|} \mu_n(dy) \right)^2.
 \end{aligned}$$

Here, note that the fourth equality holds because $\int_{\mathbb{S}^n} \frac{y_i y_j}{\|y_A\|} \mu_n(dy) = 0$ whenever $i \neq j$ because $\mu_m = (N_m)_{\#} \eta_{m+1}$ where $N_m : \mathbb{R}^{m+1} \rightarrow \mathbb{S}^m$ is the map such that $(y_1, \dots, y_{m+1}) \mapsto \frac{1}{\sqrt{\sum_{i=1}^{m+1} y_i^2}} (y_1, \dots, y_{m+1})$ and η_{m+1} is the standard Gaussian measure on $\mathbb{R}^{m+1}$.

Also,

$$\int_{\mathbb{S}^n} \frac{y_1^2}{\|y_A\|} \mu_n(dy) = \frac{1}{m+1} \int_{\mathbb{S}^n} \frac{\sum_{j=1}^{m+1} y_j^2}{\|y_A\|} \mu_n(dy) = \frac{1}{m+1} \int_{\mathbb{S}^n} \|y_A\| \mu_n(dy).$$

It remains to calculate the expectation of $\|y_A\|$, which follows from a calculation identical to the one in the proof of Theorem 1. In particular, we appeal to a characterization of μ_n in terms of standard Gaussian random variables $Z_1, \dots, Z_{n+1}$ in order to write that

$$\|y_A\|^2 \sim \text{Beta} \left(\frac{m+1}{2}, \frac{n-m}{2} \right).$$

And hence

$$\begin{aligned}
 \int_{\mathbb{S}^n} \|y_A\| \mu_n(dy) &= \frac{1}{\beta \left(\frac{m+1}{2}, \frac{n-m}{2} \right)} \int_0^1 \sqrt{t} \cdot t^{(m+1)/2-1} (1-t)^{(n-m)/2-1} dt \\
 &= \frac{\beta \left(\frac{m+2}{2}, \frac{n-m}{2} \right)}{\beta \left(\frac{m+1}{2}, \frac{n-m}{2} \right)} = \frac{\Gamma \left(\frac{m+2}{2} \right) \Gamma \left(\frac{n+1}{2} \right)}{\Gamma \left(\frac{m+1}{2} \right) \Gamma \left(\frac{n+2}{2} \right)}.
 \end{aligned}$$

This finishes the proof. $\square$

Proof of Lemma A.5 First, let $S := \{d_X(x, x') \mid x, x' \in X\} \cup \{d_Y(y, y') \mid y, y' \in Y\} \subset \mathbb{R}_+$. Since both X and Y are compact, S is also compact. Also, it is easy to verify that all Λ_r (for $r \in [1, \infty)$) induce the same topology and thus the same Borel sets on S . Therefore all $d_{W_p}^{(\mathbb{R}, \Lambda_r)}$ (for $r \in [1, \infty)$) metrize the weak topology on $\mathcal{P}(S)$. By [32, Remark 1], the following two maps are continuous with respect to the weak topology and thus measurable:

$$\Phi_1 : X \rightarrow \mathcal{P}(S), \quad x \mapsto dh_X(x)$$

and

$$\Phi_2 : Y \rightarrow \mathcal{P}(S), y \mapsto dh_Y(y).$$

Since S is a compact space, the space $(\mathcal{P}(S), d_{\mathcal{W}_p}^{(\mathbb{R}, \Lambda_q)})$ is separable [53, Theorem 6.18]. This yields that $\Sigma_{\mathcal{P}(S) \times \mathcal{P}(S)} = \Sigma_{\mathcal{P}(S)} \otimes \Sigma_{\mathcal{P}(S)}$ [19, Proposition 1.5]. Hence, the product Φ of Φ_1 and Φ_2 , defined by

$$\Phi : X \times Y \rightarrow \mathcal{P}(S) \times \mathcal{P}(S), (x, y) \mapsto (dh_X(x), dh_Y(y))$$

is measurable [19, Proposition 2.4]. Since Φ is measurable, a direct application of [53, Corollary 5.22] gives the claim. $\square$

Appendix B Calculations

B.1 Lower Bounds for $d_{\text{GW}4,2}$ Between Spheres with the Geodesic Distance

In preparation for the determination of the diameter lower bounds, we first compute the 4-diameters of $\mathbb{S}_G^m$ for $m = 0, 1, 2$ using the formula for the global distance distribution given in Example 3.4.

