Found Comput Math (2011) 11:417–487
DOI 10.1007/s10208-011-9093-5

FOUNDATIONS OF
COMPUTATIONAL
MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

Gromov–Wasserstein Distances and the Metric Approach to Object Matching

Facundo Mémoli

Received: 20 August 2010 / Accepted: 21 February 2011 / Published online: 30 April 2011

Abstract This paper discusses certain modifications of the ideas concerning the Gromov–Hausdorff distance which have the goal of modeling and tackling the practical problems of object matching and comparison. Objects are viewed as metric measure spaces, and based on ideas from mass transportation, a Gromov–Wasserstein type of distance between objects is defined. This reformulation yields a distance between objects which is more amenable to practical computations but retains all the desirable theoretical underpinnings. The theoretical properties of this new notion of distance are studied, and it is established that it provides a strict metric on the collection of isomorphism classes of metric measure spaces. Furthermore, the topology generated by this metric is studied, and sufficient conditions for the pre-compactness of families of metric measure spaces are identified. A second goal of this paper is to establish links to several other practical methods proposed in the literature for comparing/matching shapes in precise terms. This is done by proving explicit lower bounds for the proposed distance that involve many of the invariants previously reported by researchers. These lower bounds can be computed in polynomial time. The numerical implementations of the ideas are discussed and computational examples are presented.

Keywords Gromov–Hausdorff distances · Gromov–Wasserstein distances · Data analysis · Shape matching · Mass transport · Metric measure spaces

Communicated by Peter Olver.

F. Mémoli (✉)
Department of Mathematics, Stanford University, California, USA
e-mail: memoli@math.stanford.edu
url: http://math.stanford.edu/~memoli/

Mathematics Subject Classification (2000) Primary 68T10 · Secondary 53C23 · 54E35 · 60D05 · 68U05

Contents

1 Introduction

Recent years have seen great advances in the fields of data acquisition and shape modeling, and in consequence, huge amounts of data and large collections of digital models have been obtained. With the goal of organizing and analyzing these collections, it is very important to be able to define and compute meaningful notions of *similarity* between shapes and datasets that exhibit invariance or insensitivity to different poses of the shapes or deformations of the objects represented by the data. Problems of this nature arise in areas such as molecular biology, metagenomics, face recognition, matching of articulated objects, computer graphics, graph matching, and pattern recognition in general. Many of the techniques used for matching and measuring dissimilarity of shapes and datasets are common, thus in the sequel, we will use the word *objects* to denote both datasets and/or shapes.

Advances in the technology that allow the acquisition and storage of massive protein data [35, 90] suggest that a particularly relevant example nowadays is that of geometrically matching proteins. It is widely believed that *geometrically similar* proteins will likely have similar functional properties [67]. Here, 'geometrically similar' refers to the existence of a Euclidean isometry that maps one protein to a small neighborhood of the other.

Another area in which the problem of matching objects is relevant is *brain warping*, a sub-field of *computational anatomy*. Computational anatomy [49] arises in medical imaging, in particular neuro-imaging, and it involves comparison of the shape of anatomical structures between two individuals [45].

After obtaining the individual representation of the structure under consideration for each of the subjects, the main component in the analysis is the establishment of precise correspondence between (anatomically) homologous substructures across subjects.

In computer vision and computer graphics there are applications which are of a different nature in that in these one searches large databases of objects for those that resemble, in some way, a given query object. These applications may tolerate a larger rate of false positives and they do not usually require that precise correspondences between sub-parts of the objects be produced.

1.1 The proposed approach

The philosophy that is proposed in this paper can be summarized as follows:

- Firstly, one must choose a flexible *representation of objects*.
- Secondly, decide on a notion of equality or *isomorphism* between objects.
- Thirdly, put a (pseudo) *metric* on the collection of all isomorphism classes of objects, thus making the space of all objects into a metric space. This metric must be such that it two objects will be at distance zero if and only if they are equal in the sense of equality that was chosen.

The representation of objects that is ultimately chosen is that of *metric measure spaces*; equality of two objects will be *isomorphism of metric measure spaces*; and the metric between objects will be the *Gromov–Wasserstein* distance. Metric measure spaces are metric spaces that have been enriched with a (fully supported) probability measure which serves the purpose of signaling the relative importance of different parts of the shapes.

Following Gromov's ideas, this paper advocates studying the space of all objects as a metric space in its own right. This type of setting has been proven extremely useful in the theoretical literature where, starting by Gromov's compactness theorem, one is able to study families of metric spaces and their metric and topological properties [52].

One of the theses of this paper is that this formalism allows for a precise description of certain properties that are desirable from data/shape matching procedures. Furthermore, in this paper we show how expressing the problems of object matching in these metric terms allows bringing many different, apparently disconnected, previously proposed procedures into a common landscape.

The representation of objects as metric measure spaces is natural and flexible and permits modeling problems that originate in different disciplines. Indeed, in many situations, datasets obtained experimentally are readily endowed with a notion of distance between their points, which turns them into metric spaces. Summaries of the information contained in such datasets often take the form of *metric invariants*, which is important to study and characterize since the analysis of these invariants can provide insights about the nature of the underlying phenomenological science producing the data. With this formulation one is able to express in precise terms the notion that these invariants behave well under deformations or perturbations of the objects, something of interest not only in object matching, but also in other applications such as clustering, and in more generality, *data analysis*.

In problems in *shape analysis*, one can also frequently come up with notions of distance between points on an object that are useful for a given application. One example is matching objects under invariance to rigid transformations, in which the natural choice is to endow objects with the distance induced from their embedding in Euclidean space.

In addition to the metric that must be specified in order to be able to regard a given object as a metric space, the proposal in this paper also requires that users specify a *probability measure*. Without this probability measure one could directly try to use the Gromov–Hausdorff distance [80, 81] for measuring dissimilarity between two objects. It turns out, unfortunately, that the practical computation of the Gromov–Hausdorff distance is highly non-trivial [75] and practical procedures for estimating it have been developed only in the context of objects satisfying certain smoothness conditions [17, 81].

A first contribution of this paper is that it presents new results, which extend the original definition of the Gromov–Hausdorff distance in a way such that the associated discrete problems one needs to solve in practical applications are of an easier nature than yielded by previous related approaches [17, 81]. Further, the computational techniques proposed in the present paper are applicable without restriction on the nature of the objects: objects are not required to be smooth, and the methods of this paper can be used, for example, for comparing phylogenetic trees or ultrametric spaces. This extension of the notion of Gromov–Hausdorff distance which operates on metric measure spaces is called the Gromov–Wasserstein distance, and is based on concepts from *mass transportation* [111].

A second contribution of this paper is that this change of perspective allows one to recognize that many invariants previously used by applied researchers in *signature based methods* for shape matching can be proved to be *quantitatively stable* under the new notion of distance between objects that is constructed in this paper. These statements about stability obviously translate into *lower bounds* for this notion of distance. In turn, these lower bounds lead to optimization problems which are less computationally demanding than the estimation of the full shape distance.

Some of the results in this paper have been announced in [75].

1.2 Related Work

Many approaches have been proposed in the context of (pose invariant) shape classification and recognition, including the pioneering work on *size theory* by Frosini and collaborators [38], where the authors already propose a certain formalization of the problem of shape matching.

Measuring shape variation is an important problem in biostatistics, and an early reference for this topic is the work of D'Arcy Thompson [107]. In the statistical literature there have been many contributions to the understanding of how shape variation may be modeled and quantified [8–10, 25, 63, 68, 103], giving rise to the concept of *shape spaces*. This concept is related to the work of Grenander on *deformable templates* [47, 48], which provides another mathematical formalism for certain problems dealing with shape deformation.

The use of formalisms from differential geometry for the study of shape deformation has made inroads into areas dealing with applied problems such as object recognition, target detection and tracking, classification of biometric data, and automated medical diagnostics. Two of the most prominent ideas in this research area have been the use of fluid flow [28, 29, 108] and elasticity notions [98, 113, 114]. One domain of application of these methods is computational anatomy, where finding precise correspondence between homologous anatomical structures is often the goal. These developments have led to building *Riemannian structures* on many shape spaces [82–84, 115]. We should refer the reader to the recent book by Younes [112]. The present work differs from these approaches in that one builds *metric structures* on the collection of all objects, instead of the more specialized Riemannian structures.

There have been many other approaches to the problem of (pose invariant) object matching and recognition in the computer graphics and computer vision literatures: see [109, 110] for an overview. In many cases, the underlying idea revolved around the computation and comparison of certain invariants or signatures of the objects in order to ascertain whether two objects are in fact the *same object* (up to a certain notion of invariance). An overview of such methods (often referred to as *feature* or *signature* based methods) can be found in [23]. A few examples are: the *size theory* of Frosini and collaborators [33, 34, 38–40]; the Reeb graph approach of Hilaga et al. [54]; the *spin images* of Johnsson [60]; the *shape distributions* of [86]; the *canonical forms* of [37]; the Hamza–Krim approach [53]; the spectral approaches of [94, 100]; the *integral invariants* of [30, 74, 91]; the *shape contexts* of [4]. Other examples are [3, 5, 31, 62, 69, 70, 80, 94, 96].

Signature based methods are often motivated by the need for fast algorithms for detecting similarity between objects and do not always provide a correspondence between objects. The domain of application is usually very different from methods such as those used in computational anatomy, where one already knows that homologous substructures between objects exist (i.e., one is matching brain cortices against brain cortices), as opposed to matching objects in a database of CAD models where one may have to find all the chairs in a database that also contains objects very different from chairs, such as planes, fridges, cars, etc. Here one should single out for example [37, 53, 54, 94, 100], where the type of deformations that these methods were originally intended to accommodate are *intrinsic isometries*: namely, objects in this case are modeled as surfaces, and one tolerates deformations of the objects that leave geodesic distances unchanged.

Recognizing that the structure that had been used and exploited in many of the preceding examples was nothing but the metric structure on the objects, [78, 80, 81] proposed viewing the practical problem of comparing objects under certain deformations as that of comparing metric spaces with the Gromov–Hausdorff distance. The ideas put forward [78, 80, 81] found application in the hands of Kimmel's group at the Technion who carried out several experimental studies and developments in the following years [13–15, 17–19, 92].

On a different vein, the work of Gu and collaborators has emphasized deformations more general than intrinsic isometries, namely their methods can accommodate shape analysis under *conformal deformations* [58, 59, 116, 117]. They have also explored extremely interesting connections with subjects such as the Ricci flow and Teichmüller space.

1.3 Organization of the Paper

This paper is organized as follows. Section 1.4 presents the basic background concepts and the notation used in the paper, which is also summarized in the notation key, Sect. 11. Section 2 introduces the problem of Object Comparison in a general setting and presents basic elements such as notions of object similarity upon which the rest of the paper is based. Section 3 discusses the idea of introducing invariances into standard notions of similarity and explains the motivation for the metric space point of view emphasized in later sections. Section 4 reviews the notion of Gromov–Hausdorff distance and its main properties. In that section we also discuss connections with the Quadratic Assignment Problem. Section 5 delves into the core of the paper where the construction of the Gromov–Wasserstein distance is carried out. In that section we first introduce several isomorphism invariants of metric measure spaces, study their relative strength and illustrate their discriminative power. We do the latter by providing, for each of the invariants that we treat, explicit constructions of pairs of non-isomorphic metric measure spaces that have the same value of the invariant. We also explain how these invariants are related to some of the pre-existing object matching techniques. In Sect. 5.3 we find that two possible constructions of a Gromov–Wasserstein distance are possible, one of them having been analyzed by Sturm [105]. We then establish the theoretical properties of the previously unreported distance, study the Hölder equivalence with Sturm's notion in certain classes of smooth shapes, and establish connections with the Gromov–Hausdorff distance and other notions. In Sects. 5.5 and 5.6 we study in more detail the topology generated by the construction of the Gromov–Wasserstein distance that we adopt. Some basic lower and upper bounds for the Gromov–Wasserstein distances are presented in that section. Section 6 presents other more interesting lower and upper bounds for the proposed notion of distance. The goal is twofold: on one hand establishing these bounds makes apparent the connection to other approaches found in the literature, and on the other it provides lower bounds which are easily computable and consequently of practical value.

Section 7 discusses the computational aspect of the ideas, establishing that the problems one needs to solve in practice are either linear or quadratic optimization problems (with linear constraints on continuous variables). In this section we also argue that the construction of a Gromov–Wasserstein distance that we propose in this paper leads to somewhat easier optimization problems, and is therefore preferable in practical applications. We present computational illustrations in Sect. 8, and conclusions in Sect. 9.

To enhance readability, the proofs of most results are not given in the main text and are instead deferred to Sect. 10.

1.4 General Background Concepts and Notation

Throughout the presentation we use some simple concepts from measure theory and point set topology which can be consulted, for example, in [36]. We touch upon some mass transportation concepts, which can be consulted in the very friendly and up to date presentation of [111].

For each $n \in \mathbb{N}$, we denote by Π_n the set of all permutations of $\{1, 2, \dots, n\}$.

Given two metric spaces (X, d_X) and (Y, d_Y), a map $\varphi : X \to Y$ is an *isometry* if φ is surjective and $d_X(x, x') = d_Y(\varphi(x), \varphi(x'))$ for all $x, x' \in X$. If X and Y are such that there exists an isometry between them, then we say that X and Y are *isometric*. A map φ is an *isometric embedding* if it is an isometry onto its image. Given $\varepsilon > 0$, a ε-*net* for the metric space (X, d_X) is a subset $S \subset X$ s.t. for all $x \in X$ there exists $s \in S$ with $d_X(x, s) \le \varepsilon$.

Given a metric space (X, d_X) by a measure on X we will mean a measure on $(X, \mathcal{B}(X))$ where $\mathcal{B}(X)$ is the Borel σ-algebra of X.

Given measurable spaces $(X, \mathcal{B}(X))$ and $(Y, \mathcal{B}(Y))$ with measures μ_X and μ_Y, respectively, let $\mathcal{B}(X \times Y)$ be the σ-algebra on $X \times Y$ generated by subsets of the form $A \times B$ where $A \in \mathcal{B}(X)$ and $B \in \mathcal{B}(Y)$. The *product measure* $\mu_X \otimes \mu_Y$ is defined to be the unique measure on $(X \times Y, \mathcal{B}(X \times Y))$ s.t. $\mu_X \otimes \mu_Y(A \times B) = \mu_X(A)\mu_Y(B)$ for all $A \in \mathcal{B}(X)$ and $B \in \mathcal{B}(Y)$. For $x \in X$, δ_x^X denotes the Dirac measure on X.

For a measurable map $f : X \to Y$ between two compact metric spaces X and Y, and μ a measure on X, the *push-forward measure* $f_\# \mu$ on Y is given by $f_\# \mu(A) := \mu(f^{-1}(A))$ for $A \in \mathcal{B}(Y)$.

Given a metric space (X, d), a Borel measure ν on X, a function $f : X \to \mathbb{R}$, and $p \in [1, \infty]$ we denote by $\|f\|_{L^p(\nu)}$ the L^p norm of f with respect to the measure ν.

Remark 1.1 When ν is a probability measure, i.e., $\nu(X) = 1$, then $\|f\|_{L^p(\nu)} \ge \|f\|_{L^q(\nu)}$ for $p, q \in [1, \infty]$ and $p \ge q$.

A special class of metric spaces are those that arise from Riemannian manifolds. We denote by **Riem** the collection of all compact Riemannian manifolds. Given $(X,g)\in\mathbf{Riem}$, one regards X as a metric space with the *intrinsic metric* arising from the metric tensor g, see [101].

A *probability measure* on the measurable space $(X,\mathcal{B})$ is any measure μ on X s.t. $\mu(X)=1$. We now recall some concepts about convergence of probability measures on metric spaces. An excellent reference is [36, Chap. 11]. A family M of probability measures on the metric space X is called *tight* if for any $\varepsilon>0$ there exists a compact $K_\varepsilon\subset X$ s.t. $\mu(K_\varepsilon)>1-\varepsilon$ for all $\mu\in M$.

For a metric space (Z,d) let $\mathcal{P}(Z)$ denote the set of all Borel probability measures on Z. By $C_b(Z)$ we denote the set of all continuous and bounded real valued functions on Z. We say that a sequence $\{\mu_n\}_{n\in\mathbb{N}}\in\mathcal{P}(Z)$ *converges weakly* to $\mu\in\mathcal{P}(Z)$, and write $\mu_n\xrightarrow{w,n}\mu$, if and only if for all $f\in C_b(Z)$, $\int_Z f\,d\mu_n\longrightarrow\int_Z f\,d\mu$ as $n\uparrow\infty$.

2 Comparing Objects

An *object* in a compact metric space (Z,d) will be a compact subset of Z. Let $\mathcal{C}(Z)$ denote the collection of all compact subsets of Z (objects). Assume that inside the metric space (Z,d) one is trying to compare objects A and B. One possibility is to use the Hausdorff distance:

$$d_{\mathcal{H}}^Z(A,B):=\max\Big(\sup_{a\in A}\inf_{b\in B}d(a,b),\sup_{b\in B}\inf_{a\in A}d(a,b)\Big). \tag{2.1}$$

In general, whenever one intends to compare two objects, a correspondence/alignment is established for this purpose. Definition 2.1 and Proposition 2.1 make this apparent for the case of the Hausdorff distance.

Definition 2.1 (Correspondence) For non-empty sets A and B, a subset $R\subset A\times B$ is a *correspondence* (between A and B) if and only if

- $\forall a\in A$ there exists $b\in B$ s.t. $(a,b)\in R$,
- $\forall b\in B$ there exists $a\in A$ s.t. $(a,b)\in R$.

Let $\mathcal{R}(A,B)$ denote the set of all possible correspondences between sets A and B.

Example 2.1 Let X and Y be non-empty sets; then $\mathcal{R}(X,Y)$ is also non-empty, as it contains $X\times Y$.

Example 2.2 Let $X\in\mathcal{C}(Z)$ and $p\in Z$. When $Y=\{p\}$, $\mathcal{R}(X,Y)=\{X\times\{p\}\}$.

Example 2.3 Let $X=\{a,b\}$ and $Y=\{A,B,C\}$. In this case, $R_1=\{(a,A),(b,B),(a,C)\}$ is a correspondence but $R_2=\{(a,A),(b,B)\}$ is not.

Example 2.4 (Correspondences as 0, 1 matrices) Let X and Y be finite. Then, $\mathcal{R}(X,Y)$ can be represented by the set of matrices with 0 or 1 entries that satisfy the following constraints: Let $X = \{x_1, \ldots, x_{n_X}\}$ and $Y = \{y_1, \ldots, y_{n_Y}\}$. Then, consider $\widehat{\mathcal{R}}(X, Y)$ to be the set of $n_X \times n_Y$ matrices with entries 0 or 1 s.t. for all $((r_{i,j})) \in \widehat{\mathcal{R}}(X, Y)$:

$$\sum_{i=1}^{n_X} r_{i,j} \geq 1 \quad \text{for } j \in \{1, \ldots, n_Y\} \quad \text{and} \quad \sum_{j=1}^{n_Y} r_{i,j} \geq 1 \quad \text{for } i \in \{1, \ldots, n_X\}. \tag{2.2}$$

Consider the bijections $X \longleftrightarrow \{1, \ldots, n_X\}$ and $Y \longleftrightarrow \{1, \ldots, n_Y\}$ given by $x_i \leftrightarrow i$ and $y_j \leftrightarrow j$ for all $i \in \{1, \ldots, n_X\}$ and $j \in \{1, \ldots, n_Y\}$. Let $T : \mathcal{R}(X, Y) \to \widehat{\mathcal{R}}(X, Y)$ be given $R \mapsto ((r_{i,j}))$ where $r_{i,j} = 1$ if $(x_i, y_j) \in R$ and $r_{i,j} = 0$ otherwise. Then, actually, T provides a bijection. Frequently, by abuse of notation, $\mathcal{R}$ will be identified with $\widehat{\mathcal{R}}$.

Example 2.5 Let X and Y be finite s.t. $|X| = |Y| = n$. In this case, $\pi \in \widehat{\mathcal{R}}(X, Y)$ for all permutation matrices π of size n.

The following non-standard expression for the Hausdorff distance will be useful.

Proposition 2.1 *Let (Z, d) be a compact metric space. Then the Hausdorff distance between any two sets $A, B \subset Z$ can be expressed as*

$$d_{\mathcal{H}}^Z(A, B) = \inf_R \sup_{(a,b) \in R} d(a, b) \tag{2.3}$$

where the infimum is taken over all $R \in \mathcal{R}(A, B)$.

Proof Let $\varepsilon > 0$ and $R \in \mathcal{R}(A, B)$ be s.t. $d(a, b) < \varepsilon$ for all $(a, b) \in R$. Since R is a correspondence between A and B it follows that $\inf_{b \in B} d(a, b) < \varepsilon$ for all $a \in A$ and $\inf_{a \in A} d(a, b) < \varepsilon$ for all $b \in B$. Recalling (2.1) it follows that $d_{\mathcal{H}}^Z(A, B) \leq \varepsilon$.

Assume now that $d_{\mathcal{H}}^Z(A, B) < \varepsilon$. Then, for each $a \in A$ there exist $b \in B$ s.t. $d(a, b) < \varepsilon$. Then, one may define $\phi : A \to B$ s.t. $d(a, \phi(a)) < \varepsilon$ for all $a \in A$. Similarly, define $\psi : B \to A$ s.t. $d(\psi(b), b) < \varepsilon$ for all $b \in B$. Consider $R = \{(a, \phi(a), a \in A)\} \cup \{(\psi(b), b), b \in B\}$. Obviously, $R \in \mathcal{R}(A, B)$ and by construction $d(a, b) < \varepsilon$ for all $(a, b) \in R$. □

The Hausdorff distance is indeed a metric on the set of compact subsets of the (compact) metric space (Z, d).

Proposition 2.2 [22, Proposition 7.3.3]

(1) *If $A, B, C \in \mathcal{C}(Z)$ then*

$$d_{\mathcal{H}}^Z(A, B) \leq d_{\mathcal{H}}^Z(A, C) + d_{\mathcal{H}}^Z(B, C).$$

(2) *If $d_{\mathcal{H}}^Z(A, B) = 0$ for $A, B \in \mathcal{C}(Z)$ then $A = B$.*

The quality of the approximation of an object A by a finite set $\mathbb{A}_n$ can obviously be described by $d_{\mathcal{H}}^Z(A,\mathbb{A}_n)$. Indeed, assume $\mathbb{A}_n\subset A$; then $d_{\mathcal{H}}^Z(A,\mathbb{A}_n)$ can be interpreted as the *minimal covering radius* associated to covering A with balls centered at points in $\mathbb{A}_n$. This can be made precise by recalling (2.1) and using the fact that $\mathbb{A}_n\subset A$:

$$d_{\mathcal{H}}^Z(A,\mathbb{A}_n)=\inf\left\{\varepsilon>0\,\middle|\,A\subset\bigcup_{a\in\mathbb{A}_n}B_Z(a,\varepsilon)\right\}. \tag{2.4}$$

Remark 2.1 (One argument in favor of a metric structure on objects) In practice, metric properties are desirable, and one should insist on having (appropriate versions of) them for whichever notion of similarity between objects one chooses to work with. In the case of the Hausdorff distance, these properties imply, in particular, that if one is interested in comparing objects A and B, and if $\mathbb{A}_n\subset Z$ and $\mathbb{B}_m\subset Z$ are finite (maybe 'noisy'[1]) samples of A and B, respectively, then

$$\left|d_{\mathcal{H}}^Z(A,B)-d_{\mathcal{H}}^Z(\mathbb{A}_n,\mathbb{B}_m)\right|\le d_{\mathcal{H}}^Z(A,\mathbb{A}_n)+d_{\mathcal{H}}^Z(B,\mathbb{B}_m). \tag{2.5}$$

One can interpret this expression as follows. There are two objects A and B for which one desires to obtain a measure of their dissimilarity using Hausdorff distance, but one does not have access to them except via sampling. That is, there is a procedure that allows one to take an increasing number of sample points from the objects, perhaps at a certain cost. Hence, in practice, one always has to rely on finite samples $\mathbb{A}_n$ and $\mathbb{B}_m$ of the two objects. Inequality (2.5) above says that by computing the Hausdorff distance between the samples we are able to approximate the Hausdorff distance between the true objects up to an error given by how well the finite samples represent the A and B.

Invoking (2.4), one sees that (2.5) expresses the fact that comparing these finite samples gives an answer as good as the approximation of the underlying objects by these discrete sets. This can be interpreted as *consistency* and/or *stability*.

Finally, one has the following.

Theorem 2.1 (Blaschke, Theorem 7.3.8 [22]) *If* (Z,d) *is compact, then the collection of all objects* $(\mathcal{C}(Z),d_{\mathcal{H}}^Z)$ *is also compact.*

The ideas proposed in this paper rely on the idea of relaxing the notion of correspondence as given by Definition 2.1. In order to do this it is necessary to introduce a new class of objects denoted by $\mathcal{C}_w(Z)$ (and defined precisely below). Informally speaking, an object in this class will be specified by not only the set of points that constitute it, but one is also required to specify a distribution of importance over these points. This relaxed notion of correspondence between objects is called *matching measure* (or coupling) and is made precise in Definition 2.3 below.

There is a very well-known family of distances which makes use of this alternative way of pairing objects. These are the so called Mass Transportation distances

[1] By 'noisy' it is understood that it is not necessary that $\mathbb{A}_n\subset A$ and $\mathbb{B}_m\subset B$.

[111] (also known as Wasserstein–Kantorovich–Rubinstein distances as known in the Math community or Earth Mover's distance [85, 96] in the Object Recognition arena). These concepts are reviewed next.

2.1 Mass Transportation Distances

Assume as before that $A, B \in \mathcal{C}(Z)$, and, let μ_A and μ_B be Borel probability measures with *supports* A and B, respectively. Informally, this means that if a set $C \subset Z$ is s.t. $C \cap A = \emptyset$ then $\mu_A(C) = 0$. A precise definition is given below:

Definition 2.2 (Support of a measure) The support of a measure μ on a metric space (Z, d), denoted by $\mathrm{supp}[\mu]$, is the minimal closed subset $Z_0 \subset Z$ such that $\mu(Z \backslash Z_0) = 0$.

These probability measures can be thought of as acting as *weights* for each point in each of the sets. A simple interpretation is that for each $a \in A$, $r > 0$, $\mu_A(B_Z(a, r))$ quantifies the (relative) importance of the point a at scale r (in the discrete case this measure can clearly be interpreted as signaling how much one trusts the sample point). In other words, if a' is another point in A and if $\mu_A(B_Z(a, r)) \leq \mu_A(B_Z(a', r))$ one would say that a' *is more important than* a at scale r. Note that since μ_A (respectively, μ_B) is a probability measure, $\mu_A(A) = 1$ (respectively, $\mu_B(B) = 1$).

One naturally requires that $A = \mathrm{supp}[\mu_A]$ and $B = \mathrm{supp}[\mu_B]$. By taking μ_A and μ_B into account one will, therefore, be comparing not only the geometry of the sets, but also, the distribution of "importance" over the sets. Define the collection of all *weighted objects* (in the metric space (Z, d)),

$$\mathcal{C}_w(Z) := \big\{(A, \mu_A),\ A \in \mathcal{C}(Z)\big\},$$

where for each $A \in \mathcal{C}(Z)$, μ_A is a Borel probability measure with $\mathrm{supp}[\mu_A] = A$.

Definition 2.3 (Matching measure) Let $A, B \in \mathcal{C}_w(Z)$. A measure μ on the product space $A \times B$ is a *matching measure* or *coupling* of μ_A and μ_B iff

$$\mu(A_0 \times B) = \mu_A(A_0) \quad \text{and} \quad \mu(A \times B_0) = \mu_B(B_0) \tag{2.6}$$

for all Borel sets $A_0 \subset A$, $B_0 \subset B$. Denote by $\mathcal{M}(\mu_A, \mu_B)$ the set of all couplings of μ_A and μ_B.

Example 2.6 Let X and Y be objects in $\mathcal{C}_w(Z)$; then $\mathcal{M}(\mu_X, \mu_Y)$ is non-empty, as it always contains the product measure $\mu_X \otimes \mu_Y$.

Example 2.7 Pick $X \in \mathcal{C}_w(Z)$. When $Y = \{z_0\}$ for some $z_0 \in Z$, $\mu_Y = \delta^Y_{y_0}$ is a Dirac delta. In this case there is only one measure coupling: $\mathcal{M}(\mu_X, \mu_Y) = \{\mu_X \otimes \delta^Y_{y_0}\}$.

Remark 2.2 (Matching measures between finite spaces) When $X, Y \in \mathcal{C}_w(Z)$ are finite, say $n_X = |X|$ and $n_Y = |Y|$, then $\mathcal{M}(\mu_X, \mu_Y)$ is composed of matrices with non-negative entries of size $n_X \times n_Z$ satisfying $n_X + n_Y$ *linear* constraints:

$$\sum_{x\in X} \mu(x, y) = \mu_Y(y) \quad \text{for } y \in Y \quad \text{and} \quad \sum_{y\in Y} \mu(x, y) = \mu_X(x) \quad \text{for } x \in X.$$

Example 2.8 Let $X = \{x_1, x_2\}$ and $\mu_X(x_1) = \mu_X(x_2) = 1/2$. Let $Y = \{y_1, y_2, y_3\}$ and $\mu_Y(y_1) = \mu_Y(y_2) = \mu_Y(y_3) = 1/3$. Let

$$\mu_{\mathrm{I}} = \begin{array}{c} \\ x_1 \\ x_2 \end{array} \!\!\begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{pmatrix} 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 \end{pmatrix} \end{array}, \qquad \mu_{\mathrm{II}} = \begin{array}{c} \\ x_1 \\ x_2 \end{array} \!\!\begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{pmatrix} 2/9 & 1/6 & 1/9 \\ 1/9 & 1/6 & 2/9 \end{pmatrix} \end{array},$$

$$\mu_{\mathrm{III}} = \begin{array}{c} \\ x_1 \\ x_2 \end{array} \!\!\begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{pmatrix} 1/9 & 1/6 & 2/9 \\ 1/9 & 1/6 & 2/6 \end{pmatrix} \end{array}.$$

Then $\mu_{\mathrm{I}}, \mu_{\mathrm{II}} \in \mathcal{M}(\mu_X, \mu_Y)$ but $\mu_{\mathrm{III}} \notin \mathcal{M}(\mu_X, \mu_Y)$.

It turns out that for each $\mu \in \mathcal{M}(\mu_A, \mu_B)$, $\mathrm{supp}[\mu] \subset A \times B$ is a *correspondence* which we denote by $R(\mu)$. This is proved below.

For $z = (a, b) \in A \times B$ let $\pi_A(z) = a$ and $\pi_B(z) = b$ denote the coordinate projections of z. For $K \subset A \times B$ let $\pi_A(K) = \{\pi_A(z) | z \in K\}$ and $\pi_B(K) = \{\pi_B(z) | z \in K\}$.

Lemma 2.1 (König's Lemma) *Let $E \subset A \times B$ be a closed set. Then the following condition is necessary and sufficient for the existence of $\mu \in \mathcal{M}(\mu_A, \mu_B)$ such that $E = supp[\mu]$: for all $A_0 \in \mathcal{B}(A)$ the set $(B \supset) P_B(A_0) := \pi_B((A_0 \times B) \cap E)$ satisfies*

$$\mu_B\big(P_B(A_0)\big) \geq \mu_A(A_0).$$

For this lemma see [52] and also [72, §2.5]. The lemma can be obtained as a special case of [104, Theorem 11].

Lemma 2.2 *Let μ_A and μ_B be Borel probability measures on (Z, d), a compact space. If $\mu \in \mathcal{M}(\mu_A, \mu_B)$, then $R(\mu) :=$ supp$[\mu]$ belongs to $\mathcal{R}($supp$[\mu_A]$, supp$[\mu_B])$.*

Remark 2.3 (The finite case) It is particularly easy to argue in the case A and B are finite and $A = \mathrm{supp}[\mu_A]$ and $B = \mathrm{supp}[\mu_B]$. Indeed, in these case let $\mu \in \mathcal{M}(\mu_A, \mu_B)$ be given in its matricial representation as in Remark 2.2. Consider the matrix $R(\mu) = ((r_{ab}))$ where $r_{ab} = 1$ is $\mu(a, b) > 0$ and $r_{ab} = 0$ otherwise. Fix $a \in A$. Then, since by hypothesis $0 < \mu_A(a) = \sum_{b\in B} \mu(a, b)$, one must have $r_{ab} = 1$ for at least one $b \in B$. Similarly, for any $b \in B$, $r_{ab} = 1$ for at least one $a \in A$. This implies via Example 2.4 that $R(\mu)$ is a correspondence between A and B.

Proof of Lemma 2.2 Let us see that $R(\mu)$ is indeed a correspondence. We will prove that for any $a \in \mathrm{supp}[\mu_A]$ there exists $b \in \mathrm{supp}[\mu_B]$ such that $(a, b) \in R(\mu)$.

Indeed, pick $a \in \mathrm{supp}[\mu_A]$ and for $\varepsilon > 0$ let $F_\varepsilon(a) := P_B(\overline{B}(a,\varepsilon))$ and $F_0(a) = \{b \in B \mid (a,b) \in R(\mu)\}$. Notice that $F_\varepsilon(a) = \{b' \in B \mid (a',b') \in R(\mu)$, for some $a' \in \overline{B}(a,\varepsilon)\}$. It is enough to prove that $F_0(a) \neq \emptyset$.

Since $a \in \mathrm{supp}[\mu_A]$, then $\mu_A(\overline{B}(a,\varepsilon)) > 0$ for all $\varepsilon > 0$. Then, according to Lemma 2.1, the set $F_\varepsilon(a)$ has $\mu_B(F_\varepsilon(a)) > 0$ and hence $F_\varepsilon(a) \neq \emptyset$ for $\varepsilon > 0$. Also, it is easy to see that if $\varepsilon' < \varepsilon$ then $F_{\varepsilon'}(a) \subset F_\varepsilon(a)$. Also, all the sets $F_\varepsilon(a)$ are closed. Since $\mathrm{supp}[\mu_A]$ is a closed set inside the compact space A, then $\mathrm{supp}[\mu_A]$ is compact.

Consider a sequence $\{\epsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^+$ s.t. $\lim_k \epsilon_k = 0$. By construction, F_{ϵ_1} is closed and one has that $\{F_{\epsilon_k}\}_k$ is a nested family of non-empty closed sets. Then, their intersection $\bigcap_{k\geq 1} F_{\epsilon_k} = \lim_{k\to\infty} F_{\epsilon_k}$ is non-empty. Hence, there exists $b \in B$ such that $(a,b) \in R(\mu)$. By proceeding in the same way, for any $b \in \mathrm{supp}[\mu_B]$ there exist $a \in \mathrm{supp}[\mu_A]$ s.t. $(a,b) \in R(\mu)$ and hence $R(\mu)$ is a correspondence between A and B. □

Definition 2.4 (Wasserstein distances, Chap. 7 of [111]) For each $p \geq 1$ consider the following family of distances on $\mathcal{C}_w(Z)$, where (Z,d) is a compact metric space:

$$d^Z_{\mathcal{W},p}(A,B) := \inf_{\mu\in\mathcal{M}(\mu_A,\mu_B)} \left(\int_{A\times B} d^p(a,b)\, d\mu(a,b)\right)^{1/p} \tag{2.7}$$

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

$$d^Z_{\mathcal{W},\infty}(A,B) := \inf_{\mu\in\mathcal{M}(\mu_A,\mu_B)} \sup_{(a,b)\in R(\mu)} d(a,b). \tag{2.8}$$

These distances are none other than the Wasserstein–Kantorovich–Rubinstein distances between measures [36, 44, 111].

These distances have been considered for Object Comparison/Matching applications several times (for some values of p, typically $p = 1$ or 2); see for example [31, 65, 85, 96] and more recently, from a more theoretical point of view, [41].

As in the definition of $d^Z_{\mathcal{W},p}$, in the sequel, by an abuse of notation an object $(A,\mu_A) \in \mathcal{C}_w(Z)$ will also be denoted by either A or μ_A. The reader should keep in mind, however, that a measurable set $A \subset Z$ could be represented by many probability measures; all that is required is that those probability measures have support A.

Example 2.9 (Wasserstein distance between finite objects) Let (Z,d) be a compact metric space. When $A, B \subset Z$ are finite, μ_A and μ_B are linear combinations of delta measures and can be represented as vectors. In particular, (2.7) takes the following form:

$$d^Z_{\mathcal{W},p}(A,B) = \inf_{\mu}\left(\sum_{a,b} d^p(a,b)\,\mu(a,b)\right)^{1/p},$$

where the infimum is taken over all matching measures matrices between μ_A and μ_B, see Remark 2.2. The optimization problem above is a Linear Optimization problem with linear constraints on continuous variables. There exist standard specialized algorithms for numerically computing the minimal value above [88], see Sect. 7.

An initial question, which is now easy to answer, is how Wasserstein distances are related to $d^Z_{\mathcal{H}}(,)$. Upon noting that (2.1) and (2.8) are essentially the same expression and that $R(\mu) \in \mathcal{R}(A,B)$ one obtains

Corollary 2.1 (Relationship between $d_{\mathcal{H}}$ and $d_{\mathcal{W},\infty}$) *For* (A,μ_A) *and* (B,μ_B) *in* $\mathcal{C}_w(Z)$

$$d^Z_{\mathcal{H}}(A,B) \le d^Z_{\mathcal{W},\infty}\big((A,\mu_A),(B,\mu_B)\big)$$

for all choices of μ_A *and* μ_B *such that* $A = \mathrm{supp}[\mu_A]$ *and* $B = \mathrm{supp}[\mu_B]$.

Proof Pick any $\mu \in \mathcal{M}(\mu_A,\mu_B)$. Then, by hypothesis, Lemma 2.2 gives $R(\mu) \in \mathcal{R}(A,B)$. By Proposition 2.1,

$$d^Z_{\mathcal{H}}(A,B) = \inf_{R\in\mathcal{R}(A,B)} \|d\|_{L^\infty(R)} \le \|d\|_{L^\infty(R(\mu))}.$$

The conclusion follows since μ is arbitrary. □

This connection between Hausdorff and Mass Transportation distances has already been pointed out in the robotics literature, see [55].

The main properties of this family of distances are reviewed next.

Proposition 2.3 [44, 111]

(1) *For each* $1 \le p \le \infty$, $d^p_{\mathcal{W},p}$ *defines a metric on* $\mathcal{C}_w(Z)$.
(2) *For any* $1 \le q \le p \le \infty$ *and* $A,B \in \mathcal{C}_w(Z)$

$$d^Z_{\mathcal{W},q}(A,B) \le d^Z_{\mathcal{W},p}(A,B).$$

Distances $d^Z_{\mathcal{W},p}$, for finite p offer an interesting alternative to the Hausdorff distance. Note that, in the finite discrete case, computing them involves solving a Linear Optimization Problem (LOP); see Example 2.9.

Remark 2.4 (Counterpart to Remark 2.1) Assume $\mathbb{A}_n$ and $\mathbb{B}_m$ are finite (possibly noisy, i.e., $\mathbb{A}_n \not\subseteq A$, etc.) samples from A and B, respectively. Assume further that each of them is given together with a discrete probability measure μ_n and ν_n, respectively. Then Property 1 above implies that (cf. (2.5) and Remark 2.1)

$$\big|d^Z_{\mathcal{W},p}(A,B) - d^Z_{\mathcal{W},p}(\mathbb{A}_n,\mathbb{B}_m)\big| \le d^Z_{\mathcal{W},p}(A,\mathbb{A}_n) + d^Z_{\mathcal{W},p}(B,\mathbb{B}_m). \tag{2.9}$$

This is a trivial observation: the point one must make is that now the quality of the approximation of A by its discrete representative $\mathbb{A}_n$ is governed by the Wasserstein–Kantorovich–Rubinstein distance between μ_A and μ_n. In this context, $d^Z_{\mathcal{W},p}(B,\mathbb{B}_n)$ replaces $d^Z_{\mathcal{H}}(B,\mathbb{B}_n)$. For objects (A,μ_A) and (B,μ_B) it is sometimes convenient to abuse the notation by writing $d^Z_{\mathcal{W},p}(\mu_A,\mu_B)$ instead of $d^Z_{\mathcal{W},p}(A,B)$.

Similarly to Blaschke's Theorem, one has the following.

Theorem 2.2 (Prokhorov [111]) *If* (Z,d) *is compact and* $p\geq 1$, *then the collection of all weighted objects* $(\mathcal{C}_w(Z), d_{\mathcal{W},p}^Z)$ *is also compact.*

3 Introducing Invariances

This section provides some more background about the ways in which the purely metric ideas arise in practical considerations dealing with object matching. The presentation will be informal, while we delay the technicalities for Sect. 4 and subsequent sections.