Example B.1 (The global distance distributions of $\mathbb{S}_G^0, \mathbb{S}_G^1$ and $\mathbb{S}_G^2$) $\mathbb{S}_G^0$ consists of two points which are at distance π apart and so $H_{\mathbb{S}_G^0}^0(t) = \mu_0 \otimes \mu_0\{(x, x') \in \mathbb{S}_G^0 \times \mathbb{S}_G^0 \mid d_0(x, x') \leq t\}$ is

$$H_{\mathbb{S}_G^0}^0(t) = \begin{cases} \frac{1}{2} & 0 \leq t < \pi \\ 1 & t = \pi. \end{cases}$$

By Example 3.4, the global distance distributions of $\mathbb{S}_G^1$ and $\mathbb{S}_G^2$ are

$$H_{\mathbb{S}_G^1}^1(t) = \frac{t}{\pi}, \quad H_{\mathbb{S}_G^2}^2(t) = \frac{(1 - \cos t)}{2} \text{ for } t \in [0, \pi].$$

Consequently, the generalized inverses (see Eq. 3) are

$$H_{\mathbb{S}_G^0}^{-1}(u) = \begin{cases} 0 & 0 \leq u \leq \frac{1}{2} \\ \pi & \frac{1}{2} < u \leq 1 \end{cases}$$

while

$$H_{\mathbb{S}_G^1}^{-1}(u) = u\pi \quad \text{and} \quad H_{\mathbb{S}_G^2}^{-1}(u) = \arccos(1 - 2u) \quad \text{for } u \in [0, 1].$$

Example B.2 ($\text{DLB}_{4,2}(\mathbb{S}_G^0, \mathbb{S}_G^1)$ and $\text{DLB}_{4,2}(\mathbb{S}_G^1, \mathbb{S}_G^2)$) By Example B.1 the 4-diameters of $\mathbb{S}_G^0$, $\mathbb{S}_G^1$ and $\mathbb{S}_G^2$ are

$$\begin{aligned}\text{diam}_4(\mathbb{S}_G^0) &= \left(\int_0^1 \left(H_{\mathbb{S}_G^0}^{-1}(u) \right)^4 du \right)^{1/4} = \left(\int_{1/2}^1 \pi^4 du \right)^{1/4} = \frac{\pi}{2^{1/4}}, \\ \text{diam}_4(\mathbb{S}_G^1) &= \left(\int_0^1 \left(H_{\mathbb{S}_G^1}^{-1}(u) \right)^4 du \right)^{1/4} = \left(\int_0^1 (u\pi)^4 du \right)^{1/4} = \frac{\pi}{5^{1/4}}, \\ \text{diam}_4(\mathbb{S}_G^2) &= \left(\int_0^1 \left(H_{\mathbb{S}_G^2}^{-1}(u) \right)^4 du \right)^{1/4} = \left(\int_0^1 (\arccos(1-2u))^4 du \right)^{1/4} \\ &= \left(24 - 6\pi^2 + \frac{\pi^4}{2} \right)^{1/4}.\end{aligned}$$

Hence, by the definition of $\text{DLB}_{4,2}$ we have

$$\text{DLB}_{4,2}(\mathbb{S}_G^0, \mathbb{S}_G^1) = \left| \left(\frac{\pi^4}{2} \right)^2 - \left(\frac{\pi^4}{5} \right)^2 \right|^{1/2} = \pi \left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{5}} \right)^{1/2} \approx 1.602.$$

and

$$\text{DLB}_{4,2}(\mathbb{S}_G^1, \mathbb{S}_G^2) = \left| \left(\frac{\pi}{5^{1/4}} \right)^2 - \left(24 - 6\pi^2 + \frac{\pi^4}{2} \right)^{2/4} \right|^{1/2} \approx 0.861.$$