3.1 Extrinsic Similarity

Let $\mathcal{T}_Z$ be the group of isometries of the compact metric space (Z,d), i.e.,

$$\mathcal{T}_Z = \{T : Z \to Z \,|\, \forall z, z' \in Z, d\big(T(z), T\big(z'\big)\big) = d\big(z,z'\big)\}. \tag{3.1}$$

T acts on $A\in\mathcal{C}(Z)$ in the usual way: $T(A)=\{T(a), a\in A\}$. On $(A,\mu_A)\in\mathcal{C}_w(Z)$, the action of $T\in\mathcal{T}_Z$ is given by $T(A,\mu_A)=(T(A), T_\#\mu_A)$. Let $\mathcal{O}(Z)$ denote the choice for the class of objects: either $\mathcal{C}(Z)$ or $\mathcal{C}_w(Z)$. For the case of any well behaved notion of distance D on $\mathcal{O}(Z)$ one can decide to study the following problem: for $A,B\in\mathcal{O}(Z)$ consider

$$D^{\mathcal{T}_Z}(A,B) := \inf_{T\in\mathcal{T}_Z} D\big(A, T(B)\big). \tag{3.2}$$

Obviously this leads to notions of distance between objects that is invariant to *ambient space isometries*. In other words, if A and B are such that $D^{\mathcal{T}_Z}(A,B)$ is sufficiently small, one would say that A and B are *extrinsically similar*. There is of course a complementary notion of *intrinsic similarity* that is described in the rest of this section.

The most common case for the space Z is $(Z,d)=(\mathbb{R}^k, \|\cdot\|)$,[2] which has been approached by several authors in the past, with $D=d_{\mathcal{H}}^{\mathbb{R}^k}(,)$ (and $\mathcal{O}(Z)=\mathcal{C}(Z)$) in [57], and $D=d_{\mathcal{W},p}^{\mathbb{R}^k}$ (and $\mathcal{O}(Z)=\mathcal{C}_w(Z)$) [31, 96] and references therein. In this case, the underlying concept is that of similarity to rigid isometries.

It is in this context that the so called *iterative closest point* algorithm operates [99, 118]. Very similar ideas appear in the context of matching protein structures; see [66, 67] and references therein.

Approaches of this nature typically incur a high computational cost. Whenever possible, one tries to use computationally cheaper alternatives to rule out dissimilar objects, and rely on finer but more costly techniques only once one has reason to believe that the objects are similar (according to the cheaper methods).

[2] Or a compact subset of $\mathbb{R}^k$.

3.2 Comparing Invariants: Signature Based Methods

A different idea, used to a certain extent in the signature based methods [11, 37, 53, 54, 86], consists of computing and comparing invariants to $T \in \mathcal{T}_Z$. This leads to comparing the *metric* information of A and B more directly. These methods are sometimes referred to as *signature based methods*. More precisely, if $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_n\}$ are finite sets of points in $Z \subset \mathbb{R}^k$, *ideally* one would like to meaningfully compare the distance matrices $\mathbf{D}_A = ((\|a_i - a_j\|))$ and $\mathbf{D}_B = ((\|b_i - b_j\|))$ in a fashion compatible with the particular choice of D (and $\mathcal{O}(Z)$) that one has made. By this it is meant that one would hope to come up with a notion of distance $d_{\mathbf{D}}$ between distance matrices such that $d_{\mathbf{D}}(\mathbf{D}_A, \mathbf{D}_B)$ provides a lower bound for $D(A, B)$ in some precise sense.

Researchers have been resorting to comparisons between simple invariants constructed from the distance matrices. For example, in [53] the authors use (essentially) the histograms of row sums of the distance matrices as the invariants they compare. In [86], on the other hand, the authors compare the histograms of all distance values present in $\mathbf{D}_A$ and $\mathbf{D}_B$.

In [51] (under the assumption that $n = m$) the authors propose attaching to each point $a \in A$, the (sorted) vector $V_A(a) \in \mathbb{R}^n$ of distances from a to all other $a' \in A$. They then propose a measure of dissimilarity between A and B which is roughly based on finding $\pi \in \Pi_n$ s.t. $\|V_A - V_B \circ \pi\|_1$ is minimized.

In [4, 74, 91], similar invariants were proposed where to each point $a \in A$ one attaches the histogram or distribution of distances $\|a - a'\|$ for all $a' \in A$.[3]

At any rate, directly comparing the distances matrices associated to two objects A and B amounts to considering $(A, \mathbf{D}_A)$ and $(B, \mathbf{D}_B)$ as metric spaces *without any reference to* Z (which in this case is a compact subset of $\mathbb{R}^k$).

3.3 Intrinsic Similarity

The idea of computing metric invariants and comparing them in order to measure dissimilarity between objects has found applications in situations more general than the ones dealing with extrinsic similarity that have been discussed so far. It was recognized by Hilaga et al. [54] that deformations that may change the Euclidean distance are of interest for certain applications. Hilaga and coauthors deal with the problem of measuring dissimilarity between objects when the geodesic distance between two points may remain approximately constant. The procedure that they propose attaches to each given (triangulated) object the Reeb graph [93] arising from a certain function defined on the object. The function is at each point is defined as a certain average of the geodesic distance to all other points on the object. The dissimilarity value between two objects is defined as a certain notion of dissimilarity between their corresponding Reeb graphs.

[3] We follow the standard usage of the words distribution and histogram: histogram is an estimate of density and distribution is a cumulative version of this estimate.

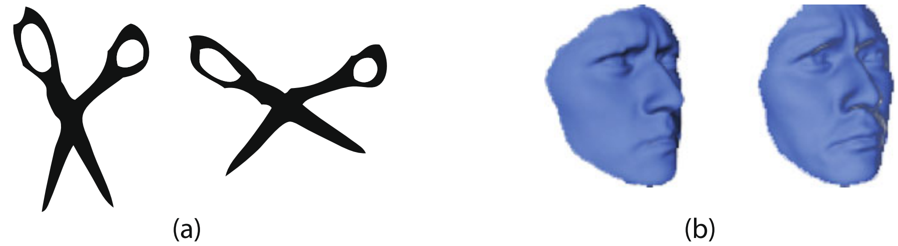

Fig. 1 (**a**) An articulated object in two different articulations. (**b**) Three-dimensional surface representing a human face in two different expressions

A similar idea was used by Hamza and Krim in [53] where instead of looking at the Reeb graph, they attach to each object the histogram of the values of the aforementioned function. The measure of dissimilarity between two objects is obtained by computing a certain notion of distance between probability densities.

3.3.1 The Case of Planar Articulated Objects

In the problem of comparing *articulated objects* [70] the idea of computing $D^{\mathcal{T}_Z}$ does not appear to be useful. Indeed, take for example the scissors in Fig. 1: the two articulations of the scissors cannot be matched with small Hausdorff distance by using any rigid transformation. An example in a similar spirit is given by the problem of comparing faces under different expressions or objects in different poses [37].

In either of these situations, it is more meaningful to compute invariants for metrics other than ambient space metrics. For instance, in [70] the authors endow articulated objects such as those in Fig. 1(a) with a metric coming from the path length distance generated by the restriction of the Euclidean metric to the interior of the object A. In this manner they obtain a distance matrix $\mathbf{D}_A$ which is in general different from the matrix $((\|a_i - a_j\|))$.[4] The authors cogently argue that this metric is (approximately) insensitive to articulations and the procedure that they propose relies on

(1) computing invariants out of the distance matrix $\mathbf{D}_A$, and
(2) comparing these invariants in order to obtain a measure of dissimilarity between two objects.

3.3.2 An Example from 3D Object Recognition

In [16] the authors claimed that one can model the deformations suffered by the surface representing the face under different expressions as those that may change the object in such a manner that geodesic distances remain approximately constant. Some experimental evidence for this claim was presented. Figure 1(b) depicts two expressions of the same human face. It is easy to believe that the Euclidean distance

[4]It is actually, point-wise greater than or equal to. There is equality when the object bounds a convex region of the plane.

between some pairs of points on the face will change noticeably between two different expressions.

The practical procedure of [16] is the same as that of [37]: given a collection of points A sampled from a surface, they computed the matrix $\mathbf{D}_A$ of inter-point geodesic distances. Then, they applied metric MDS (multidimensional scaling) [32] in order to find points A' in some Euclidean space whose inter-point (Euclidean) distances resemble those in $\mathbf{D}_A$ as closely as possible. By mapping into some (possibly high-dimensional) Euclidean space, they unlocked the possibility of computing several standard Euclidean invariants, such as different moments of inertia. Then, they described each object A by a vector I_A of Euclidean invariants of the point sets A', and the comparison of two objects A and B was carried out by computing some norm of the difference of the corresponding vectors of invariants, I_A and I_B.

It is not clear, and to the best of our knowledge not explored by the authors of [37], whether one gains anything in terms of classification error by performing this approximate embedding into some Euclidean space, with respect to the possibility of directly computing invariants out of the geodesic distance matrices $\mathbf{D}_A$ for each face A. It is clear, however, that from the point of view of deciding which invariants to compute, one does indeed gain, given that the moments of an object in Euclidean space offer a complete set of invariants.

In subsequent papers [20, 21], the authors study the possibility of performing a variant of metric MDS where the target space is a spherical space $\mathbb{S}^n$. In this case, the vector of invariants I_A for a given face surface A that the authors propose comes from computing coefficients of the expansion in spherical harmonics of a certain object dependent function $f_A : \mathbb{S}^n \to \mathbb{R}$ that they define.

3.4 A Critique to Signature Based Methods

Although direct comparison of invariants is general enough to accommodate object discrimination under deformations that are not necessarily produced by ambient space isometries, a criticism that applies to all the methods mentioned in this section is that the comparison of invariants does not necessarily lead to a *strict metric* on classes of objects. Constructions such as the ones in [37, 53, 54, 86, 94] propose measures of dissimilarity between objects that by construction are invariant (or approximately invariant) to isometries (be it with respect to Euclidean or geodesic distances) but could yield zero dissimilarity for two non-isometric objects.

What seems to be lacking is a natural language for expressing and reasoning about questions such as the stability and discriminative power of the signature based methods. Furthermore, even though at the intuitive level many of the methods discussed appear to be related, there is no formal, precise understanding of what the relationship could be.

Moreover, a possibility that remained unexplored by the methods above was that of attempting to directly compute a certain notion of distance between distance matrices $\mathbf{D}_A$ and $\mathbf{D}_B$ for objects A and B. We should mention, however, that in the context of protein structure comparison, some ideas regarding the direct comparison of distance matrices can be found for example in [56].

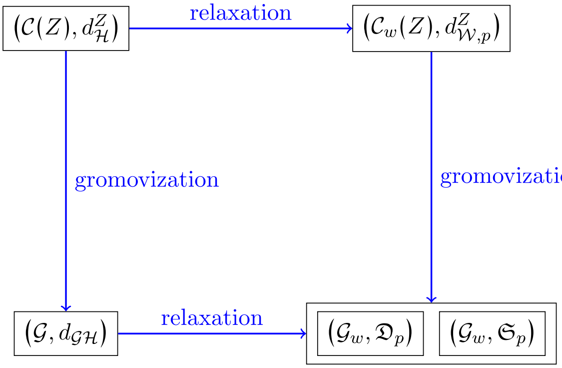

Fig. 2 This figure represents the construction of the metrics $\mathfrak{D}_p$ and $\mathfrak{S}_p$ on $\mathcal{G}_w$, the collection of all metric measure spaces. In this figure, $\mathcal{G}$ is the collection of all compact metric spaces and $d_{\mathcal{GH}}$ the Gromov–Hausdorff distance. For a given compact metric space Z, $\mathcal{C}(Z)$ denotes the collection of all objects in Z (closed subsets of Z). Similarly, $\mathcal{C}_w(Z)$ is the collection of all weighted objects on Z (probability measures on Z). The *horizontal lines* represent the process of relaxing/softening the notion of correspondence, whereas *vertical arrows*, carrying the label "gromovization", represent the process by which one gets rid of the ambient space

The ideas put forward in [78, 80, 81] bypassed the computation and comparison of invariants by using the Gromov–Hausdorff distance to directly compare the distance matrices associated to the objects.

In a nutshell, the idea of the Gromov–Hausdorff distance is that in the absence of a common metric space where both A and B are embedded, one first looks for a sufficiently rich, abstract metric space Z that admits isometric copies A' and B' of A and B, respectively. Then, a notion of dissimilarity D between A' and B' is computed and the arbitrariness is eliminated by optimizing over the choice of Z, where one informally calls the process by which this arbitrariness is eliminated *gromovization*. This is the basic idea of the so called Gromov–Hausdorff distances (which arises from the choice $D = d_{\mathcal{H}}^Z(,)$ and $\mathcal{O}(Z) = \mathcal{C}(Z)$) and Gromov–Wasserstein distances (arising from the choice $D = d_{\mathcal{W},p}^Z$ and $\mathcal{O}(Z) = \mathcal{C}_w(Z)$). As discussed in the upcoming sections of the paper, for each $p \in [1, \infty]$ there are two possible definitions of a Gromov–Wasserstein type of distance: $\mathfrak{S}_p$ and $\mathfrak{D}_p$: see Fig. 2.

The formal treatment coming from considering the Gromov–Wasserstein distances on the collection of objects allows for precise statements about the stability and interrelationship between different methods. Those ideas are discussed in the remainder of the paper.

4 The Gromov–Hausdorff Distance

Following [52], the *Gromov–Hausdorff distance* between (compact) metric spaces X and Y is defined as

$$d_{\mathcal{GH}}(X,Y) := \inf_{Z,f,g} d_{\mathcal{H}}^Z\bigl(f(X), g(Y)\bigr) \tag{4.1}$$

where $f : X \to Z$ and $g : Y \to Z$ are *isometric embeddings* (distance preserving transformations) into the metric space Z. This expression seems daunting from the computational point of view since if one chose to compute the Gromov–Hausdorff distance appealing to (4.1), apparently, one would have to optimize over huge spaces defining Z, f and g. There are equivalent tamer expressions which are called below. Nevertheless, as was already pointed out in [81], this expression helps framing the procedure of [37] inside the Gromov–Hausdorff realm.

Let $\mathcal{G}$ denote the collection of all (isometry classes of) compact metric spaces. As seen below in Proposition 4.1, $\mathcal{G}$ can be made into a metric space in its own right by endowing it with the Gromov–Hausdorff metric.

There is an alternative expression for $d_{\mathcal{GH}}$.

Definition 4.1 (Metric coupling) Given $X, Y \in \mathcal{G}$, let $\mathcal{D}(d_X, d_Y)$ denote the set of all possible *metrics* on the disjoint union of X and Y, $X \sqcup Y$ that extend the metrics d_X and d_Y. This means that besides satisfying all triangle inequalities, it also holds that if $d \in \mathcal{D}(d_X, d_Y)$ then $d(x, x') = d_X(x, x')$ and $d(y, y') = d_Y(y, y')$ for all $x, x' \in X$ and $y, y' \in Y$.

Example 4.1 Consider the case when X and Y are finite metric spaces. Let $n_X = |X|$ and $n_Y = |Y|$. In this case, $d \in \mathcal{D}(X, Y)$ can be regarded as a matrix of size $(n_X + n_Y) \times (n_X + n_Y)$ with the following block structure:

$$d = \left(\begin{array}{c|c} d_X & \mathbf{t} \\ \hline \mathbf{t}^T & d_Y \end{array}\right),$$

where $\mathbf{t} \in \mathbb{R}_+^{n_X \times n_Y}$ is s.t. d is a distance matrix. Note that this means that $\mathbf{t}$ is defined by $2(n_Y\binom{n_X}{2} + n_X\binom{n_Y}{2})$ linear constraints arising from all triangle inequalities involving either two points in X and one point in Y, or two points in Y and one in X.

Example 4.2 Let $X = \{x_1, \ldots, x_{n_X}\}$ be a finite metric space and $Y = \{p\}$. In this case, $\mathbf{t} \in \mathbb{R}_+^{n_X}$ must be s.t. $|\mathbf{t}_i - \mathbf{t}_j| \leq d_X(x_i, x_j) \leq \mathbf{t}_i + \mathbf{t}_j$ for $i, j \in \{1, 2, \ldots, n_X\}$.

Example 4.3 Let X, Y be compact metric spaces and $e := \max(\mathrm{diam}(X), \mathrm{diam}(Y))$. For $x \in X$ and $y \in Y$ let $d(x, y) = e/2$, and let d reduce to d_X (respectively, d_Y) on $X \times X$ (respectively, $Y \times Y$). Then, clearly, $d \in \mathcal{D}(d_X, d_Y)$. This means that for all X, Y compact, $\mathcal{D}(d_X, d_Y) \neq \emptyset$.

Remark 4.1 (Alternative form of $d_{\mathcal{GH}}$) One can equivalently (in the sense of equality) define the Gromov–Hausdorff distance between metric spaces (X, d_X) and (Y, d_Y) as (see [22, Remark 7.3.12])

$$d_{\mathcal{GH}}(X, Y) = \inf_{R,d} \sup_{(x,y)\in R} d(x, y) \Big(= \inf_{R,d} \|d\|_{L^\infty(R)} \Big) \tag{4.2}$$

where the infimum is taken over $R \in \mathcal{R}(X, Y)$ and $d \in \mathcal{D}(d_X, d_Y)$. The reader should compare (4.2) with (2.3): informally, the difference is that in the former one optimizes not only on the choice of R, but also on the choice of $d \in \mathcal{D}(d_X, d_Y)$, as well.

The following well-known properties of the Gromov–Hausdorff distance $d_{\mathcal{GH}}$ will be essential for the presentation. From now on, for metric spaces (X, d_X) and (Y, d_Y) let $\Gamma_{X,Y} : X \times Y \times X \times Y \to \mathbb{R}^+$ be given by

$$\Gamma_{X,Y}\big(x, y, x', y'\big) := \big|d_X\big(x, x'\big) - d_Y\big(y, y'\big)\big|. \tag{4.3}$$

Proposition 4.1 (Chap. 7, [22])

(1) *Let* (X, d_X), (Y, d_Y) *and* (Z, d_Z) *be metric spaces then*

$$d_{\mathcal{GH}}(X, Y) \le d_{\mathcal{GH}}(X, Z) + d_{\mathcal{GH}}(Y, Z).$$

(2) *If* $d_{\mathcal{GH}}(X, Y) = 0$ *and* (X, d_X), (Y, d_Y) *are compact metric spaces*, *then* (X, d_X) *and* (Y, d_Y) *are isometric*.

(3) *Let* $\mathbb{X}$ *be a subset of the compact metric space* (X, d_X). *Then*

$$d_{\mathcal{GH}}\big((X, d_X), (\mathbb{X}, d_{X|_{\mathbb{X}\times\mathbb{X}}})\big) \le d_{\mathcal{H}}^X(\mathbb{X}, X).$$

(4) *For compact metric spaces* (X, d_X) *and* (Y, d_Y):

$$\frac{1}{2}\big|\mathrm{diam}(X) - \mathrm{diam}(Y)\big| \le d_{\mathcal{GH}}(X, Y) \le \frac{1}{2}\max\big(\mathrm{diam}(X), \mathrm{diam}(Y)\big) \tag{4.4}$$

where $\mathrm{diam}(X) := \max_{x,x' \in X} d_X(x, x')$ *stands for the* Diameter *of the metric space* (X, d_X).

(5) *For compact metric spaces* (X, d_X) *and* (Y, d_Y),

$$d_{\mathcal{GH}}(X, Y) = \frac{1}{2} \inf_{R \in \mathcal{R}(X,Y)} \sup_{\substack{x_1, x_2 \in X \\ y_1, y_2 \in Y \\ s.t.\, (x_i, y_i) \in R}} \Gamma_{X,Y}(x_1, y_1, x_2, y_2)$$

$$\left(= \frac{1}{2} \inf_R \|\Gamma_{X,Y}\|_{L^\infty(R\times R)}\right). \tag{4.5}$$

Remark 4.2 Note that (4.5) makes direct use of the metrics (distance matrices in the finite setting) of X and Y, which is seemingly more computationally appealing than the standard definition (4.1).

Remark 4.3 (About stability of the Gromov–Hausdorff distance computation) One can obtain the following nice stability condition similar to Remark 2.1. Assume A and B are compact metric spaces, and $\mathbb{A}_n \subset A$ and $\mathbb{B}_m \subset B$ are finite samples of A and B, respectively, then

$$\big|d_{\mathcal{GH}}(A, B) - d_{\mathcal{GH}}(\mathbb{A}_n, \mathbb{B}_m)\big| \le d_{\mathcal{H}}^A(A, \mathbb{A}_n) + d_{\mathcal{H}}^B(B, \mathbb{B}_m). \tag{4.6}$$

The interpretation is very simple. One is interested in computing $d_{\mathcal{GH}}(A, B)$, but one only has access to $\mathbb{A}_n$ and $\mathbb{B}_m$, and therefore one can only attempt to compute $d_{\mathcal{GH}}(\mathbb{A}_n, \mathbb{B}_m)$. Hence, reassuringly, (4.6) above expresses the fact that the error in the answer can be controlled at will by increasing the sampling density.

Remark 4.4 It is possible to use Gromov–Hausdorff ideas to define a certain notion of *partial similarity* between two objects; see [81, Remark 9].