Example B.3 ($\text{SLB}_{4,2}(\mathbb{S}_G^0, \mathbb{S}_G^1)$ and $\text{SLB}_{4,2}(\mathbb{S}_G^1, \mathbb{S}_G^2)$) By Example B.1 and the definition of $\text{SLB}_{4,2}$ we obtain

$$\begin{aligned}\text{SLB}_{4,2}(\mathbb{S}_G^0, \mathbb{S}_G^1) &= \left(\int_0^1 |(H_{\mathbb{S}_G^0}^{-1}(u))^2 - (H_{\mathbb{S}_G^1}^{-1}(u))^2|^2 du \right)^{1/4} \\ &= \left(\int_0^{\frac{1}{2}} |u^2\pi^2|^2 du + \int_{\frac{1}{2}}^1 |\pi^2 - u^2\pi^2|^2 du \right)^{1/4} \\ &= \pi \left(\frac{1}{2} + \frac{1}{5} - \frac{7}{12} \right)^{1/4} \approx 1.836,\end{aligned}$$

and similarly

$$\begin{aligned}\text{SLB}_{4,2}(\mathbb{S}_G^1, \mathbb{S}_G^2) &= \left(\int_0^1 |(H_{\mathbb{S}_G^1}^{-1}(u))^2 - (H_{\mathbb{S}_G^2}^{-1}(u))^2|^2 du \right)^{1/4} \\ &= \left(\int_0^1 |u^2\pi^2 - (\arccos(1-2u))^2|^2 du \right)^{1/4} \approx 0.931.\end{aligned}$$

B.2 Lower Bounds for $d_{\text{GW},2}$ Between Spheres with the Euclidean Metric

Example B.4 (The global distance distributions of $\mathbb{S}_E^0, \mathbb{S}_E^1$ and $\mathbb{S}_E^2$) $\mathbb{S}_E^0$ consists of two points which are at distance 2 apart.

$$H_{\mathbb{S}_E^0}(t) = \begin{cases} \frac{1}{2} & 0 \leq t < 2 \\ 1 & t = 2. \end{cases}$$

Consequently,

$$H_{\mathbb{S}_E^0}^{-1}(u) = \begin{cases} 0 & 0 \leq u \leq \frac{1}{2} \\ 2 & \frac{1}{2} < u \leq 1. \end{cases}$$

Similarly, the global distance distributions for $\mathbb{S}_G^1$ and $\mathbb{S}_G^2$ are

$$H_{\mathbb{S}_E^1}(t) = \frac{2}{\pi} \arcsin \frac{t}{2}, \quad H_{\mathbb{S}_E^2}(t) = \frac{t^2}{4} \quad \text{for } t \in [0, 2].$$

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

$$H_{\mathbb{S}_E^1}^{-1}(u) = 2 \sin \left(\frac{u\pi}{2} \right) \quad \text{and} \quad H_{\mathbb{S}_E^2}^{-1}(u) = 2\sqrt{u}.$$

Example B.5 ($\text{DLB}_{4,2}(\mathbb{S}_E^0, \mathbb{S}_E^1)$ and $\text{DLB}_{4,2}(\mathbb{S}_E^1, \mathbb{S}_E^2)$) By Example B.4, the 4-diameters of $\mathbb{S}_E^0, \mathbb{S}_E^1$ and $\mathbb{S}_E^2$ are

$$\begin{aligned} \text{diam}_4(\mathbb{S}_E^0) &= \left(\int_0^1 \left(H_{\mathbb{S}_E^0}^{-1}(u) \right)^4 du \right)^{1/4} = \left(\int_{1/2}^1 2^4 du \right)^{1/4} = 2^{3/4}. \\ \text{diam}_4(\mathbb{S}_E^1) &= 2 \left(\int_0^1 \sin^4 \left(\frac{u\pi}{2} \right) du \right)^{1/4} = 2 \left(\frac{3}{8} \right)^{1/4}. \\ \text{diam}_4(\mathbb{S}_E^2) &= 2 \left(\int_0^1 (\sqrt{u})^4 du \right)^{1/4} = \frac{2}{3^{1/4}}. \end{aligned}$$