Remark 4.5 (Yet another expression for $d_{\mathcal{GH}}$) It was proved in [61] that Property (5) in Proposition 4.1 above can be recast in a somewhat clearer form: For functions $\phi: X \to Y$ and $\psi: Y \to X$ consider the numbers $A(\phi) := \sup_{x_1,x_2 \in X} |d_X(x_1,x_2) - d_Y(\phi(x_1),\phi(x_2))|$, $B(\psi) := \sup_{y_1,y_2\in Y} |d_X(\psi(y_1),\psi(y_2)) - d_Y(y_1,y_2)|$ and $C(\phi,\psi) := \sup_{x\in X,\, y\in Y} |d_X(x,\psi(y)) - d_Y(\phi(x),y)|$, then

$$d_{\mathcal{GH}}(X,Y) = \inf_{\substack{\phi: X\to Y\\ \psi: Y\to X}} \frac{1}{2}\max\bigl(A(\phi), B(\psi), C(\phi,\psi)\bigr). \tag{4.7}$$

Formula (4.7) is suggestive from the computational point of view and leads to considering certain algorithmic procedures such as those in [17, 80, 81].

Remark 4.6 (Connection to the Quadratic Assignment Problem) As seen next, the expression (4.5) is reminiscent of the QAP (Quadratic Assignment Problem). This will permit inferring something about the inherent complexity of computing the Gromov–Hausdorff distance. Consider finite metric spaces, $\mathbb{X} = \{x_1,\ldots,x_n\}$ and $\mathbb{Y} = \{y_1,\ldots,y_m\}$ with metrics $d_{\mathbb{X}}$ and $d_{\mathbb{Y}}$, respectively. Recall the representation of correspondences between finite sets as $\{0,1\}$-matrices described in Example 2.4. Then one has

$$d_{\mathcal{GH}}(X,Y) = \frac{1}{2}\min_R \max_{i,k,j,l} \Gamma_{ijkl}\, r_{ij} r_{kl},$$

where $\Gamma_{ijkl} := |d_X(x_i,x_k) - d_Y(y_j,y_l)|$. Now, one can obtain a family of related problems by relaxing the max to a sum as follows. Fix $p \ge 1$; then one can also consider the problem

$$(\mathcal{P}_p) \qquad \min_R K(R)$$

where $K(R) := \sum_{ij}\sum_{kl}(\Gamma_{ikjl})^p r_{ij} r_{kl}$.

Problem $(\mathcal{P}_p)$ can be regarded as a generalized version of the QAP. In the standard QAP ([89]) $n = m$ and the inequalities (2.2) defining $\widehat{\mathcal{R}}$ in Example 2.4 are actually equalities, which forces each $R = ((r_{i,j})) \in \widehat{\mathcal{R}}(\mathbb{X},\mathbb{Y})$ to be a permutation matrix.

Actually, as is argued next, when $n = m$, $(\mathcal{P}_p)$ reduces to a QAP, which is known to be an NP-hard problem [89]. Indeed, it is clear that for any $R \in \widehat{\mathcal{R}}(\mathbb{X},\mathbb{Y})$ there exist $\pi \in \Pi_n$ ($n \times n$ permutations matrices) such that $r_{ij} \ge \pi_{ij}$ for all $1 \le i, j \le n$. Then, since $(\Gamma_{ijkl})^p$ is non-negative for all $1 \le i,j,k,l \le n$, it follows that $K_p(R) \ge K_p(\pi)$. Therefore, the minimal value of $K_p(R)$ is attained at some $R \in \Pi_n$.

One also has the following two theorems [22, Chap. 7], cf. Blaschke's and Prokhorov's theorems.

Theorem 4.1 (Gromov's pre-compactness theorem) *Let $\mathcal{F} \subset \mathcal{G}$ be a class of compact metric spaces s.t* (1) $\mathrm{diam}(X) \le D$ *for all* $X \in \mathcal{F}$; *and* (2) *for every* $\varepsilon > 0$ *there exists a natural number N s.t. every $X \in \mathcal{F}$ admits an ε-net with no more than N points. Then, $\mathcal{F}$ is pre-compact in the topology generated by $d_{\mathcal{GH}}$.*

Table 1 Four expressions for the GH distance

(4.1)	$d_{\mathcal{GH}}(X,Y)=\inf_{Z,f,g} d_{\mathcal{H}}^Z(f(X),g(Y))$
(4.2)	$d_{\mathcal{GH}}(X,Y)=\inf_{R,d}\|d\|_{L^\infty(R)}$
(4.5)	$d_{\mathcal{GH}}(X,Y)=\frac{1}{2}\inf_R\|\varGamma_{X,Y}\|_{L^\infty(R\times R)}$
(4.7)	$d_{\mathcal{GH}}(X,Y)=\frac{1}{2}\inf_{\phi,\psi}\max(A(\phi),B(\psi),C(\phi,\psi))$

Theorem 4.2 (Completeness) *The space* $(\mathcal{G},d_{\mathcal{GH}})$ *is complete.*

Remark 4.7 (Summing up: different expressions for the Gromov–Hausdorff distance) Up to this point in the presentation, four expressions for the Gromov–Hausdorff distance have been discussed, and these are put together in Table 1.

These expressions are all related by equalities (for compact metric spaces):

$$(4.1)=(4.2)=(4.5)=(4.7).$$

4.1 The Plan: From Gromov–Hausdorff to Gromov–Wasserstein

In this paper, a modification of the original formulation of [78, 80, 81] is carried out, namely, the proposal is to substitute the underlying Hausdorff component in the definition of the Gromov–Hausdorff distance by a *relaxed* notion of proximity between objects (more precisely by the Wasserstein–Kantorovich–Rubinstein distance) and then find what the equivalent version of Property 5 in Proposition 4.1 would be.

If one refers to the process of substitution of correspondences and max operations by coupling measures and L^p norms, respectively, as *relaxation*, then the goal is to figure out how to complete the diagram below:

$$\begin{tikzcd} d_{\mathcal{H}} \arrow[r, "\text{relax}"] \arrow[d] & d_{\mathcal{W},p} \arrow[d] \\ d_{\mathcal{GH}} \arrow[r, "\text{relax}"] & \boxed{?} \end{tikzcd} \tag{4.8}$$

In the diagram above, vertical lines refer to the *gromovization* process; cf. p. 435.

The main idea in this paper idea is to consider *three* out of the (four) different expressions that one has available for the Gromov–Hausdorff distance and try to pick the one that will provide the most computationally tractable framework without sacrificing the theoretical underpinnings. These three expressions are (4.5),(4.2) and (4.7).

The path starting at (4.7) has been explored first in [78, 80, 81] and later in [17]. This paper concentrates, therefore, on (4.5) and (4.2). As discussed below, these two options are natural. One of them is singled out, based on computational cost considerations of the associated discrete problem. Interestingly, as will be shown below, the two expressions that one obtains from the relaxation procedure applied to (4.2) and (4.5), call them (4.5)* and (4.2)*, respectively, are not related by an equality, in contrast with the fact that (4.5) = (4.2), see Remark 5.14 below. This is represented diagrammatically in (4.9) below where vertical arrows represent the "relaxation":

$$\begin{tikzcd}
d^{(4.2)}_{\mathcal{GH}} \arrow[rr, equal] \arrow[dd] & & d^{(4.5)}_{\mathcal{GH}} \arrow[dd] \\
& d_{\mathcal{H}} \arrow[ul] \arrow[ur] \arrow[dd, Rightarrow] & \\
(4.2)^* \arrow[rr, equal, "?"] & & (4.5)^* \\
& d_{\mathcal{W},p} \arrow[ul] \arrow[ur] &
\end{tikzcd} \tag{4.9}$$

The final version of this diagram is given in (5.12).

The line followed in [17, 80, 81] for developing algorithmic procedures that compute the Gromov–Hausdorff distance was justified only for points sampled from smooth surfaces. In contrast, the formalism of metric measure spaces used in this paper allows the approach in this paper to work in more (theoretical and practical) generality.

One important observation is that both expressions (4.5) and (4.2) make use of the notion of *correspondence*. As discussed in previous sections, in fact, at the level of Hausdorff distances, the formal substitution of correspondences for measure couplings, and of max for L^p norms ($p \geq 1$) leads to Wasserstein–Kantorovich–Rubinstein distances. Now, carrying out the same program on the Gromov–Hausdorff distance will lead to the so called *Gromov–Wasserstein* distances.

The foregoing correspondence in favor of measure couplings is computationally advantageous. This is so because whereas the nature of correspondences is essentially combinatorial, measure couplings take continuous values, even in the case of discrete spaces. This fact simplifies the optimization problems that one must solve in real applications. In addition, no modification of the framework will be needed when solving this practical optimization tasks, in the sense that the implementation is straightforward, as opposed to [17, 81].

These points are further discussed in Sect. 7.

5 Gromov–Wasserstein Distances

In order to carry out the goal of obtaining a more computationally tractable alternative to the Gromov–Hausdorff distance, it is necessary to require more structure than just a set of points with a metric on them: assume that a probability measure is given on the (sets of) points, as was the case in Sect. 2.1. Again, this probability measure can be thought of as signaling the importance of the different points in the dataset, check the definition in Sect. 2.1.

5.1 Metric Measure Spaces

Definition 5.1 [52] A *metric measure space* (mm-space for short) will always be a triple (X, d_X, μ_X) where

- (X, d_X) is a compact metric space.
- μ_X is a Borel probability measure on X i.e., $\mu_X(X) = 1$, and μ_X has full support: $\mathrm{supp}[\mu_X] = X$.

When it is clear from the context, the triple (X, d_X, μ_X) will be denoted by only X. The reason for imposing $\mu_X(X) = 1$ is that one thinks of μ_X as a modelization of the acquisition process or sampling procedure of an object.[5]

Two mm-spaces (X, d_X, μ_X) and (Y, d_Y, μ_Y) are called *isomorphic* iff there exists an isometry $\psi : X \to Y$ such that

$(\psi)_\# \mu_X = \mu_Y$. Furthermore, we will denote by $\mathcal{G}_w$ denote the collection of all mm-spaces.

Example 5.1 Consider the mm-spaces $(\{a, b\}, \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right), \{\frac{1}{2}, \frac{1}{2}\})$ and $(\{a', b'\}, \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right), \{\frac{1}{4}, \frac{3}{4}\})$. These two spaces are isometric but they are not isomorphic; see Fig. 3.

Example 5.2 (The $(n-1)$-simplex as a mm-space) For each $n \in \mathbb{N}$, consider the mm-space (Δ_n, d_n, ν_n) where Δ_n consists of n points $\{1, \dots, n\}$, $d_n(i, j) := 1 - \delta_{i,j}$ and $\nu_n(i) = \frac{1}{n}$. In words, the mm-space Δ_n is the $(n-1)$-simplex where the distance between any two distinct points is 1 and the probability measure is uniform, see Fig. 4.

Example 5.3 (Riemannian manifolds as mm-spaces) Let (M, g) be a compact Riemannian manifold. Consider the metric d_M on M induced by the metric tensor g and the normalized measure μ_M, that is, for all measurable $C \subset M$, $\mu_M(C) = \frac{\mathrm{vol}_M(C)}{\mathrm{Vol}(M)}$. Here, $\mathrm{vol}_M(\cdot)$ is the Riemannian volume measure on M and $\mathrm{Vol}(M) = \mathrm{vol}_M(M)$. Then (M, d_M, μ_M) is a mm-space. Finally, note that since the Riemannian volume measure is entirely determined by the metric, it follows that within **Riem** $\subset \mathcal{G}_w$, *isomorphism reduces to isometry.*

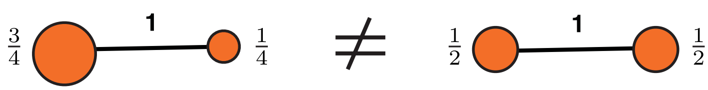

Fig. 3 Two mm-spaces that are isometric but not isomorphic

[5] Think for example of acquisition of a 3D object by a range scanner.

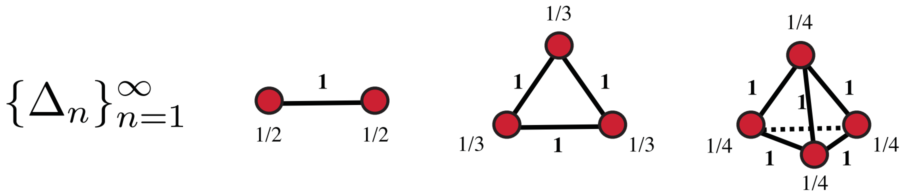

Fig. 4 Family of mm-spaces $\{\Delta_n\}$ where Δ_n is the standard $(n-1)$-simplex with uniform probability measure and metric s.t. the distance between any two distinct points is 1

5.2 Some Invariants of mm-Spaces

Next, a number of simple isomorphism invariants of mm-spaces are defined. Many of these will be used in Sect. 6 to establish lower bounds for the metrics we will impose on $\mathcal{G}_w$.

Definition 5.2 (p-Diameters) Given a mm-space (X, d_X, μ_X) and $p \in [1, \infty]$ we define its *p-diameter* as

$$\operatorname{diam}_p(X) := \left(\int_X \int_X \left(d_X\left(x, x'\right)\right)^p \mu_X(dx)\mu_X\left(dx'\right)\right)^{1/p} \left(= \|d_X\|_{L^p(\mu_X \otimes \mu_X)}\right)$$

for $1 \le p < \infty$, and $\operatorname{diam}_\infty(X) := \operatorname{diam}(\operatorname{supp}[\mu_X])$.

Example 5.4 For each $n \in \mathbb{N}$ consider the $(n-1)$-simplex (Δ_n, d_n, ν_n) of Example 5.2. Then, for $p \ge 1$, $\operatorname{diam}_p(\Delta_n) = (\sum_{i,j=1}^n (1-\delta_{ij})\frac{1}{n^2})^{1/p} = (1-1/n)^{1/p}$. It follows that p-diameters are able to discriminate between simplexes of different dimension.

Definition 5.3 Given $p \in [1, \infty]$ and an mm-space (X, d_X, μ_X) define the *p-eccentricity function* of X by

$$s_{X,p} : X \to \mathbb{R}^+ \quad \text{given by} \quad x \mapsto \left(\int_X d_X\left(x, x'\right)^p \mu\left(dx'\right)\right)^{1/p} \left(= \left\|d_X(x, \cdot)\right\|_{L^p(\mu_X)}\right)$$

for $1 \le p < \infty$, and by

$$s_{X,\infty} : X \to \mathbb{R}^+ \quad \text{given by} \quad x \mapsto \sup_{x' \in \operatorname{supp}[\mu_X]} d_X\left(x, x'\right)$$

for $p = \infty$.

Remark 5.1 Hamza and Krim proposed using eccentricity functions (with $p = 2$) for describing objects in [53].

Figure 5 shows two three-dimensional shapes of animals colored (in the online version) by the value of the eccentricity.

Fig. 5 (Color online) Three-dimensional models of a horse and a camel colored with the values of the eccentricity function $s_{X,1}$ where the metric is an estimate of the geodesic metric and the probability measure was chosen to be uniform. The color code is that *blue* represents low values, while *red* represents high values

Remark 5.2 (From eccentricities to diameters) Note that p-diameters can be computed from eccentricities: indeed, it is easy to check that for any mm-space (X, d_X, μ_X), $\mathrm{diam}_p(X) = \|s_{X,p}\|_{L^p(\mu_X)}$.

Definition 5.4 (Distribution of distances) To an mm-space (X, d_X, μ_X) associate its *distribution of distances*:

$$f_X : \big[0, \mathrm{diam}(X)\big] \to [0,1] \quad \text{given by} \quad t \mapsto \mu_X \otimes \mu_X\big(\big\{\big(x,x'\big)|d_X\big(x,x'\big) \le t\big\}\big).$$

Remark 5.3 (Probabilistic interpretation of the distribution of distances) Note that $f_X(t)$ can be interpreted as follows. Assume that one randomly samples two points $\mathbf{x}$ and $\mathbf{x}'$ from X independently, and each distributed according to the law μ_X; then, $f_X(t)$ equals the probability[6] that the distance between these two random samples is not greater than t, that is, $f_X(t) = \mathbb{P}_{\mu_X \otimes \mu_X}(\{(\mathbf{x}, \mathbf{x}')|\, d_X(\mathbf{x}, \mathbf{x}') \le t\})$. Distributions of distances have been proposed and successfully used in the applied literature [86] under the name of *shape distributions*.

Example 5.5 Consider the mm-spaces $\Delta_2 = (\{p,q\}, \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right), \{\frac{1}{2}, \frac{1}{2}\})$ and $\Delta_3 = (\{p_1, p_2, p_3\}, \left(\begin{smallmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{smallmatrix}\right), \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\})$ from Example 5.2. One easily finds their distributions of distances to be the functions in Fig. 6. In more generality, for $n \in \mathbb{N}$,

$$f_{\Delta_n}(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{n} & \text{for } t \in [0,1) \\ 1 & \text{for } t \ge 1. \end{cases}$$

[6] For some probability space.

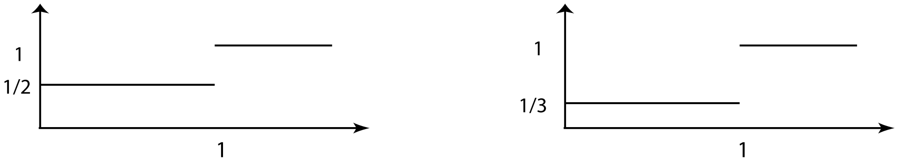

Fig. 6 Distributions of distances of the mm-spaces $\varDelta_2$ and $\varDelta_3$

Remark 5.4 (Distributions of distances and p-diameters) For a given $X \in \mathcal{G}_w$, all p-diameters can be recovered from f_X, since by Cavalieri's principle (cf. [2, Lemma 5.2.7]), for all $p \geq 1$

$$\mathrm{diam}_p(X) = \left(\int_0^\infty t^p \, df_X(dt) \right)^{1/p}.$$

Conversely, one has

Proposition 5.1 *For any $X \in \mathcal{G}_w$, the collection $\{\mathrm{diam}_p(X),\ p \in \mathbb{N}\}$ of all p-diameters (for natural p) determines f_X completely.*

Proof Let $M = \mathrm{diam}(X) < \infty$. Then, $(\mathrm{diam}_p(X))^p \leq M^p$ for $p \geq 1$. Hence all moments $m_k := \int_0^\infty t^k \, df_X(dt)$ of df_X exist and are bounded by M^k, $k \in \mathbb{N}$. In particular, the series $\sum_k m_k \frac{r^k}{k!}$ has radius of convergence equal to $+\infty$, and by [6, Theorem 30.1] df_X is determined by $\{m_k,\ k \in \mathbb{N}\}$. □

This result justifies the following simple procedure for obtaining a signature of and comparing objects: let $S(X) := (s_1, s_2, \ldots, s_p, \ldots)$ be the element of $\mathbb{R}^\infty$ where $s_p = \mathrm{diam}_p(X)$. Then, in order to "compare" two objects X and Y one could merely compute some distance between $S(X)$ and $S(Y)$. The proposition above tells us that $S(X)$ contains the same information as f_X. In this sense, the choice of a metric on $\mathbb{R}^\infty$ seems crucial and at best moot. For this reason it appears to be more sound to use the *SLB* function (and related lower bounds) described in Sect. 6.

Remark 5.5 (Distributions of distances do not discriminate all mm-spaces) Consider three different examples:

- Consider the metric spaces from Fig. 7. Both have the same distribution of distances with associated measure $\frac{1}{4}\delta_0^{\mathbb{R}} + \frac{1}{4}\delta_{\sqrt{2}}^{\mathbb{R}} + \frac{1}{8}\delta_2^{\mathbb{R}} + \frac{1}{4}\delta_{\sqrt{10}}^{\mathbb{R}} + \frac{1}{8}\delta_4^{\mathbb{R}}$. Yet, these two mm-spaces are clearly not isomorphic, since they are not isometric. Boutin and Kemper [11] have analyzed the discrimination of finite sets in Euclidean space using distributions of distances.[7]
- Another example can be constructed easily. For $\alpha, \beta, \gamma \geq 0$ with sum equal to 1 consider the space with three points $X = \left(\{a, b, c\}, \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \{\alpha, \beta, \gamma\}\right)$ and let

[7]They express their results in the language of multisets. Therefore, the implicit assumption is that the sets are endowed with uniform probability measures.

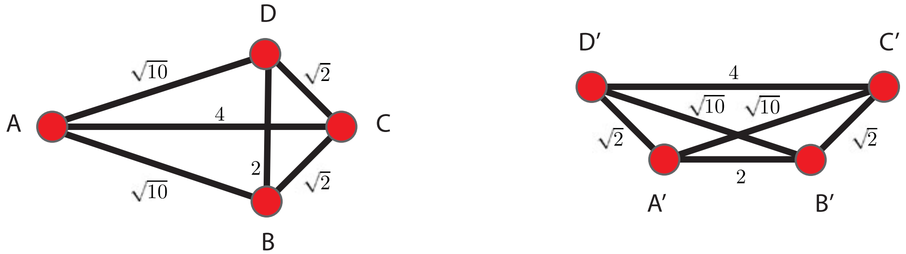

Fig. 7 Two non-isomorphic mm-spaces (both have uniform probability measures attached to them) with the same distributions of distances. Note that these spaces can be realized in $\mathbb{R}^2$ (example taken from [11])

$Y = \Delta_2$. Then,

$$df_X = \delta_0^{\mathbb{R}}\big(\alpha^2 + \beta^2 + \gamma^2\big) + \delta_1^{\mathbb{R}}(2\alpha\beta + 2\beta\gamma + 2\gamma\alpha).$$

and $df_Y = \frac{1}{2}\delta_0^{\mathbb{R}} + \frac{1}{2}\delta_1^{\mathbb{R}}$. Imposing $\alpha^2 + \beta^2 + \gamma^2 = 2\alpha\beta + 2\beta\gamma + 2\gamma\alpha = \frac{1}{2}$ one finds that, modulo a permutation, all solutions are of the form

$$\left(\alpha, \frac{(1-\alpha) \pm \sqrt{\alpha(2-3\alpha)}}{2}, \frac{(1-\alpha) \mp \sqrt{\alpha(2-3\alpha)}}{2}\right)$$

for $\alpha \in [0, 2/3]$. For example, for $\alpha = \frac{1}{4}$, one finds $(\beta, \gamma) = (\frac{3\pm\sqrt{5}}{8}, \frac{3\mp\sqrt{5}}{8})$.

- Another, strikingly simple, example is the following one adapted from [7], which constructs two non-isomorphic finite sets of points on the real line which have the same distribution of distance: Let $X = \{0, 1, 4, 10, 12, 17\} \subset \mathbb{R}$ and $Y = \{0, 1, 8, 11, 13, 17\} \subset \mathbb{R}$ both with uniform probability measures. Let

$$A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 16, 17\},$$

then

$$df_X = df_Y = \delta_0^{\mathbb{R}} \frac{3}{18} + \sum_{a \in A} \delta_a^{\mathbb{R}} \frac{1}{18}.$$

The next definition formalizes the idea of considering a point-wise shape descriptor that assigns to each point the (cumulative) "histogram" of distances from all points to the point of interest.

Definition 5.5 (Local distribution of distances) To a mm-space (X, d_X, μ_X) associate its *local distribution of distances* defined by

$$h_X : X \times \big[0, \operatorname{diam}(X)\big] \to [0, 1] \quad \text{given by}$$
$$(x, t) \mapsto \mu_X\big(\big\{x' | d_X\big(x, x'\big) \le t\big\}\big) = \mu_X\big(\overline{B_X(x, t)}\big).$$

For each $x \in X$, denote by $dh_X(x, \cdot)$ the unique probability measure on $\mathbb{R}^+$ defined by $dh_X(x, [a, b]) = h_X(x, b) - h_X(x, a)$, for all $a, b \ge 0$ with $a \le b$.

Remark 5.6 (Probabilistic interpretation of the local distribution of distances) Note that $h_X(x,t)$ can be interpreted as follows. Assume that one randomly samples a point $\mathbf{x}'$ from X distributed according to the law μ_X, then $h_X(x,t)$ equals the probability that the distance between x and this random sample does not exceed t: $h_X(x,t) = \mathbb{P}_{\mu_X}(d_X(x,\mathbf{x}') \leq t)$.

This type of invariants have been considered in the applied literature before. The earliest mention of this known to the author is in the work of German researchers [3, 5, 62].

The so called *shape context* [4, 23, 97, 102] invariant is closely related to h_X.[8] More similar to h_X is the invariant proposed by Manay et al. in [74] in the context of planar objects. This type of invariant has also been used for three-dimensional objects [30, 42]. More recently, in the context of planar curves, similar constructions have been analyzed in [12].

Remark 5.7 (Local distribution of distances as a proxy for scalar curvature) There is an interesting observation that in the class **Riem** $\subset \mathcal{G}_w$ of closed Riemannian manifolds local distributions of distance are intimately related to curvatures. Let M be an n-dimensional closed Riemannian manifold which by the construction of Example 5.3 one can regard as an mm-space by endowing it with the geodesic metric and with probability measure given by the normalized volume measure. The Riemannian volume of a ball of radius t centered at $x \in M$ has the following expansion [101]:

$$\mathrm{vol}_M\big(B_M(x,t)\big) = \omega_n(t) \cdot \left(1 - \frac{S_M(x)}{6(n+1)}t^2 + O\big(t^4\big)\right),$$

where $S_M(x)$ is the *scalar curvature* of M at x, $\omega_n(t)$ is the volume of a ball of radius t in $\mathbb{R}^n$ and $O(t^4)$ is a term whose decay to 0 as $t \downarrow 0$ is faster than t^4. It then follows that

$$h_M(x,t) = \frac{\omega_n(t)}{\mathrm{Vol}(M)}\left(1 - \frac{S_M(x)}{6(n+2)}t^2 + O\big(t^4\big)\right).$$

One may then argue that local shape distributions play a role of generalized notions of curvature. In Sect. 6, lower bounds are established for the different metrics on $\mathcal{G}_w$ that make explicit use of this generalized notion of curvatures.

Remark 5.8 (From local distributions of distance to distributions of distances) Notice that one can express the distribution of distances f_X of a mm-space (X, d_X, μ_X) in terms of its local distribution of distances. Indeed, note that for $t \geq 0$,

$$Q_X(t) := \big\{\big(x,x'\big) \text{ s.t. } d_X\big(x,x'\big) \leq t\big\} = \big\{\big(x,x'\big) \text{ s.t. } x' \in \overline{B_X(x,t)},\ x \in X\big\},$$

and then

$$f_X(t) = \mu_X \otimes \mu_X\big(Q_X(t)\big) = \int_X \left(\int_{\overline{B_X(x,t)}} \mu_X\big(dx'\big)\right)\mu_X(dx) = \int_X h_X(x,t)\,\mu_X(dx). \tag{5.1}$$

[8] More precisely, h_X corresponds to the cumulative version of the intrinsic shape context.

More suggestive from the probabilistic point of view is the notation $f_X(t) = \mathbb{E}_{\mu_X}(h_X(\mathbf{x}, t))$.

Remark 5.9 (From local distributions of distance to eccentricities) Notice that for $p \geq 1$ one can also express the p-eccentricity $s_{X,p}$ of an mm-space (X, d_X, μ_X) in terms of its local distribution of distances. Clearly, for $x \in X$,

$$s_{X,p}(x) = \left(\int_0^\infty t^p \, dh_X(x, dt)\right)^{1/p}. \tag{5.2}$$

In words, the p-eccentricity $s_{X,p}(x)$ at x is just the pth moment of the measure $dh_X(x)$.

The following counterexample proves that, in general, local shape distributions can confound two non-isomorphic mm-spaces.

Example 5.6 (Local distributions of distances do not discriminate mm-spaces) There are large families of mm-spaces that are confounded by local distributions of distances. Below, two non-isomorphic finite mm-spaces X and Y, with the same cardinality, are constructed s.t. there exist a permutation P with $dh_Y(y) = dh_X(P(y))$ for all $y \in Y$. Consider the mm-spaces $X = (\Delta_3 \sqcup \Delta_3 \sqcup \Delta_3, d, w_X)$ and $Y = (\Delta_3 \sqcup \Delta_3 \sqcup \Delta_3, d, w_Y)$ given in Fig. 8 where

$$d = \left(\begin{array}{ccc|ccc|ccc} 0 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 0 & 1 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 1 & 0 & 2 & 2 & 2 & 2 & 2 & 2 \\ \hline 2 & 2 & 2 & 0 & 1 & 1 & 2 & 2 & 2 \\ 2 & 2 & 2 & 1 & 0 & 1 & 2 & 2 & 2 \\ 2 & 2 & 2 & 1 & 1 & 0 & 2 & 2 & 2 \\ \hline 2 & 2 & 2 & 2 & 2 & 2 & 0 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 & 1 & 0 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 0 \end{array}\right) \tag{5.3}$$

and $w_X, w_Y \in \mathcal{M}_1^+(9)$ are given by

$$w_X = \left[\frac{23}{140}, \frac{1}{105}, \frac{67}{420}, \frac{2}{15}, \frac{1}{15}, \frac{2}{15}, \frac{4}{21}, \frac{1}{28}, \frac{3}{28}\right]^T \tag{5.4}$$

and

$$w_Y = \left[\frac{3}{28}, \frac{1}{15}, \frac{67}{420}, \frac{23}{140}, \frac{2}{15}, \frac{1}{28}, \frac{1}{105}, \frac{4}{21}, \frac{2}{15}\right]^T. \tag{5.5}$$

That X and Y cannot be isomorphic can be seen from Fig. 8, but an explicit argument follows. If they were isomorphic, then there would exist an isometry $\Psi : Y \to X$ that would also respect the weights: $w_Y \circ \Psi = w_X$. Isometries are necessarily bijective, i.e., they are permutations. Notice that there exist only two permutations Π_1 and

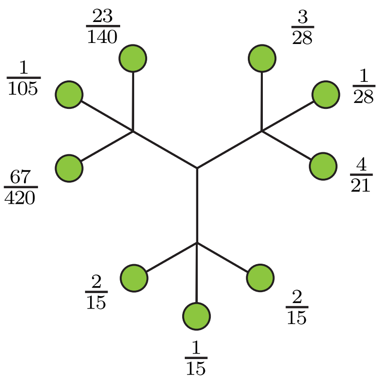

Fig. 8 (Color online) Two non-isomorphic mm-spaces with the same local shape distributions (up to a permutation), see text. In *green* is X and Y is in *orange*. The numbers represent the mass assigned to each node. The length of all edges is $\frac{1}{2}$. Note that by construction, the sum of masses of any three nodes in a branch is $\frac{1}{3}$

$\varPi_2$ s.t. $w_X = w_Y \circ \varPi_i$ for $i \in \{1, 2\}$. These permutations are given by

$$(1, 2, 3, 4, 5, 6, 7, 8, 9) \stackrel{\varPi_1}{\longmapsto} \big((4, 7, 3, \mathbf{5}, 2, \mathbf{9}, 8, 6, 1)\big)$$

and

$$(1, 2, 3, 4, 5, 6, 7, 8, 9) \stackrel{\varPi_2}{\longmapsto} \big((4, 7, 3, \mathbf{9}, 2, \mathbf{5}, 8, 6, 1)\big).$$

Neither of these permutations gives rise to an isometry. Indeed, direct computation shows that

$$\max_{i,j} \big|d(i, j) - d\big(\varPi(i), \varPi(j)\big)\big| = 1 \quad \text{for } \varPi \in \{\varPi_1, \varPi_2\}.$$

Nonetheless, the local shape distributions of X and Y agree up to a permutation. Indeed, notice that there are only three distance values that are possible in the metric d given by (5.3): 0, 1 and 2. Take for example x_1 the first point in X, which as listed in (5.4) has weight $\frac{23}{140}$. Then, for $\rho \in [0, 1)$, $h_X(x, \rho) = \frac{23}{140}$. For $\rho \in [1, 2)$, there are two points besides x_1 s.t. $d_X(x_1, \cdot) \le \rho$ and these have weights $\frac{1}{105}$ and $\frac{67}{420}$ (see Fig. 8) and thus $h_X(x_1, \rho) = \frac{1}{3}$. Finally, for $\rho \ge 2$, $h_X(x_1, \rho) = 1$. Hence one can write that

$$\begin{aligned} dh_X(x_1) &= \omega_X(x_1)\delta_0^{\mathbb{R}} + \left(\frac{1}{105} + \frac{67}{420}\right)\delta_1^{\mathbb{R}} + \left(1 - \frac{1}{3}\right)\delta_2^{\mathbb{R}} \\ &= \omega_X(x_1)\delta_0^{\mathbb{R}} + \frac{71}{420}\delta_1^{\mathbb{R}} + \frac{2}{3}\delta_2^{\mathbb{R}}. \end{aligned}$$

The same procedure can be applied to computing the values of $dh_X(x)$ for all $x \in X$, and similarly for all $dh_Y(y)$, $y \in Y$. The results of carrying out these computations for all points in X and for all points in Y are

$$dh_X = \delta_0^{\mathbb{R}} w_X + \delta_1^{\mathbb{R}} V_X + \delta_2^{\mathbb{R}} \frac{2}{3} U \quad \text{and} \quad dh_Y = \delta_0^{\mathbb{R}} w_Y + \delta_1^{\mathbb{R}} V_Y + \delta_2^{\mathbb{R}} \frac{2}{3} U,$$

where

$$V_X = \left[\frac{71}{420}, \frac{34}{105}, \frac{73}{420}, \frac{1}{5}, \frac{4}{15}, \frac{1}{5}, \frac{1}{7}, \frac{25}{84}, \frac{19}{84}\right]^T,$$

$$V_Y = \left[\frac{19}{84}, \frac{4}{15}, \frac{73}{420}, \frac{71}{420}, \frac{1}{5}, \frac{25}{84}, \frac{34}{105}, \frac{1}{7}, \frac{1}{5}\right]^T,$$

and

$$U = [1, 1, 1, 1, 1, 1, 1, 1, 1]^T.$$

Recall that by construction $w_X = w_Y \circ \Pi$ for $\Pi \in \{\Pi_1, \Pi_2\}$. By direct computation one sees that $V_X = V_Y \circ \Pi$ as well and hence that $dh_X = \delta_0^{\mathbb{R}} w_Y \circ \Pi + \delta_1^{\mathbb{R}} V_Y \circ \Pi + \delta_2^{\mathbb{R}} \frac{2}{3} U \circ \Pi = dh_Y \circ \Pi$ for $\Pi \in \{\Pi_1, \Pi_2\}$. This proves that X and Y have the same local shape distributions up to a permutation.

It is possible to construct a four-parameter family of counterexamples with a similar structure. This counterexample was found using intensive symbolic computation in Matlab.

Example 5.7 (Spheres) For $n \in \mathbb{N}$ consider spheres endowed with geodesic metric and normalized area measure $(\mathbb{S}^n, d, \nu)$. In this case, the following claims are true:

(1) $s_{\mathbb{S}^n,1}(x) = \pi/2$ for all $x \in \mathbb{S}^n$ and for all $n \in \mathbb{N}$.
(2) $\operatorname{diam}_1(\mathbb{S}^n) = \pi/2$ for all $n \in \mathbb{N}$.
(3) $\operatorname{diam}_\infty(\mathbb{S}^n) = \pi$ for all $n \in \mathbb{N}$.
(4) $\operatorname{diam}_2(\mathbb{S}^2) = \sqrt{\pi^2/2 - 1}$.
(5) $\operatorname{diam}_2(\mathbb{S}^1) = \pi/\sqrt{3}$.
(6) $h_{\mathbb{S}^n}(x, t) = f_{\mathbb{S}^n}(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi}\,\Gamma(\frac{n}{2})} \int_0^t (\sin r)^{n-1}\, dr$ for all $x \in \mathbb{S}^n$ and $t \in [0, \pi]$.

One can easily prove (1), as follows. Let $a : \mathbb{S}^n \to \mathbb{S}^n$ denote the *antipodal map*. Fix $x \in \mathbb{S}^n$. Note that for all $x' \in \mathbb{S}^n$,

$$d(x, x') + d(x', a(x)) = \pi$$

and in particular, integrating out x' with respect to ν, one obtain $s_{\mathbb{S}^n,1}(x) + s_{\mathbb{S}^n,1}(a(x)) = \pi$. But by symmetry of $\mathbb{S}^n$, $s_{\mathbb{S}^n,1}$ must be constant, and hence the claim. Claim (2) follows from (1). Claim (3) is obvious, and (4) and (5) follow from simple computations. Finally, for (6) first note that by symmetry $h_{\mathbb{S}^n}(x, t)$ does not depend on x. That $h_{\mathbb{S}^n}(\cdot, t) = f_{\mathbb{S}^n}(t)$ follows from (5.1) and the fact that $h_{\mathbb{S}^n}(x, t)$ is independent of x. Finally, the explicit expression for $\nu(\overline{B_{\mathbb{S}^n}(\cdot, t)})$ follows from standard formulas for the area of spheres [46].

Remark 5.10 The preceding remarks (Remarks 5.8, 5.9 and 5.2) prove that local distributions of distances are the strongest amongst the invariants that have been considered so far. See also Sect. 6.

5.3 Two Distances on $\mathcal{G}_w$

This section introduces two notions of distance on mm-spaces that are central to all subsequent considerations. As was done for metric spaces in Sect. 2, one needs to introduce a notion of correspondence/coupling between the mm-spaces involved in the comparison. The definition below is essentially the same as Definition 2.3.

Definition 5.6 (Measure coupling) Given two metric measure spaces (X, d_X, μ_X) and (Y, d_Y, μ_Y) one says that a measure μ on the product space $X \times Y$ is a *coupling* of μ_X and μ_Y iff

$$\mu(A \times Y) = \mu_X(A), \quad \text{and} \quad \mu\big(X \times A'\big) = \mu_Y\big(A'\big) \tag{5.6}$$

for all measurable sets $A \subset X$, $A' \subset Y$. Denote by $\mathcal{M}(\mu_X, \mu_Y)$ the set of all couplings of μ_X and μ_Y.

Starting from (4.5), we now construct a new, tentative notion of distance between metric spaces. The idea is to use (4.5) as the starting point because the goal is for the new distance to directly compare the metrics of X and Y (in a meaningful way). Roughly speaking, the idea is to substitute the L^∞ norm in (4.5) by L^p norms, and correspondences by coupling measures. For $p \in [1, \infty)$ and $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ let

$$\begin{aligned}\mathbf{J}_p(\mu) &:= \frac{1}{2}\left(\int_{X\times Y}\int_{X\times Y}\big(\Gamma_{X,Y}\big(x, y, x', y'\big)\big)^p \,\mu(dx \times dy)\mu\big(dx' \times dy'\big)\right)^{1/p} \\ &\left(= \frac{1}{2}\|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)}\right)\end{aligned} \tag{5.7}$$

and also let

$$\mathbf{J}_\infty(\mu) := \frac{1}{2} \sup_{\substack{x,x'\in X\\ y,y'\in Y\\ \text{s.t. } (x,y),(x',y')\in R(\mu)}} \Gamma_{X,Y}\big(x, y, x', y'\big)\left(= \frac{1}{2}\|\Gamma_{X,Y}\|_{L^\infty(R(\mu)\times R(\mu))}\right) \tag{5.8}$$

Definition 5.7 For $\infty \geq p \geq 1$ one defines the *distance* $\mathfrak{D}_p$ between two mm-spaces X and Y by

$$\mathfrak{D}_p(X, Y) := \inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)} \mathbf{J}_p(\mu). \tag{5.9}$$

One needs to prove that expression (5.9) in fact defines a metric on the set of all isomorphism classes of mm-spaces, which constitutes an interesting technical step in itself. These and other properties of $\mathfrak{D}_p$, of similar spirit to those reported for $d_{\mathcal{GH}}(,)$ in Proposition 4.1, are treated in Theorem 5.1 below.

Remark 5.11 ($\mathfrak{D}_p$ as a "relaxation" of (4.5)) Recall that $d_{\mathcal{GH}}^{(4.5)}(X,Y) = \frac{1}{2}\inf_R \|\varGamma_{X,Y}\|_{L^\infty(R\times R)}$. Since by definition $\mathfrak{D}_p(X,Y) = \frac{1}{2}\inf_\mu \|\varGamma_{X,Y}\|_{L^p(\mu\otimes\mu)}$, in view of the informal discussion summarized in (4.9), one sees that (4.5)* = (5.9). Indeed, notice that, formally, the structure of $\mathfrak{D}_p$ can be obtained from that of (4.5) by substitution of correspondences by measure couplings and L^∞ norms by L^p norms.

Definition 5.8 (Another distance on $\mathcal{G}_w$: Sturm's construction) In [105] K.T. Sturm introduced and studied the following distance for mm-spaces (for each $p \geq 1$):[9]

$$\mathfrak{S}_p(X,Y) := \inf_{\mu,d}\left(\int_{X\times Y} d(x,y)^p\,\mu(dx\times dy)\right)^{1/p}\left(= \inf_{d,\mu}\|d\|_{L^p(\mu)}\right) \tag{5.10}$$

where the infimum is taken over all $d \in \mathcal{D}(d_X, d_Y)$ (recall Definition 4.1) and $\mu \in \mathcal{M}(\mu_X, \mu_Y)$.

The corresponding definition for $p = \infty$ is

$$\mathfrak{S}_\infty(X,Y) = \inf_{\mu,d}\sup_{(x,y)\in R(\mu)} d(x,y)\left(= \inf_{d,\mu}\|d\|_{L^\infty(R(\mu))}\right). \tag{5.11}$$

Remark 5.12 Recall that $d_{\mathcal{GH}}^{(4.2)}(X,Y) = \inf_{d,R}\|d\|_{L^\infty(R)}$, hence one sees that Sturm's proposal (5.10) corresponds to what was called (4.2)* in Sect. 4.1, i.e., (4.2)* = (5.10). Since $\{R(\mu)|\,\mu \in \mathcal{M}(\mu_X,\mu_Y)\} \subset \mathcal{R}(X,Y)$ it is clear that $\mathfrak{S}_\infty(X,Y) \geq d_{\mathcal{GH}}(X,Y)$. Notice that is bound is dual to the inequality $d_{\mathcal{W},\infty} \geq d_{\mathcal{H}}$ contained in Proposition 2.3.