Hence by the definition of $\text{DLB}_{4,2}$ we have

$$\text{DLB}_{4,2}(\mathbb{S}_E^0, \mathbb{S}_E^1) = \left| 2^{3/2} - 2^2 \left(\frac{3}{8} \right)^{2/4} \right|^{1/2} \approx 0.616.$$

and

$$\text{DLB}_{4,2}(\mathbb{S}_E^1, \mathbb{S}_E^2) = \left| 2^2 \left(\frac{3}{8} \right)^{2/4} - \left(\frac{2}{3^{1/4}} \right)^2 \right|^{1/2} \approx 0.374.$$

Example B.6 ($\text{SLB}_{4,2}(\mathbb{S}_E^0, \mathbb{S}_E^1)$ and $\text{SLB}_{4,2}(\mathbb{S}_E^1, \mathbb{S}_E^2)$) By Example B.4 and the definition of $\text{SLB}_{4,2}$ we obtain

$$\begin{aligned} \text{SLB}_{4,2}(\mathbb{S}_E^0, \mathbb{S}_E^1) &= \left(\int_0^1 |(H_{\mathbb{S}_E^0}^{-1}(u))^2 - (H_{\mathbb{S}_E^1}^{-1}(u))^2|^2 du \right)^{1/4} \\ &= \left(\int_0^{\frac{1}{2}} |H_{\mathbb{S}_E^1}^{-1}(u)|^4 ds + \int_{\frac{1}{2}}^1 |2^2 - (H_{\mathbb{S}_E^1}^{-1}(u))^2|^2 du \right)^{1/4} \\ &= \left(\int_0^{\frac{1}{2}} \left| 2^2 \sin^2 \left(\frac{u\pi}{2} \right) \right|^2 du + \int_{\frac{1}{2}}^1 \left| 2^2 - 2^2 \sin^2 \left(\frac{u\pi}{2} \right) \right|^2 du \right)^{1/4} \\ &\approx 0.976, \end{aligned}$$

and similarly

$$\text{SLB}_{4,2}(\mathbb{S}_E^1, \mathbb{S}_E^2) = \left(\int_0^1 \left| 2^2 \sin^2 \left(\frac{u\pi}{2} \right) - 2^2 u \right|^2 du \right)^{1/4} \approx 0.549.$$

B.3 Distortion $\text{dis}_{4,2}$ Under the Equatorial Coupling

Example B.7 ($\text{dis}_{4,2}(\gamma_{0,1}, \mathbb{S}_G^0, \mathbb{S}_G^1)$) By Remark 1.18 and Example B.2, we have

$$\begin{aligned} &(\text{dis}_{4,2}(\gamma_{0,1}, \mathbb{S}_G^0, \mathbb{S}_G^1))^4 \\ &= (\text{diam}_4(\mathbb{S}_G^0))^4 + (\text{diam}_4(\mathbb{S}_G^1))^4 \\ &\quad - 2 \int (d_0(e_{1,0}(y), e_{1,0}(y')))^2 (d_1(y, y'))^2 \mu_1(dy) \mu_1(dy') \\ &= \frac{\pi^4}{2} + \frac{\pi^4}{5} - 2 \int_{\mathbb{S}^1} \int_{\mathbb{S}^1} (d_0(e_{1,0}(y), e_{1,0}(y')))^2 (d_1(y, y'))^2 \mu_1(dy) \mu_1(dy'). \end{aligned} \tag{18}$$

We use polar coordinates to compute the integral on the right hand side. We write $y = (\cos \theta, \sin \theta)$ and $y' = (\cos \theta', \sin \theta')$, where $\theta, \theta' \in [0, 2\pi]$. Note also that with this parametrization and by the definition of the map $e_{1,0}$, we can write

$$e_{1,0}(y) = \text{sign}(\cos \theta); \quad e_{1,0}(y') = \text{sign}(\cos \theta')$$