Remark 5.13 (From measure couplings to correspondences) In practical applications one seeks not only a measure of dissimilarity between objects, but also the knowledge of the precise *matching* between them may be important. In the context of Gromov–Wasserstein distances, the (optimal) matching between objects X and Y is encoded by a measure coupling μ which seems to provide nothing but a *fuzzy* type of matching. However, by Lemma 2.2, given $R(\mu) = \mathrm{supp}[\mu]$ is in fact a correspondence between X and Y. This correspondence obviously induces two maps $\phi : X \to Y$ and $\psi : Y \to X$ that can be used to map back an forth between the two objects. There is a sense in which this is a reasonable setting. As is proved in Theorem 5.1 below, $\mathbf{J}_p(\mu) = 0$ implies that $R(\mu)$ describes an isometry between X and Y and in this case ϕ and ψ are forced to be inverses of each-other.

Remark 5.14 ($\mathfrak{S}_p$ and $\mathfrak{D}_p$ are not equal in general) At this point it becomes clear that in the construction outlined in Sect. 4.1, (4.5)* = (5.9) and (4.2)* = (5.10), see (4.9). Note that since (4.5) and (4.2) are equal, one could conjecture (5.9) and (5.10) to be equal as well. In this respect, Theorem 5.1 below proves that $\mathfrak{S}_p \geq \mathfrak{D}_p$ for $1 \leq p \leq \infty$ and that $\mathfrak{S}_\infty = \mathfrak{D}_\infty$. However, for $p < \infty$ the equality does not hold in general.

[9] He presented the case that $p = 2$.

We now exhibit $X, Y \in \mathcal{G}_w$ s.t. $\mathfrak{D}_p(X,Y) < \mathfrak{S}_p(X,Y)$. For each $n \geq 2$ let $X = \Delta_n$ be the $(n-1)$-simplex (see Example 5.2) and let Y be a single point $\{y\}$. Let us consider for simplicity the case $p = 1$. From Example 2.7 it is known that $\mathcal{M}(\nu_n, \{1\}) = \{\nu_n\}$, and from Example 4.1:

$$\mathcal{D}(\Delta_n, Y) = \left\{ \begin{pmatrix} d_X & f \\ f^T & 0 \end{pmatrix}, f \in \mathbb{R}^n_+ \,|\, f_i - f_j \leq 1 \leq f_i + f_j, i \neq j \right\}.$$

Hence, $\mathfrak{S}_1(\Delta_n, Y) = \inf_f \sum_i f_i \nu_n(i) = \inf_f \frac{\sum_i f_i}{n}$ where the infimum is taken among all f satisfying the conditions in the definition of $\mathcal{D}(\Delta_n, Y)$. In the constraints for f there are $\binom{n}{2}$ different inequalities of the form $1 \leq f_i + f_j$ $(i \neq j)$. Adding them and noting that each f_i appears in exactly $n-1$ of those inequalities yields $\binom{n}{2} \leq (n-1) \sum_i f_i$. This implies that for all $n \geq 2$,

$$\mathfrak{S}_1(\Delta_n, Y) \geq \frac{1}{2}.$$

From Theorem 5.1 below it follows that

$$\mathfrak{D}_1(\Delta_n, Y) = \frac{1}{2} \operatorname{diam}_1(\Delta_n) = \frac{n-1}{2n}$$

where the last equality follows from Example 5.4. Hence, one sees that $\mathfrak{S}_1(\Delta_n, Y) > \mathfrak{D}_1(\Delta_n, Y)$ for all $n \geq 2$. The next interesting question is whether $\mathfrak{S}_p$ and $\mathfrak{D}_p$ are "comparable" in the Lipschitz sense. The arguments in Remark 5.17 prove this to be in general false. However, as explained in Sect. 5.5, a weaker comparability statement does hold, namely that they both generate the same topology on $\mathcal{G}_w$. Proposition 5.2 below proves that within **Riem** a bi-Hölder type of comparability does hold.

The diagram below summarizes the proposed construction, cf. (4.9), vertical arrows signify the process of *relaxing* (that is, correspondences are substituted by measure couplings, and L^∞ norms by L^p norms).

$$\begin{tikzcd}
\boxed{d_{\mathcal{GH}}^{(4.2)} = \inf_{d,R} \|d\|_{L^\infty(R)}} \arrow[rr, equal] \arrow[dd] & & \boxed{d_{\mathcal{GH}}^{(4.5)} = \frac{1}{2} \inf_R \|\Gamma_{X,Y}\|_{L^\infty(R \times R)}} \arrow[dd] \\
& d_{\mathcal{H}} \arrow[ul] \arrow[ur] \arrow[dd, Rightarrow] & \\
\boxed{\mathfrak{S}_p = \inf_{d,\mu} \|d\|_{L^p(\mu)}} \arrow[rr, dotted, no head] & & \boxed{\mathfrak{D}_p = \frac{1}{2} \inf_\mu \|\Gamma_{X,Y}\|_{L^p(\mu \otimes \mu)}} \\
& d_{\mathcal{W},p} \arrow[ul] \arrow[ur] &
\end{tikzcd} \tag{5.12}$$

As is argued in Sect. 7, $\mathfrak{D}_p$ is more amenable to numerical computations than $\mathfrak{S}_p$.

Remark 5.15 One may wonder what is the relationship between $d_{\mathcal{GH}}(X,Y)$ and (some of) the $\mathfrak{D}_p(X,Y)$'s. In this respect, Theorem 5.1 (b) below asserts that $d_{\mathcal{GH}}(X,Y) \leq \mathfrak{D}_\infty(X,Y)$.

5.4 Properties of $\mathfrak{D}_p$

The theorem below summarizes the main properties of $\mathfrak{D}_p$; its proof is postponed to Sect. 10.1. See [105] for a treatment of similar properties for $\mathfrak{S}_p$.

In Sect. 6 we present lower bounds for $\mathfrak{D}_p$ based on the invariants discussed in Sect. 5.2.

Theorem 5.1 (Properties of $\mathfrak{D}_p$) *Let $p \in [1,\infty]$, then*

(a) *$\mathfrak{D}_p$ defines a metric on the collection of all isomorphism classes of mm-spaces.*
(b) (*Relationship with $d_{\mathcal{GH}}$*) *Let X and Y be two mm-spaces,*[10] *then one has*

$$d_{\mathcal{GH}}(X,Y) \leq \mathfrak{D}_\infty(X,Y),$$

where it is understood that on the left-hand side X and Y are the canonical projections of $X, Y \in \mathcal{G}_w$ onto $\mathcal{G}$.
(c) (*What happens under two probability measures on the same space* (*keeping the same metric*)*?*) *Let (Z,d) be a compact metric space and α and β two different Borel probability measures on Z. Let $X = (Z,d,\alpha)$ and $Y = (Z,d,\beta)$ then*

$$\mathfrak{D}_p(X,Y) \leq d^Z_{\mathcal{W},p}(\alpha,\beta).$$

(d) (*What happens under two different metrics on the same space* (*keeping the same probability measure*)*?*) *Let (Z,α) be a measure space and $d, d' : X \times X \to \mathbb{R}^+$ be two measurable metrics on Z.*

Let $X = (Z,d,\alpha)$ and $Y = (Z,d',\alpha)$, then

$$\mathfrak{D}_p(X,Y) \leq \frac{1}{2}\|d - d'\|_{L^p(Z\times Z,\alpha\otimes\alpha)}.$$

(e) (*What happens for a random sampling of the metric space?*) *Let $p \in [1,\infty)$ and $\mathbb{X}_m \subset X$ be a set of m random variables $\mathbf{x}_i : \Omega \to X$ defined on some probability space Ω with law μ_X. Let $\mu_m(\omega,\cdot) := \frac{1}{m}\sum_{i=1}^m \delta^X_{\mathbf{x}_i(\omega)}$ denote the empirical measure. For each $\omega \in \Omega$ consider the mm-spaces (X,d_X,μ_X) and $(\mathbb{X}_m,d_X,\mu_m)$, then, for μ_X-almost all $\omega \in \Omega$,*

$$(\mathbb{X}_m,d_X,\mu_m) \xrightarrow{\mathfrak{D}_p} (X,d_X,\ \mu_X) \quad \text{as } m \uparrow \infty.$$

[10] Note that in the definition of mm-spaces, $(X,d_X,\mu_X) \in \mathcal{G}_w$, then $\mathrm{supp}[\mu_X] = X$.

(f) (*Distance to a point*) *Let* $Z = \{z\}$, *then*

$$\mathfrak{D}_p(X, Z) = \frac{\operatorname{diam}_p(X)}{2}.$$

From this and Property (*a*) (*triangle inequality*), *it holds that for all* $X, Y \in \mathcal{G}_w$:

$$\frac{\operatorname{diam}_p(X) + \operatorname{diam}_p(Y)}{2} \geq \mathfrak{D}_p(X, Y) \geq \left| \frac{\operatorname{diam}_p(X) - \operatorname{diam}_p(Y)}{2} \right|. \tag{5.13}$$

(g) (*Relationship to* $\mathfrak{S}_p$) *For any* $X, Y \in \mathcal{G}_w$, *it holds that* $\mathfrak{S}_p(X, Y) \geq \mathfrak{D}_p(X, Y)$ *for* $p \geq 1$, *and* $\mathfrak{S}_\infty(X, Y) = \mathfrak{D}_\infty(X, Y)$.
(h) (*Ordering of the different distances*). $\mathfrak{D}_p \geq \mathfrak{D}_q$ *and* $\mathfrak{S}_p \geq \mathfrak{S}_q$ *whenever* $\infty \geq p \geq q \geq 1$.
(i) (*Equivalence of the different distances*). *If* $p \geq q \geq 1$, *then* $\mathfrak{D}_p(X, Y) \leq M^{1-\frac{q}{p}} \cdot (\mathfrak{D}_q(X, Y))^{\frac{q}{p}}$ *where* $M = \max(\operatorname{diam}(X), \operatorname{diam}(Y))$.

Remark 5.16 (The parameter p) Writing the framework with p as a parameter is not superfluous. In fact even the simple bound (5.13) will be useful for discriminating between certain spaces that the corresponding Gromov–Hausdorff bound (4.4) cannot. For example, consider the case when $X = (\mathbb{S}^1, d_1, \mu_1)$ and $X = (\mathbb{S}^2, d_2, \mu_2)$ where d_1 and d_2 are the usual spherical distance metrics and μ_1 and μ_2 stand for normalized area on $\mathbb{S}^1$ and $\mathbb{S}^2$, respectively. Since $\operatorname{diam}(\mathbb{S}^n) = \pi$ for all $n \in N$, then (4.4) vanishes as $(5.13)_{p=\infty}$ also does. However, since by Example 5.7, $\operatorname{diam}_2(\mathbb{S}^1) = \pi/\sqrt{3}$ and $\operatorname{diam}_2(\mathbb{S}^2) = \sqrt{\frac{\pi^2}{2} - 2}$, it follows that $(5.13)_{p=2}$ does permit telling $\mathbb{S}^1$ and $\mathbb{S}^2$ apart. In fact, invoking (5.13), one sees that $\mathfrak{D}_2(\mathbb{S}^1, \mathbb{S}^2) \geq 0.0503$.

Definition 5.9 Let X be a set and $d, d' : X \times X \to \mathbb{R}^+$ two metrics on X. One says that d and d' are *bi-Hölder equivalent* with exponents $\alpha, \beta > 0$, whenever there exist constants $c_2 \geq c_1 > 0$ s.t.

$$c_1 \cdot \big(d\big(x, x'\big)\big)^\alpha \leq d'\big(x, x'\big) \leq c_2 \cdot \big(d\big(x, x'\big)\big)^\beta \quad \text{for all } x, x' \in X.$$

When the inequality holds for $\alpha = \beta = 1$, one says that d and d' are *bi-Lipschitz equivalent*. Note that if d and d' are bi-Hölder equivalent, then a sequence $\{x_n\}_n \subset X$ is Cauchy w.r.t. d *if and only if* $\{x_n\}_n$ is Cauchy w.r.t. d'.

Remark 5.17 ($\mathfrak{S}_p$ and $\mathfrak{D}_p$ are not bi-Lipschitz equivalent) For all $n \in \mathbb{N}$ and $p \in [1, \infty)$,

$$\mathfrak{S}_p(\Delta_n, \Delta_{2n}) \geq \frac{1}{4} \quad \text{and} \quad \mathfrak{D}_p(\Delta_n, \Delta_{2n}) \leq \frac{1}{2}\left(\frac{3}{2n}\right)^{1/p}.$$

Indeed, the following facts are true for $p \in [1, \infty)$:

Claim 5.1 *For all $m, n \in \mathbb{N}$,*

$$\mathfrak{D}_p(\Delta_n, \Delta_m) = \frac{1}{2}\left(\frac{1}{n} + \frac{1}{m} - 2 \cdot \max_{\mu \in \mathcal{M}(\nu_n, \nu_m)} \sum_{x,y} \mu_{x,y}^2\right)^{1/p} \tag{5.14}$$

and in particular, from (5.14) it follows that

$$\mathfrak{D}_p(\Delta_n, \Delta_m) \leq \frac{1}{2}\left(\frac{1}{n} + \frac{1}{m}\right)^{1/p}. \tag{5.15}$$

Proof Indeed, note that for all $x, x' \in \Delta_n$ and $y, y' \in \Delta_m$ one has that $\Gamma_{\Delta_n,\Delta_m}(x, y, x', y') = |\delta_{x,x'} - \delta_{y,y'}|$. Note that $\Gamma_{\Delta_n,\Delta_m}$ is 0 or 1 and hence $\Gamma_{\Delta_n,\Delta_m}(x, y, x', y') = |\delta_{x,x'} - \delta_{y,y'}|^2 = \delta_{x,x'} + \delta_{y,y'} - 2\delta_{x,x'} \cdot \delta_{y,y'} \in \{0, 1\}$. Clearly, for all $p \in [1, \infty)$, $(\Gamma_{\Delta_n,\Delta_m}(x, y, x', y'))^p = \Gamma_{\Delta_n,\Delta_m}(x, y, x', y')$. Then, for any $\mu \in \mathcal{M}(\nu_n, \nu_m)$,

$$\begin{aligned}
\left(2\mathbf{J}_p(\mu)\right)^p &= \sum_{x,y} \mu_{x,y} \sum_{x',y'} \mu_{x',y'}(\delta_{x,x'} + \delta_{y,y'} - 2\delta_{x,x'}\delta_{y,y'}) \\
&= \sum_{x,y} \mu_{x,y}\left(\nu_n(x) + \nu_m(y) - 2\mu_{x,y}\right) \\
&= \sum_x \nu_n^2(x) + \sum_y \nu_m^2(y) - 2\sum_{x,y} \mu_{x,y}^2 \\
&= \frac{1}{n} + \frac{1}{m} - 2\sum_{x,y} \mu_{x,y}^2
\end{aligned}$$

from which (5.14) follows. □

Claim 5.2 *For all $n, m \in \mathbb{N}$, one has*

$$\max_{\mu \in \mathcal{M}(\nu_n, \nu_m)} \sum_{x,y} \mu_{x,y}^2 \leq \frac{1}{2}\left(1/n + 1/m - |1/n - 1/m|\right) = \frac{1}{\max(n, m)}. \tag{5.16}$$

Proof For all $n \in \mathbb{N}$ and $p \in [1, \infty)$ from Example 5.4 it follows that $\operatorname{diam}_p(\Delta_n) = (1 - 1/n)^{1/p}$. Then, from Theorem 5.1(h) and (f), $\mathfrak{D}_p(\Delta_n, \Delta_m) \geq \mathfrak{D}_1(\Delta_n, \Delta_m) \geq \frac{1}{2}|1/n - 1/m|$. Now, from (5.14) for $p = 1$, one obtains (5.16). □

Claim 5.3 *For all $n, m \in \mathbb{N}$:*

$$\mathfrak{S}_p(\Delta_n, \Delta_m) \geq \frac{1}{2}\left(1 - \frac{\min(n, m)}{\max(n, m)}\right). \tag{5.17}$$

Proof Let $d \in \mathcal{D}(d_n, d_m)$ and $\mu \in \mathcal{M}(\nu_n, \nu_m)$, then, for all $x, x' \in \Delta_n$ and $y \in \Delta_m$, $d(x, y) + d(x', y) \geq d_n(x, x') = 1 - \delta_{x,x'}$. Then, for all $y \in \Delta_m$,

$$\sum_{x,x'} \mu_{x,y}\mu_{x',y}\big(d(x, y) + d\big(x', y\big)\big) \geq \sum_{x,x'} \mu_{x,y}\mu_{x',y}(1 - \delta_{x,x'})$$
$$\Downarrow$$
$$2\sum_{x} \mu_{x,y} d(x, y)\nu_m(y) \geq \big(\nu_m(y)\big)^2 - \sum_{x} \mu_{x,y}^2.$$

Now, sum over all $y \in \Delta_m$, and substitute $\nu_m(y) = 1/m$ to obtain:

$$\sum_{x,y} d(x, y)\mu_{x,y} \geq \frac{1}{2}\bigg(1 - n\sum_{x,y} \mu_{x,y}^2\bigg).$$

By symmetry and the fact that $\mu \in \mathcal{M}(\nu_n, \nu_m)$ was arbitrary, one finds

$$\mathfrak{S}_1(\Delta_n, \Delta_m) \geq \frac{1}{2}\bigg(1 - \min(n, m) \cdot \max_{\mu\in\mathcal{M}(\nu_n,\nu_m)} \sum_{x,y} \mu_{x,y}^2\bigg).$$

Combining this expression with (5.16) and recalling Theorem 5.1(h) one obtains the claim. □

Claim 5.4 *$\mathfrak{S}_p$ and $\mathfrak{D}_p$ are not bi-Lipschitz equivalent.*

Proof Let $m = 2n$, then by (5.15) and (5.17) one sees that $\mathfrak{S}_p(\Delta_n, \Delta_{2n}) \geq 1/4$ and

$$\mathfrak{D}_p(\Delta_n, \Delta_{2n}) \leq \frac{1}{2}\bigg(\frac{3}{2n}\bigg)^{1/p}$$

whose the right-hand side vanishes as $n \uparrow \infty$. Then, the condition of Definition 5.9 cannot be satisfied. □

Remark 5.18 It follows from the previous remark that $(\mathcal{G}_w, \mathfrak{D}_p)$ *is not complete*. Indeed, a potential limit object for the sequence $\{\Delta_n\}_{n\in\mathbb{N}}$ is a space with infinitely many points and with the discrete metric. Such a space would not be compact.

Remark 5.19 (Gromov's box distance) In Chap. $3\frac{1}{2}$ of [52] Gromov proposes at least two more notions of distance on $\mathcal{G}_w$. In particular, for each $\lambda \geq 0$ he defines the "box metric" $\underline{\Box}_\lambda$ which Sturm [105] proved to be bi-Lipschitz equivalent to $\mathfrak{S}_p$ within families of compact mm-spaces with uniform bound on the diameters. Gromov's box metric, nonetheless, appears to lead to hard combinatorial problems; this has been another motivation for the construction of $\mathfrak{D}_p$.

5.5 Equivalence of $\mathfrak{D}_p$ and $\mathfrak{S}_p$

This section establishes the topological equivalence of $\mathfrak{S}_p$ and $\mathfrak{D}_p$ for $p < \infty$. Recall that one already has that $\mathfrak{D}_p \leq \mathfrak{S}_p$ on $\mathcal{G}_w \times \mathcal{G}_w$.

It is useful to introduce another invariant of mm-spaces.

Definition 5.10 (Modulus of mass distribution) For $\delta > 0$, define the *modulus of mass distribution* of $X \in \mathcal{G}_w$ as

$$v_\delta(X) := \inf\big\{\varepsilon > 0 | \mu_X\big(\big\{x | \mu_X\big(B_X(x,\varepsilon)\big) \leq \delta\big\}\big) \leq \varepsilon\big\}. \tag{5.18}$$

Example 5.8 It is easy to see that

- (*Simplices*) $v_\delta(\Delta_n) = 1$ if $\delta \geq n^{-1}$ and $v_\delta(\Delta_n) = 0$ if $\delta < n^{-1}$,
- (*Spheres*) $v_\delta(\mathbb{S}^n) = \min(1, \rho_n^{-1}(\delta))$, where $\rho_n(\varepsilon) = \nu(B_{\mathbb{S}^n}(\cdot, \varepsilon))$ and ν is the normalized area measure on $\mathbb{S}^n$, see Example 5.7.