Hence by definition of d_0 ,

$$d_0(e_{1,0}(y), e_{1,0}(y')) = \pi \mathbb{1}(\cos \theta \cdot \cos \theta' < 0).$$

Moreover,

$$d_1(y, y') = \arccos(\cos \theta \cos \theta' + \sin \theta \sin \theta') = \arccos(\cos(\theta - \theta')).$$

We then have

$$\begin{aligned}
 & \int_{\mathbb{S}^1} \int_{\mathbb{S}^1} (d_0(e_{1,0}(y), e_{1,0}(y')))^2 (d_1(y, y'))^2 \mu_1(dy) \mu_1(dy') \\
 &= \int_0^{2\pi} \int_0^{2\pi} (\pi \mathbb{1}(\cos \theta \cdot \cos \theta' < 0))^2 (\arccos(\cos(\theta - \theta')))^2 \times \left(\frac{1}{2\pi}\right)^2 d\theta d\theta' \\
 &= \frac{1}{4} \left(\int_0^{\pi/2} \int_{\pi/2}^{3\pi/2} (\arccos(\cos(\theta' - \theta)))^2 d\theta d\theta' \right. \\
 &\quad \left. + \int_{3\pi/2}^{2\pi} \int_{\pi/2}^{3\pi/2} (\arccos(\cos(\theta - \theta')))^2 d\theta d\theta' \right) \times 2 \\
 &= \frac{1}{2} \left(\int_0^{\pi/2} \int_{\pi/2-\theta}^{3\pi/2-\theta} (\arccos(\cos(t)))^2 dt d\theta \right. \\
 &\quad \left. + \int_{3\pi/2}^{2\pi} \int_{\theta-3\pi/2}^{\theta-\pi/2} (\arccos(\cos(t)))^2 dt d\theta \right).
 \end{aligned}$$

The first integral becomes

$$\begin{aligned}
 & \int_0^{\pi/2} \int_{\pi/2-\theta}^{3\pi/2-\theta} (\arccos(\cos(t)))^2 dt d\theta \\
 &= \int_0^{\pi/2} \left[\int_{\pi/2-\theta}^{\pi} t^2 dt + \int_{\pi}^{3\pi/2-\theta} (2\pi - t)^2 dt \right] d\theta \\
 &= \int_0^{\pi/2} \left[\left(\frac{\pi^3}{3} - \frac{(\pi/2 - \theta)^3}{3} \right) + \left(\frac{\pi^3}{3} - \frac{(\theta + \pi/2)^3}{3} \right) \right] d\theta \\
 &= \frac{\pi^4}{3} - \int_0^{\pi} \frac{u^3}{3} du = \frac{\pi^4}{3} - \frac{\pi^4}{12} = \frac{\pi^4}{4}.
 \end{aligned}$$

Similarly, the second integral also evaluates to

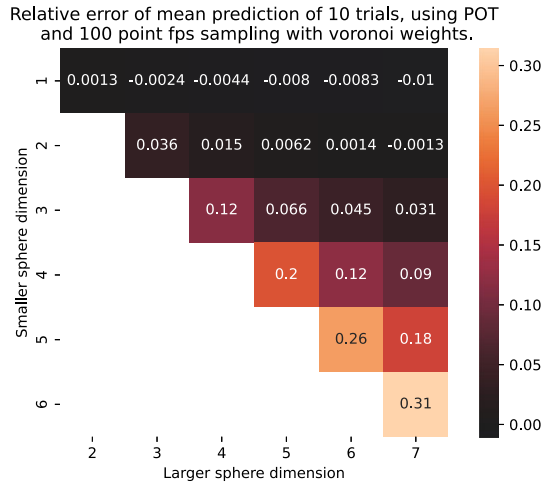
$$\begin{aligned}
 & \int_{3\pi/2}^{2\pi} \int_{\theta-3\pi/2}^{\theta-\pi/2} (\arccos(\cos(t)))^2 dt d\theta \\
 &= \int_{3\pi/2}^{2\pi} \left[\int_{\theta-3\pi/2}^{\pi} t^2 dt + \int_{\pi}^{\theta-\pi/2} (2\pi - t)^2 dt \right] d\theta = \frac{\pi^4}{4}.
 \end{aligned}$$

Thus,

$$\int_{\mathbb{S}^1} \int_{\mathbb{S}^1} (d_0(e_{1,0}(y), e_{1,0}(y')))^2 (d_1(y, y'))^2 \mu_1(dy) \mu_1(dy') = \frac{\pi^4}{4}.$$

The rest of the calculations are given in Example 2.4.

Fig. 6 Relative errors of computed and true differences using the CGD solver from POT, with FPS as the sampling procedure and Voronoi weights



Appendix C Another Experiment: Varying Dimensions Experiment

In this experiment, we fixed the number of samples taken at 100 and varied the dimensions of the two spheres between 1 and 7. The subsampling method was chosen to be FPS and the weights were those produced by the Voronoi method. Finally, we fixed the solver to the CGD solver from POT.

Using the results of ten trials ($n_{\text{trials}} = 10$) for fixed sphere dimensions, m and n , $\hat{d}_{m,n}^i$, where $i = 1, \dots, n_{\text{trials}}$, we estimated the true distance via the average over trials,

$$d_{m,n} \approx \hat{d}_{m,n} := \frac{1}{n_{\text{trials}}} \sum_{i=1}^{n_{\text{trials}}} \hat{d}_{m,n}^i.$$

We then recorded the relative error of this estimator: $\text{relative-error}_{m,n} := \frac{\hat{d}_{m,n} - d_{m,n}}{d_{m,n}}$ in the corresponding entry of the heatmap shown in Fig. 6. We observe a dramatic decrease in accuracy as the dimensions of both spheres increase which of course one would expect to reduce by using a larger number of points.

References

1. Henry Adams, Johnathan Bush, Nate Clause, Florian Frick, Mario Gómez, Michael Harrison, R Amzi Jeffs, Evgeniya Lagoda, Sunhyuk Lim, and Facundo Mémoli. Gromov-Hausdorff distances, Borsuk-Ulam theorems, and Vietoris-Rips complexes. *arXiv preprint arXiv:2301.00246*, 2022.
2. 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, Brussels, Belgium, October–November 2018. Association for Computational Linguistics.

3. Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*, Lectures in Mathematics. ETH Zürich, 2nd edn. Springer Science & Business Media, 2005.
4. Kendall Atkinson and Weimin Han. *Spherical harmonics and approximations on the unit sphere: an introduction*, volume 2044. Springer Science & Business Media, 2012.
5. Fred L Bookstein. The study of shape transformation after D'Arcy Thompson. *Mathematical Biosciences* 34:177–129. [https://doi.org/10.1016/0025-5564\(77\)90101-8](https://doi.org/10.1016/0025-5564(77)90101-8), 1977.
6. Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33. American Mathematical Society, 2022.
7. 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.
8. Samir Chowdhury and Tom Needham. Gromov–Wasserstein averaging in a Riemannian framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 842–843, 2020.
9. Michael L Collyer, David J Sekora, and Dean C Adams. A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity*, 115(4):357–365, 2015.
10. Marco Cuturi, Laetitia Meng-Papaxanthos, Yingtao Tian, Charlotte Bunne, Geoff Davis, and Olivier Teboul. Optimal transport tools (ott): A jax toolbox for all things Wasserstein. *arXiv preprint arXiv:2201.12324*, 2022.
11. Morris H DeGroot and Mark J Schervish. *Probability and statistics*. Pearson Education, 2012.
12. Julie Delon, Agnes Desolneux, and Antoine Salmona. Gromov–Wasserstein distances between Gaussian distributions. *Journal of Applied Probability*, 59(4):1178–1198, 2022.
13. Ian L Dryden and Kanti V Mardia. *Statistical shape analysis: with applications in R*, volume 995. John Wiley & Sons, 2016.
14. Lester Dubins and Gideon Schwarz. Equidiscontinuity of Borsuk–Ulam functions. *Pacific Journal of Mathematics*, 95(1):51–59, 1981.
15. 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.
16. Yuval Eldar, Michael Lindenbaum, Moshe Porat, and Yehoshua Y Zeevi. The farthest point strategy for progressive image sampling. *IEEE transactions on image processing*, 6(9):1305–1315, 1997.
17. Chao Fan, Wenwen Li, Levi J Wolf, and Soe W Myint. A spatiotemporal compactness pattern analysis of congressional districts to assess partisan gerrymandering: a case study with California and North Carolina. *Annals of the Association of American Geographers*, 105(4):736–753, 2015.
18. Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. POT: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.
19. Gerald B Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.
20. Teofilo F Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical computer science*, 38:293–306, 1985.
21. Colin Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):285–321, 1991.
22. Mikhail Gromov, Misha Katz, Pierre Pansu, and Stephen Semmes. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152. Springer, 1999.
23. Reigo Hendrikson. Using Gromov–Wasserstein distance to explore sets of networks. 2016.
24. Aaron R Kaufman, Gary King, and Mayya Komisarchik. How to measure legislative district compactness if you only know it when you see it. *American Journal of Political Science*, 65(3):533–550, 2021.
25. Christian Peter Klingenberg and Grant S McIntyre. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution*, 52(5):1363–1375, 1998.
26. Marius Loeordeanu and Martial Hebert. A spectral technique for correspondence problems using pairwise constraints. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1482–1489, 2005.