Proposition 5.2 (Lemma 6.5, [50]) *For any fixed $X \in \mathcal{G}_w$, the map that sends $\delta \geq 0$ to $v_\delta(X)$ is non-decreasing, right-continuous and bounded by* 1. *Moreover*, $\lim_{\delta \downarrow 0} v_\delta(X) = 0$.

The proposition below, together with the facts that always $\mathfrak{S}_p \geq \mathfrak{D}_p$ (Theorem 5.1(g)) and $v_\delta(X) \stackrel{\delta\downarrow 0}{\to} 0$ for all $X \in \mathcal{G}_w$, establishes the topological equivalence between $\mathfrak{S}_p$ and $\mathfrak{D}_p$ on $\mathcal{G}_w$ for $p \geq 1$.

Proposition 5.3 *Let $X, Y \in \mathcal{G}_w$, $p \in [1,\infty)$ and $\delta \in (0, 1/2)$. Then*,

$$\mathfrak{S}_p(X,Y) \leq \big(4 \cdot \min\big(v_\delta(X), v_\delta(Y)\big) + \delta\big)^{1/p} \cdot M$$

whenever $\mathfrak{D}_p(X,Y) < \delta^5$, where $M = 2 \cdot \max(\operatorname{diam}(X), \operatorname{diam}(Y)) + 45$.

Remark 5.20 If $\mathcal{F} \subset \mathcal{G}_w$ is a family for which there exists a surjective function $\rho_{\mathcal{F}} : [0,\infty) \to [0, \delta_{\mathcal{F}}]$ with $\delta_{\mathcal{F}} > 0$, such that

$$\mu_X\big(B_X(x,\varepsilon)\big) \geq \rho_{\mathcal{F}}(\varepsilon) \quad \text{for all } \varepsilon \geq 0, x \in X, \text{all } X \in \mathcal{F},$$

then for all $\delta \in (0, \delta_{\mathcal{F}})$:

$$\sup_{X \in \mathcal{F}} v_\delta(X) \leq \inf\big\{\varepsilon > 0 | \rho_{\mathcal{F}}(\varepsilon) > \delta\big\}.$$

Indeed, assume that $\varepsilon > 0$ is such that $\rho_{\mathcal{F}}(\varepsilon) > \delta$ and fix any $X \in \mathcal{F}$. Then, clearly, $L_\delta := \{x \in X | \mu_X(B_X(x,\varepsilon)) \leq \delta\} = \emptyset$, and hence $\mu_X(L_\delta) = 0 \leq \varepsilon$. Thus $v_\delta(X) \leq \varepsilon$ and the claim follows since ε (s.t. $\delta < \rho_{\mathcal{F}}(\varepsilon)$) and X were arbitrary.

When X and Y are in suitable classes of mm-spaces, one can refine the result in Proposition 5.3. In particular, the following theorem asserts that when X and Y are restricted to certain classes of Riemannian manifolds one essentially obtains bi-Hölder equivalence.

Theorem 5.2 *Let $p \in [1,\infty)$ and $\mathcal{F} = \mathcal{F}(n,K,V,I,D) \subset \mathcal{G}_w$ denote the collection of all n-dimensional Riemannian manifolds endowed with normalized volume measures, with uniform upper bound $K > 0$ on their sectional curvatures, uniform upper bound V on their volumes, uniform lower bound $I > 0$ on their injectivity radii, and uniform upper bound D on their diameters. Then, there exist constants $\delta_{\mathcal{F}} = \delta(n,K,V,I) > 0$ and $C_{\mathcal{F}} = C(n,K,V,I,D) > 0$ with the property that whenever $X, Y \in \mathcal{F}(n,K,V,I,D)$ are such that $\mathfrak{D}_p(X,Y) < \delta_{\mathcal{F}}$, then*

$$\mathfrak{S}_p(X,Y) \le C_{\mathcal{F}} \cdot \big(\mathfrak{D}_p(X,Y)\big)^{\frac{1}{5np}}.$$

5.6 Pre-compact Families in $(\mathcal{G}_w, \mathfrak{D}_p)$

One has the following sufficient conditions for pre-compactness.

Theorem 5.3 *Let $\mathcal{F}$ be a family of mm-spaces such that there exist $D > 0$ and $\rho_{\mathcal{F}} : [0,\infty) \to [0,\delta_{\mathcal{F}}]$ surjective, $\delta_{\mathcal{F}} > 0$, with*

(1) $\sup\{\mathrm{diam}(X),\ X \in \mathcal{F}\} \le D$, *and*
(2) $\inf\{\mu_X(B_X(x,\varepsilon)),\ x \in X\} \ge \rho_{\mathcal{F}}(\varepsilon)$ *for all $X \in \mathcal{F}$ and $\varepsilon \ge 0$.*

Then, $\mathcal{F}$ is pre-compact for the $\mathfrak{D}_p$ topology, for each $p \ge 1$.

This theorem is analogous to Blaschke's Theorem (Theorem 2.1), Prokhorov's Theorem (Theorem 2.2), and Gromov's Pre-compactness Theorem (Theorem 4.1).

Example 5.9 Recall the family $\{\Delta_n,\ n \in \mathbb{N}\}$ of Example 5.2. Condition (1) is satisfied trivially. However, $\{\Delta_n,\ n \in \mathbb{N}\}$ violates condition (2). Indeed, notice that for $\varepsilon \in [0,1)$ one has $\mu_{\Delta_n}(B_X(x,\varepsilon)) = \frac{1}{n}$ for all $x \in \Delta_n$. Hence, $\inf_n \mu_{\Delta_n}(B_X(x,\varepsilon)) = 0$ and there can be no surjective non-negative function from $[0,\infty)$ to $[0,\delta)$, $\delta > 0$, s.t. condition (2) is satisfied.

Remark 5.21 Condition (2) may be ensured by assuming that in addition to satisfying (1), all spaces X in $\mathcal{F}$ are *doubling* with the same *doubling constant* $C_{\mathcal{F}}$:

$$\mu_X\big(B_X(x,2\varepsilon)\big) \le C_{\mathcal{F}} \cdot \mu_X\big(B_X(x,\varepsilon)\big), \quad \text{for all } x \in X,\ \varepsilon \ge 0.$$

Indeed, if this is the case, then $X = B_X(x,D)$ for any $x \in X$ and hence by [2, Theorem 5.2.2]

$$\mu_X\big(B_X(x,\varepsilon)\big) \ge \left(\frac{\varepsilon}{2D}\right)^N$$

for $N = N(\mathcal{F}) := \frac{\log C_{\mathcal{F}}}{\log 2}$. From this point onwards, the discussion in Remark 5.20 applies. A similar Theorem has been proved by Sturm in [105, Theorem 3.12] for $\mathfrak{S}_p$. In his statement, Sturm assumes the uniform doubling condition for all spaces in the family in addition to uniform control of diameters. His conclusion is, actually, compactness, since he also proves that the doubling property is stable with respect to $\mathfrak{S}_p$-convergence.

6 Lower and Upper Bounds and Connections to Other Approaches

In practice, having lower bounds that are easy to compute (in the sense that they are not computationally expensive) is very important as they facilitate classification tasks: If during a query the value of the lower bound is above a certain threshold one would say that the answer to the query is negative without incurring the potentially higher computational cost of evaluating the full metric. In addition to this, it is usual practice to somehow use the "matching" coming from these lower bounds as an initial condition for the iterative algorithm that will optimize the full underlying notion of proximity, see Sect. 7.

Also, with the goal of relating the framework proposed in this paper to other proposals that can be found in the literature, it will be established that several notions of dissimilarity between objects based on comparison of mm-space isomorphism invariants (see Sect. 5.2) are lower or upper bounds to $\mathfrak{D}_p$ (5.9). Some lower and upper bounds have already been presented in Sect. 5, and in particular the relation to $d_{\mathcal{GH}}$ and Sturm's proposal ($\mathfrak{S}_p$) has been discussed there.

In this section, for each $p \in [1,\infty]$, we construct functions $\mathbf{FLB}_p$, $\mathbf{SLB}_p$ and $\mathbf{TLB}_p$ from $\mathcal{G}_w \times \mathcal{G}_w$ to $\mathbb{R}$ for which the following inequalities hold:

$$\mathfrak{S}_p \geq \mathfrak{D}_p \geq \begin{cases} \mathbf{TLB}_p \geq \mathbf{FLB}_p \\ \mathbf{SLB}_p \end{cases}$$

These functions involve computing certain distances between the invariants that we defined in Sect. 5.2. None of these lower bounds is tight: the mm-spaces X and Y of Fig. 8 are such that $\mathbf{FLB}_p(X,Y) = \mathbf{SLB}_p(X,Y) = \mathbf{TLB}_p(X,Y) = 0$, but $\mathfrak{D}_p(X,Y) > 0$.

Remark 6.1 The lower bounds for $\mathfrak{D}_p$ (and hence, by Theorem 5.1 (g), also $\mathfrak{S}_p$) that are computed in this section admit in turn lower bounds of a simplified nature when $p = 1$—related bounds for such case are also discussed. Note that by Theorem 5.1(h), these lower bounds for $\mathfrak{D}_1$ will also be lower bounds to $\mathfrak{D}_p$ for all $p > 1$.

The following technical lemma will be used in this section:

Lemma 6.1 *Let (X, d_X, μ_X) and (Y, d_Y, μ_Y) be two mm-spaces in $\mathcal{G}_w$. Let $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ be continuous and $\phi : \mathbb{R} \to [0,\infty)$ be convex. Then*

$$\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)} \int_{X\times Y} \phi\big(f(x) - g(y)\big)\,\mu(dx \times dy) \geq \int_0^1 \phi\big(F^{-1}(t) - G^{-1}(t)\big)\,dt$$

where $F(t) := \mu_X\{x \in X \mid f(x) \leq t\}$ and $G(t) := \mu_Y\{y \in Y \mid g(y) \leq t\}$ are the distributions *of f and g, respectively, and their* generalized inverses *under μ_X and μ_Y, respectively, are defined as*:

$$F^{-1}(t) = \inf\big\{u \in \mathbb{R} \mid F(u) > t\big\}.$$

Furthermore, when $\phi(u)=|u|$, $u\in\mathbb{R}$, *one can dispense with the inverses*:

$$\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\int_{X\times Y}\big|f(x)-g(y)\big|\mu(dx\times dy)\geq\int_{\mathbb{R}}\big|F(u)-G(u)\big|\,du.$$

Proof Fix $\mu\in\mathcal{M}(\mu_X,\mu_Y)$. Let $h:\mathbb{R}^2\to\mathbb{R}^2$ given by $h=(f,g)$ and consider the measure $\nu=h_\#\mu$ on $\mathbb{R}^2$. By Theorem 4.1.11 of [36] (applied to $T:X\times Y\to\mathbb{R}\times\mathbb{R}$, $(x,y)\mapsto(f(x),g(y))$) one has

$$\int_{X\times Y}\phi\big(f(x)-g(y)\big)\,\mu(dx\times dy)=\int_{\mathbb{R}\times\mathbb{R}}\phi(t-s)\,\nu(dt\times ds).$$

Now, $\nu(I\times\mathbb{R})=\mu(f^{-1}(I)\times g^{-1}(\mathbb{R}))=\mu(f^{-1}(I)\times Y)=\mu_X(f^{-1}(I))$, for any $I\in\mathcal{B}(\mathbb{R})$. Similarly, $\nu(\mathbb{R}\times J)=\mu_Y(g^{-1}(J))$ for any $J\in\mathcal{B}(\mathbb{R})$. Hence, from the equality above,

$$\int_{X\times Y}\phi\big(f(x)-g(y)\big)\mu(dx\times dy)\geq\inf_{\nu\in\mathcal{M}(f_\#\mu_X,g_\#\mu_Y)}\int_{\mathbb{R}\times\mathbb{R}}\phi(t-s)\,\nu(dt\times ds).$$

The conclusion follows from results on the transportation problem on the real line, see Remark 2.19 in [111], and then from the fact that $\mu\in\mathcal{M}(\mu_X,\mu_Y)$ was arbitrary and the right-hand side does not depend on it. □

6.1 First Lower Bound

This section establishes a lower bound for $\mathfrak{D}_p$ using eccentricities.

Recall the definition of eccentricities: for $\infty>p\geq 1$ $s_{X,p}:X\to\mathbb{R}^+$ is given by $x\mapsto\|d_X(x,\cdot)\|_{L^p(\mu_X)}$ and $s_{X,\infty}:X\to\mathbb{R}^+$ given by $x\mapsto\|d_X(x,\cdot)\|_{L^p(R(\mu))}$.

Definition 6.1 (First Lower Bound) For $X,Y\in\mathcal{G}_w$ define

- for $p\in[1,\infty)$:

$$\mathbf{FLB}_p(X,Y):=\frac{1}{2}\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\left(\int_{X\times Y}\big|s_{X,p}(x)-s_{Y,p}(y)\big|^p\,\mu(dx\times dy)\right)^{1/p}.$$

- for $p=\infty$:

$$\mathbf{FLB}_\infty(X,Y):=\frac{1}{2}\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\sup_{(x,y)\in R(\mu)}\big|s_{X,\infty}(x)-s_{Y,\infty}(y)\big|.$$

Then one obtains the following lower bound for $\mathfrak{D}_p$:

Proposition 6.1 *Let* $X,Y\in\mathcal{G}_w$ *and* $p\in[1,\infty]$. *Then*,

$$\mathfrak{D}_p(X,Y)\geq\mathbf{FLB}_p(X,Y).\tag{6.1}$$

Remark 6.2 Note that for $1\leq p<\infty$, solving for $\mathbf{FLB}_p$ leads to a *Mass Transportation Problem* [111] for the cost $c(x,y):=|s_{X,p}(x)-s_{Y,p}(y)|^p$.

Proof of Proposition 6.1 The case $p < \infty$ in (6.1) is a simple consequence of Minkowski's inequality. Indeed, fix $x \in X$, $y \in Y$ and $\mu \in \mathcal{M}(\mu_X, \mu_Y)$. Then, by Minkowski's inequality,

$$\left\|d_Y(y,\cdot)\right\|_{L^p(\mu)} \leq \left\|d_X(x,\cdot)\right\|_{L^p(\mu)} + \left\|\Gamma_{X,Y}(x,y,\cdot,\cdot)\right\|_{L^p(\mu)}.$$

But since μ is a coupling of μ_X and μ_Y, $\|d_X(x,\cdot)\|_{L^p(\mu)} = s_{X,p}(x)$ and $\|d_Y(y,\cdot)\|_{L^p(\mu)} = s_{Y,p}(y)$. From the inequality above, one has $s_{Y,p}(y) \leq s_{X,p}(x) + \|\Gamma_{X,Y}(x,y,\cdot,\cdot)\|_{L^p(\mu)}$. By exchanging the roles of X and Y one obtains

$$\left|s_{X,p}(x) - s_{Y,p}(y)\right| \leq \left\|\Gamma_{X,Y}(x,y,\cdot,\cdot)\right\|_{L^p(\mu)}. \tag{6.2}$$

Now, since $\|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)} \leq 2\max(\mathrm{diam}(X), \mathrm{diam}(Y)) < \infty$, by Fubini's theorem, $\|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)} = (\int_{X\times Y}(\|\Gamma_{X,Y}(x,y,\cdot,\cdot)\|_{L^p(\mu)})^p\,\mu(dx\times dy))^{1/p}$. From (6.2) one finds

$$\|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)} \geq \left(\int_{X\times Y}\left|s_{X,p}(x) - s_{Y,p}(y)\right|^p \mu(dx\times dy)\right)^{1/p}.$$

The claim follows upon noting that the right-hand side admits $2\mathbf{FLB}_p(X,Y)$ as a lower bound and that then in the resulting inequality one can use the fact that μ is arbitrary.

The case $p = \infty$ is a consequence of this simple fact: for functions $f, g : Z \to \mathbb{R}$ it holds that $|\sup f - \sup g| \leq \sup|f - g|$. $\square$

Let $S_{X,p} : \mathbb{R} \to [0,1]$ be given by $t \mapsto \mu_X(\{x \in X \,|\, s_{X,p}(x) \leq t\})$, i.e., $S_{X,p}$ is the distribution function of $s_{X,p}$ under μ_X. Then, invoking Lemma 6.1 with $\phi(u) = |u|^p$ one obtains

Corollary 6.1 (Lower bound based on distribution of eccentricities) *For $p \in [1,\infty)$ and $X, Y \in \mathcal{G}_w$,*

$$\mathbf{FLB}_p(X,Y) \geq \frac{1}{2}\left(\int_0^1 \left|S_{X,p}^{-1}(u) - S_{Y,p}^{-1}(u)\right|^p du\right)^{1/p}.$$

When $p = 1$, one obtains

$$\mathbf{FLB}_1(X,Y) \geq \frac{1}{2}\int_{\mathbb{R}} \left|S_{X,1}(t) - S_{Y,1}(t)\right| dt. \tag{6.3}$$

The lower bounds given in Corollary 6.1 compute distances between distributions of eccentricities. The use of such invariants was introduced in the applied literature by Hamza and Krim in [53].

6.2 Second Lower Bound

This section provides a lower bound for $\mathfrak{D}_p$ (and hence also for $\mathfrak{S}_p$) which relies on a certain comparison of the *distribution of distances* of X and Y (recall Definition 5.4).

Definition 6.2 For $X, Y \in \mathcal{G}_w$ define

- for $p \in [1, \infty)$:

$$\mathbf{SLB}_p(X, Y) := \frac{1}{2} \inf_{\gamma \in \widehat{\mathcal{M}}} \left(\int_{X \times X} \int_{Y \times Y} \left| d_X(\mathcal{X}) - d_Y(\mathcal{Y}) \right|^p \gamma(d\mathcal{X} \times d\mathcal{Y}) \right)^{1/p}.$$

- for $p = \infty$:

$$\mathbf{SLB}_\infty(X, Y) := \frac{1}{2} \inf_{\gamma \in \widehat{\mathcal{M}}} \sup_{((x,x'),(y,y')) \in R(\gamma)} \left| d_X\big(x, x'\big) - d_Y\big(y, y'\big) \right|,$$

where $\widehat{\mathcal{M}} = \mathcal{M}(\mu_X \otimes \mu_X, \mu_Y \otimes \mu_Y)$ stands for the collection of probability measures on $X \times X \bigtimes Y \times Y$ with marginals $\mu_X \otimes \mu_X$ and $\mu_Y \otimes \mu_Y$.

Then one obtains the following bound for $\mathfrak{D}_p$.

Proposition 6.2 *Let $X, Y \in \mathcal{G}_w$ and $p \in [1, \infty]$. Then,*

$$\mathfrak{D}_p(X, Y) \geq \mathbf{SLB}_p(X, Y). \tag{6.4}$$

Proof Fix $\mu \in \mathcal{M}(\mu_X, \mu_Y)$. Let $\mathcal{B}(X \times X)$ denote the σ-algebra on $X \times X$ generated by sets of the form $A \times A'$ where $A, A' \in \mathcal{B}(X)$. Define $\mathcal{B}(Y \times Y)$ in the same way. Then, let $\mathcal{B}(X \times X \bigtimes Y \times Y)$ be the σ-algebra generated by sets of the form $L \times M$ where $L \in \mathcal{B}(X \times X)$ and $M \in \mathcal{B}(Y \times Y)$. Let $\mu \widehat{\otimes} \mu$ be the (unique) measure on $(X \times X \bigtimes Y \times Y, \mathcal{B}(X \times X \bigtimes Y \times Y))$ s.t. $\mu \widehat{\otimes} \mu(A \times A' \bigtimes B \times B') = \mu(A \times B) \cdot \mu(A' \times B')$ for all $A, A' \in \mathcal{B}(X)$ and $B, B' \in \mathcal{B}(Y)$.

Clearly, $\mu \widehat{\otimes} \mu \in \mathcal{M}(\mu_X \otimes \mu_X, \mu_Y \otimes \mu_Y)$. Thus, $\mathbf{SLB}_p(X, Y)$ provides a lower bound for $\mathfrak{D}_p(X, Y)$. A similar claim is true for $p = \infty$. □

It is clear now that this bound is nothing but a measure of distance between the distribution of inter-point distances in X and Y. In fact, one has the following more explicit result.