27. Sunhyuk Lim, Facundo Mémoli, and Zane Smith. The Gromov–Hausdorff distance between spheres. *Geometry & Topology*, 27(9):3733–3800, 2023.
28. Yaron Lipman, Jesus Puente, and Ingrid Daubechies. Conformal Wasserstein distance: II. computational aspects and extensions. *Mathematics of Computation*, 82:331–381. <https://doi.org/10.1090/S0025-5718-2012-02569-5>, 2013.
29. Facundo Mémoli. On the use of Gromov–Hausdorff distances for shape comparison. *Proceedings of Point Based Graphics*, 2007.
30. Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011.
31. Facundo Mémoli, Axel Munk, Zhengchao Wan, and Christoph Weitkamp. The ultrametric Gromov–Wasserstein distance. *Discrete and Computational Geometry* 70(4):1378–1450. <https://doi.org/10.1007/s00454-023-00583-0>, 2023.
32. Facundo Mémoli and Tom Needham. Distance distributions and inverse problems for metric measure spaces. *Studies in Applied Mathematics*, 149(4):943–1001, 2022.
33. Facundo Mémoli and Guillermo Sapiro. A theoretical and computational framework for isometry invariant recognition of point cloud data. *Foundations of Computational Mathematics*, 5:313–347, 2005.
34. Facundo Mémoli, Anastasios Sidiropoulos, and Kritika Singhal. Sketching and clustering metric measure spaces. *arXiv preprint arXiv:1801.00551*, 2018.
35. Facundo Mémoli and Zhengchao Wan. On p -metric spaces and the p -Gromov–Hausdorff Distance. *p-Adic Numbers, Ultrametric Analysis and Applications*, 14(3):173–223, 2022.
36. Michael I Miller. Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. *NeuroImage*, 23:S19–S33, 2004.
37. Leandro R Monteiro, José Alexandre F Diniz-Filho, Sérgio F dos Reis, and Edilson D Araújo. Geometric estimates of heritability in biological shape. *Evolution*, 56(3):563–572, 2002.
38. Mervin E Muller. A note on a method for generating points uniformly on n -dimensional spheres. *Communications of the ACM*, 2(4):19–20, 1959.
39. Daniel Packer. Gromov–Wasserstein distance between spheres. <https://github.com/Daniel-Packer/gw-between-spheres>, 2024.
40. Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
41. Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov–Wasserstein averaging of kernel and distance matrices. In *International conference on machine learning*, pages 2664–2672. PMLR, 2016.
42. Gabriel Rioux, Ziv Goldfeld, and Kengo Kato. Entropic Gromov–Wasserstein distances: Stability and algorithms. *arXiv preprint arXiv:2306.00182*, 2023.
43. DL Robinson, PG Blackwell, EC Stillman, and AH Brook. Impact of landmark reliability on the planar Procrustes analysis of tooth shape. *Archives of oral biology*, 47(7):545–554, 2002.
44. F James Rohlf. Morphometrics. *Annual Review of Ecology and Systematics*, 21(1):299–316, 1990.
45. Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover’s distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.
46. Meyer Scetbon, Gabriel Peyré, and Marco Cuturi. Linear-time Gromov–Wasserstein distances using low rank couplings and costs. In *International Conference on Machine Learning*, pages 19347–19365. PMLR, 2022.
47. Ann I Scher, Yuan Xu, ESC Korf, Lon R White, Philip Scheltens, Arthur W Toga, Paul M Thompson, SW Hartley, MP Witter, Daniel J Valentino, et al. Hippocampal shape analysis in Alzheimer’s disease: a population-based study. *Neuroimage*, 36(1):8–18, 2007.
48. Karl-Theodor Sturm. The Space of Spaces: Curvature Bounds and Gradient Flows on the Space of Metric Measure Spaces. *Memoirs of the American Mathematical Society*, volume 290, Number 1443. American Mathematical Society. <https://doi.org/10.1090/memo/1443>, 2012.
49. Titouan Vayer. *A contribution to Optimal Transport on incomparable spaces*. PhD thesis, Lorient, 2020.
50. Titouan Vayer, Laetitia Chapel, Rémi Flamary, Romain Tavenard, and Nicolas Courty. Fused Gromov–Wasserstein distance for structured objects. *Algorithms*, 13(9):212, 2020.
51. Titouan Vayer, Nicolas Courty, Romain Tavenard, and Rémi Flamary. Optimal transport for structured data with application on graphs. In *International Conference on Machine Learning*, pages 6275–6284. PMLR, 2019.

52. Titouan Vayer, Rémi Flamary, Nicolas Courty, Romain Tavenard, and Laetitia Chapel. Sliced Gromov–Wasserstein. *Advances in Neural Information Processing Systems*, 32, 2019.
53. Cédric Villani. Optimal transport: old and new, Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], volume 338. Springer-Verlag, Berlin, <https://doi.org/10.1007/978-3-540-71050-9>, 2009.
54. Cédric Villani. Topics in optimal transportation, Graduate studies in mathematics, volume 58. American Mathematical Society, Providence, RI, <https://doi.org/10.1090/gsm/058>, 2003.
55. Lei Wang, Faisal Beg, Tilak Ratnanather, Can Ceritoglu, Laurent Younes, John C Morris, John G Csernansky, and Michael I Miller. Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type. *IEEE transactions on medical imaging*, 26(4):462–470, 2007.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Shreya Arya¹ · Arnab Auddy² · Ranthony A. Clark³ · Sunhyuk Lim⁴ · Facundo Mémoli⁵ · Daniel Packer⁵

✉ Facundo Mémoli
facundo.memoli@gmail.com

Shreya Arya
smarya@upenn.edu

Arnab Auddy
auddy.1@osu.edu

Ranthony A. Clark
ranthony.clark@duke.edu

Sunhyuk Lim
lsh3109@skku.edu

Daniel Packer
daniel.the.packer@gmail.com

¹ Department of Mathematics, University of Pennsylvania, 209 S 33rd St, Philadelphia, PA 19104, USA

² Department of Statistics, The Ohio State University, 1958 Neil Ave, Columbus, OH 43210, USA

³ Department of Mathematics, Duke University, 120 Science Dr, Durham, NC 27710, USA

⁴ Department of Mathematics, Sungkyunkwan University, Suwon-si, Gyeonggi-do 16419, Republic of Korea

⁵ Department of Mathematics, The Ohio State University, 231 W 18th Ave, Columbus, OH 43210, USA