Corollary 6.2 (Lower bound based on distribution of distances) *For $p \in [1, \infty)$ and any $X, Y \in \mathcal{G}_w$,*

$$\mathbf{SLB}_p(X, Y) \geq \frac{1}{2} \left(\int_0^1 \left| f_X^{-1}(u) - f_Y^{-1}(u) \right|^p du \right)^{1/p}$$

and for $p = 1$ the expression simplifies to

$$\mathbf{SLB}_1(X, Y) \geq \frac{1}{2} \int_0^\infty \left| f_X(t) - f_Y(t) \right| dt. \tag{6.5}$$

Proof The proof follows after a direct application of Lemma 6.1. □

Remark 6.3 (Connection to Shape distributions) The distribution of distances f_X is related to one of the shape signatures computed by Osada et al. in the influential *shape distribution* approach to comparing objects [86]. In [86] the authors propose to characterize an object X by (an estimate of) the probability density of the distance between two randomly selected points on the object X. Of course, f_X is the *cumulative* version of the shape distribution. The line started in [86] was pursued in more theoretical terms, for the case of finite Euclidean metric sets by Boutin and Kemper in [11]. The second lower bound, $\mathbf{SLB}_p$, together with the counterexamples given in Sect. 5.1 make explicit the fact that the distributions of distances are important invariants for the discrimination of objects; yet, they are not complete.

Example 6.1 Recall the mm-spaces Δ_2 and Δ_3 of Example 5.5. Applying the bound (6.5) given by the Corollary above for $p=1$, one finds $\mathbf{SLB}_1(\Delta_2,\Delta_3)\geq\frac{1}{12}$.

Example 6.2 The following lower bound holds: $\mathfrak{D}_1(\mathbb{S}^1,\mathbb{S}^2)\geq\frac{1}{2}(1-\pi/4)$. Indeed, notice that according to Example 5.7, $f_{\mathbb{S}^1}(t)=\frac{t}{\pi}$ and $f_{\mathbb{S}^2}(t)=\frac{1-\cos t}{2}$, for $t\in[0,\pi]$. Let $h(t):=f_{\mathbb{S}^1}(t)-f_{\mathbb{S}^2}(t)$ and note that $h(\pi/2+\alpha)+h(\pi/2-\alpha)=0$ for all $\alpha\in[0,\pi/2]$. Hence

$$\int_0^\infty \left|f_{\mathbb{S}^1}(t)-f_{\mathbb{S}^2}(t)\right|dt=2\int_0^{\pi/2}\left(\frac{t}{\pi}-\frac{1-\cos t}{2}\right)dt=1-\frac{\pi}{4}.$$

Therefore, by Proposition 6.2 and (6.5) one obtains that $\mathfrak{D}_1(\mathbb{S}^1,\mathbb{S}^2)\geq\mathbf{SLB}_1(\mathbb{S}^1,\mathbb{S}^2)\geq\frac{1}{2}(1-\frac{\pi}{4})\simeq 0.1073$. Note that since $\mathfrak{D}_2(\mathbb{S}^1,\mathbb{S}^2)\geq\mathfrak{D}_1(\mathbb{S}^1,\mathbb{S}^2)$ this lower bound is tighter than the one computed via 2-diameters in Remark 5.16.

6.3 Third Lower Bound

This section established a lower bound for $\mathfrak{D}_p$ based on local distributions of distances. For $p\in[1,\infty]$ and $X,Y\in\mathcal{G}_w$ let $\Omega_p: X\times Y\to\mathbb{R}$ be given by

$$\Omega_p(x,y):=\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\left\|\Gamma_{X,Y}(x,y,\cdot,\cdot)\right\|_{L^p(\mu)}\quad\text{for } p<\infty \tag{6.6}$$

and

$$\Omega_\infty(x,y):=\inf_{\mu\in\mathcal{M}}\left\|\Gamma_{X,Y}(x,y,\cdot,\cdot)\right\|_{L^\infty(R(\mu))}. \tag{6.7}$$

Definition 6.3 Let $X,Y\in\mathcal{G}_w$. Define

- for $p\in[1,\infty)$:

$$\mathbf{TLB}_p(X,Y):=\frac{1}{2}\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\left(\int_{X\times Y}\left(\Omega_p(x,y)\right)^p\mu(dx\times dy)\right)^{1/p}.$$

- for $p=\infty$:

$$\mathbf{TLB}_\infty(X,Y):=\frac{1}{2}\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\sup_{(x,y)\in R(\mu)}\Omega_\infty(x,y).$$

Proposition 6.3 *Let $X, Y \in \mathcal{G}_w$ and $p \in [1, \infty]$. Then,*

$$\mathfrak{D}_p(X, Y) \geq \mathbf{TLB}_p(X, Y). \tag{6.8}$$

Proof We only show details for $p \in [1, \infty)$. Clearly, for any $\mu \in \mathcal{M}(\mu_X, \mu_Y)$, by Fubini's theorem, the definition of Ω_p, and the definition of $\mathbf{TLB}_p$:

$$\begin{aligned} 2 \cdot \mathbf{J}_p(\mu) &= \|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)} = \left(\int_{X\times Y} \left(\left\|\Gamma_{X,Y}(x, y, \cdot, \cdot)\right\|_{L^p(\mu)}\right)^p \mu(dx \times dy)\right)^{1/p} \\ &\geq \left(\int_{X\times Y} \left(\Omega_p(x, y)\right)^p \mu(dx \times dy)\right)^{1/p} \\ &\geq 2 \cdot \mathbf{TLB}_p(X, Y). \end{aligned}$$

The conclusion follows since μ was arbitrary in $\mathcal{M}(\mu_X, \mu_Y)$. □

Remark 6.4 ($\mathbf{TLB}_p$ and Lawler's lower bound for the QAP) This lower bound is reminiscent of Lawler's lower bound in the QAP literature [89]; see Remark 4.6. To the best of our knowledge this is the first time this bound is used in the context of Object Matching/Comparison.

Remark 6.5 From (6.2) in the proof of Proposition 6.1, it is easy to see that

$$\mathbf{TLB}_p(X, Y) \geq \mathbf{FLB}_p(X, Y)$$

for all $p \geq 1$ and $X, Y \in \mathcal{G}_w$.

From the lower bound (6.8) one can go one step further and produce another lower bound of more practical appeal. In fact, Corollary 6.3 below relates the framework discussed in this paper to the very attractive idea of estimating the $\mathfrak{D}_p$ distance from below by comparing the local shape distributions that were introduced in Definition 5.5. Recall that one denotes the local shape distribution of an mm-space X at point $x \in X$ by $h_X(x, \cdot)$. For fixed $x \in X$, by $h_X^{-1}(x, \cdot)$ we denote the generalized inverse of $h_X(x, \cdot)$ with respect to its (second) argument, see Lemma 6.1.

Corollary 6.3 (Lower bound based on local distributions of distances) *Let $X, Y \in \mathcal{G}_w$ and $p \in [1, \infty)$. Then*

$$\begin{aligned} &\mathbf{TLB}_p(X, Y) \\ &\quad \geq \frac{1}{2} \min_{\mu\in\mathcal{M}(\mu_X,\mu_Y)} \int_{X\times Y} \left(\int_0^1 \left|h_X^{-1}(x, u) - h_Y^{-1}(y, u)\right|^p du\right)^{1/p} \mu(dx \times dy). \end{aligned} \tag{6.9}$$

For $p = 1$ we obtain

$$\mathbf{TLB}_1(X, Y) \geq \frac{1}{2} \min_{\mu\in\mathcal{M}(\mu_X,\mu_Y)} \int_{X\times Y} \left(\int_0^\infty \left|h_X(x, t) - h_Y(y, t)\right| dt\right) \mu(dx \times dy). \tag{6.10}$$

Proof Recall (6.7) and note that by Lemma 6.1 (for $\phi(u) = |u|^p$) for all $(x, y) \in X \times Y$,

$$\big(\Omega_p(x, y)\big)^p \geq \int_0^1 \big|h_X^{-1}(x, u) - h_Y^{-1}(y, u)\big|^p \, du.$$

For $p = 1$, by Fubini's Theorem, the right-hand side above admits, as before, a simpler expression:

$$\Omega_1(x, y) \geq \int_0^\infty \big|h_X(x, t) - h_Y(y, t)\big| \, dt.$$

The conclusion follows by recalling Definition 6.3 and by using routine arguments. □

Remark 6.6 It is obvious that the right-hand side of expression (6.10) above computes a certain dissimilarity between the "cumulative histograms of distances" (local shape distributions), h_X and h_Y. Recall the definition of local shape distributions in Definition 5.5 and see there the connection with the "shape context" idea.

Ideas of this kind have been used in protein docking [95] and matching of hippocampal surfaces [102], see Remark 5.6 in Sect. 5.2. More recently, an ad hoc procedure based on the comparison of invariants similar to the local shape distributions has been investigated in [97]. The lower bound (6.10) thus embodies a method that is compatible with the metrics $\mathfrak{D}_p$ and $\mathfrak{S}_p$. Furthermore, this idea leads to solving a transportation problem with the cost $c(x, y) := \|h_X(x, \cdot) - h_X(y, \cdot)\|_{L^1(\mathbb{R}^+)}$ which can be solved efficiently in practice. Indeed, transportation problems, also known as *Hitchcock transportation problems* are quite common in the optimization literature and standard references such as [88] describe specialized algorithms for solving these problems which run in polynomial time.

6.4 Upper Bounds

Assume that one wants to compare compact subsets of Euclidean space $\mathbb{R}^d$ under invariance to the group $E(d)$ of rigid isometries. Then, following Sect. 3, a suitable notion of similarity between objects (X, μ_X) and (Y, μ_Y) in $\mathcal{C}_w(\mathbb{R}^d)$ is

$$E_{\mathcal{W},p}(X, Y) := \inf_{T \in \mathcal{T}_Z} d_{\mathcal{W},p}^{\mathbb{R}^k}\big(X, T(Y)\big).$$

Consider the mm-spaces $X' = (X, \|\cdot\|, \mu_X)$ and $Y' = (Y, \|\cdot\|, \mu_Y)$, then by simple application of Minkowski's inequality (for the norm $\|\cdot\|$) one can easily prove that (recall Theorem 5.1(c))

$$\mathfrak{D}_p\big(X', Y'\big) \leq E_{\mathcal{W},p}\big(X', Y'\big).$$

This bound relates $\mathfrak{D}_p$ to ideas in [31, 65] and references therein. A similar claim holds true when one use $d_{\mathcal{H}}^{\mathbb{R}^d}$ instead of $d_{\mathcal{W},p}^{\mathbb{R}^d}$ and $d_{\mathcal{GH}}$ instead of $\mathfrak{D}_p$. Based on work by Alestalo et al. [1], reverse inequalities have been obtained in [76] for the case $d_{\mathcal{H}}$ and $d_{\mathcal{GH}}(,)$; and $d_{\mathcal{W},p}$ and $\mathfrak{S}_p$. Precise statement for the pairs $d_{\mathcal{H}}^{\mathbb{R}^d}$ and $d_{\mathcal{GH}}$, and $d_{\mathcal{W},p}^{\mathbb{R}^d}$ and $\mathfrak{S}_p$ are

Theorem 6.1 [76] *Let $X, Y \subset \mathbb{R}^d$ be compact. Then*

$$d_{\mathcal{GH}}\big((X, \|\cdot\|), (Y, \|\cdot\|)\big) \leq \inf_{T\in E(d)} d_{\mathcal{H}}^{\mathbb{R}^d}\big(X, T(Y)\big)$$
$$\leq c_d \cdot M^{\frac{1}{2}} \cdot \big(d_{\mathcal{GH}}\big((X, \|\cdot\|), (Y, \|\cdot\|)\big)\big)^{\frac{1}{2}}$$

where $M = \max(\mathrm{diam}(X), \mathrm{diam}(Y))$ and c_d is a constant that depends only on d.

The $\frac{1}{2}$-exponent is optimal, see [76].

Theorem 6.2 [76] *Let $X, Y \subset \mathbb{R}^d$ be compact weighted objects. Then*

$$\mathfrak{S}_p\big((X, \|\cdot\|), (Y, \|\cdot\|)\big) \leq \inf_{T\in E(d)} d_{\mathcal{W},p}^{\mathbb{R}^d}\big(X, T(Y)\big)$$
$$\leq c'_d \cdot M^{\frac{3}{4}} \cdot \big(\mathfrak{S}_p\big((X, \|\cdot\|), (Y, \|\cdot\|)\big)\big)^{\frac{1}{4}}$$

where $M = \max(\mathrm{diam}(X), \mathrm{diam}(Y))$ and c'_d is a constant that depends only on d.

7 Computational Technique

This section deals with the practical computation of $\mathfrak{D}_p$. We recast the discrete counterpart of the ideas we propose as (continuous) optimization problems.

Assume that finite mm-spaces $\mathbb{X} = \{x_1, \ldots, x_{n_\mathbb{X}}\}$ and $\mathbb{Y} = \{y_1, \ldots, y_{n_\mathbb{Y}}\}$ with metrics $d_\mathbb{X}$ and $d_\mathbb{Y}$, respectively, and probability measures $\mu_\mathbb{X}$ and $\mu_\mathbb{Y}$, respectively, are given. Let

$$\mathcal{M} := \left\{ \mu \in \mathbb{R}_+^{n_\mathbb{X}\times n_\mathbb{Y}} \,\middle|\, 0 \leq \mu_{ij} \leq 1,\ \text{where} \left\{ \begin{array}{c} \sum_j \mu_{i,j} = \mu_\mathbb{X}(x_i), \\ \text{and} \\ \sum_i \mu_{i,j} = \mu_\mathbb{Y}(y_j), \end{array} \right\} \right.$$
$$\left. \text{for all } 1 \leq i \leq n_\mathbb{X}, 1 \leq j \leq n_\mathbb{Y} \right\}.$$

Note that the number of *linear* constraints in $\mathcal{M}$ is $(n_\mathbb{X} + n_\mathbb{Y})$. Let $p \in [1, \infty)$. Then the problem that one needs to solve is

$$(P_p) \left\{ \begin{array}{l} \min_{\mu\in\mathcal{M}} \mathbf{H}_p(\mu) \\ \mathbf{H}_p(\mu) := \sum_{i,i'=1}^{n_\mathbb{X}} \sum_{j,j'=1}^{n_\mathbb{Y}} \mu_{i,j}\, \mu_{i',j'} |d_\mathbb{X}(x_i, x_j) - d_\mathbb{Y}(y_{i'}, y_{j'})|^p. \end{array} \right.$$

Problem (P_p) is a QOP (with linear constraints), albeit not necessarily convex. Nevertheless, there exist many techniques in the literature for handling this kind of problems. For the computation of examples presented in Sect. 8, we implemented an alternate optimization procedure [73] which relies on solving successive LOPs and which was initialized by solving the problem $\mathbf{FLB}_p$ (see below). We used the Matlab interface [43] for the open source LOP solver `glpk` and `YALMIP` as an interpreter [71].

Let μ^* be the measure coupling that one obtains upon convergence of the method. We then estimate $\mathfrak{D}_p(\mathbb{X},\mathbb{Y}) \simeq \frac{1}{2}(\mathbf{H}_p(\mu^*))^{1/p}$.

Remark 7.1 (About the computational cost of $\mathfrak{D}_p$ vs. $\mathfrak{S}_p$) If one were to try to compute $\mathfrak{S}_p(\mathbb{X},\mathbb{Y})$, the resulting optimization problem would be significantly harder. In fact, the problem would read

$$(S_p)\begin{cases}\min_{(\mu,d)\in\mathcal{M}\times\mathcal{D}}\mathbf{I}_p(\mu,d)\\ \mathbf{I}_p(\mu,d):=\sum_{i=1}^{n_{\mathbb{X}}}\sum_{j=1}^{n_{\mathbb{Y}}}\mu_{i,j}\,d_{i,j}^p\end{cases}$$

where

$$\mathcal{D}:=\left\{d\in\mathbb{R}_+^{n_{\mathbb{X}}\times n_{\mathbb{Y}}}\text{s.t.}\begin{Bmatrix}|d_{ij}-d_{i'j}|\le d_{\mathbb{X}}(x_i,x_{i'})\le d_{ij}+d_{i'j}\\ \text{and}\\ |d_{ij}-d_{ij'}|\le d_{\mathbb{Y}}(y_j,y_{j'})\le d_{ij}+d_{ij'}\end{Bmatrix}\right.$$
$$\left.\text{for all } 1\le i,i'\le n_{\mathbb{X}},\ 1\le j,j'\le n_{\mathbb{Y}}\right\}.$$

Note that the number of linear constraints in $\mathcal{D}$ is $N_S:=2(n_{\mathbb{Y}}\binom{n_{\mathbb{X}}}{2}+n_{\mathbb{X}}\binom{n_{\mathbb{Y}}}{2})\gg(n_{\mathbb{X}}+n_{\mathbb{Y}})$. Hence, for solving (S_p) one needs $2(n_{\mathbb{X}}\times n_{\mathbb{Y}})$ variables and N_S constraints as opposed to $(n_{\mathbb{X}}\times n_{\mathbb{Y}})$ variables and $(n_{\mathbb{X}}+n_{\mathbb{Y}})$ constraints for solving (P_p). Therefore, from the practical point of view, this justifies singling out $\mathfrak{D}_p$ as a more convenient choice. Also, it is worth mentioning that problem (S_p) is a Bilinear Optimization problem, which can obviously be reformulated as a QOP.

Nevertheless, there are situations when (S_p) is more tractable; see the discussion in [76].

7.1 Computation of the Lower Bounds

It is useful to recast the lower bounds discussed in Sect. 6.1 in this optimization setting. Consider the optimization problem:

$$(\mathbf{FLB}_p)\begin{cases}\min_{\mu\in\mathcal{M}}\mathbf{L}_p(\mu)\\ \mathbf{L}_p(\mu):=\frac{1}{2}\sum_{i=1}^{n_{\mathbb{X}}}\sum_{j=1}^{n_{\mathbb{Y}}}\mu_{ij}|s_{\mathbb{X},p}(i)-s_{\mathbb{Y},p}(j)|\end{cases}$$

where $s_{\mathbb{X},p}(i):=(\sum_{k=1}^{n_{\mathbb{X}}}\mu_{\mathbb{X}}(k)\,(d_{\mathbb{X}}(x_i,x_k))^p)^{1/p}$ and $s_{\mathbb{Y},p}(j)=(\sum_{k=1}^{n_{\mathbb{Y}}}\mu_{\mathbb{Y}}(y_k)\times(d_{\mathbb{Y}}(y_j,y_k))^p)^{1/p}$ for $1\le i\le n_{\mathbb{X}}$ and $1\le j\le n_{\mathbb{Y}}$.

While we do not do it here explicitly, it is clear that the discrete formulations of $\mathbf{SLB}_p$ and $\mathbf{TLB}_p$ also lead to LOPs. In the latter case, however, one needs to solve $n_{\mathbb{X}}\times n_{\mathbb{Y}}$ LOPs over the variable $\mu\in\mathcal{M}$.

8 Computational Examples

This section presents some computational examples that exemplify the use of the mm-space framework based on the $\mathfrak{D}_p$ distance.

8.1 An Example with Ultrametric Spaces

In many applications one has to deal with very large datasets that cannot be processed at once. Clustering is one of the standard tools for data analysis and hierarchical clustering is one of standard ways of performing clustering. Hierarchical methods of clustering give outputs in the form of *dendrograms*: nested partitions of the input set which are parametrized by a scale parameter arising from the metric structure of the input data. These dendrograms are rooted trees, and as such admit a representation using *ultrametrics* [24], which are a special type of metrics that satisfy a stronger version of the triangle inequality.

In order to deal with the massiveness of the datasets, often, practitioners apply methods akin to the *bootstrap* that consist of taking several different samples of the underlying dataset, and applying the data analysis procedure individually to each of these smaller, more manageable, datasets. The problem that must be dealt with later on is that of detecting whether there is some sort of agreement between the different outputs that correspond to each of the samples. This requires being able to measure dissimilarity between such outputs, and in the case of hierarchical clustering, one needs to measure dissimilarity between finite ultrametric spaces. Notice that there may be no way of using the labels of the points in any reasonable manner. It is the case, namely, that the measure of dissimilarity that one chooses must be able to recognize structure in a manner that is insensitive to labels of the points.

This example shows an example of computing distance between ultrametric spaces that correspond to a collection of 12 dendrograms that exhibit six clearly different behaviors. The different dendrograms correspond to different number of points and also different "shapes". They were generated by clustering randomly generated Euclidean point clouds with varying number of clusters. See Fig. 9 for more details. The algorithm applied to each finite metric space was single linkage hierarchical clustering.

Let $\mathbf{U}$ be a 12×12 matrix where the entry $\mathbf{U}(i, j)$ is the result of computing the distance $\mathfrak{D}_1$ between dendrograms i and j.

Figure 9 shows all the dendrograms and in the middle it shows a dendrogram corresponding to Single Linkage clustering applied to $\mathbf{U}$ (this would be a clustering of clusterings). Here, all ultrametric spaces representing the dendrograms were normalized to have diameter one. This makes the comparison more challenging and also more meaningful, since then the use of the GW distance will now only distinguish by "shape" and not by scale. The choice of probability measures used was uniform, that is, all points in each ultrametric space have weight equal to $1/(\text{cardinality of space})$.

Handling ultrametric spaces is possible and justified with the approach that is described in this paper. The implementation of the Gromov–Wasserstein distance that was described does not impose any restriction on the nature of the metric structure

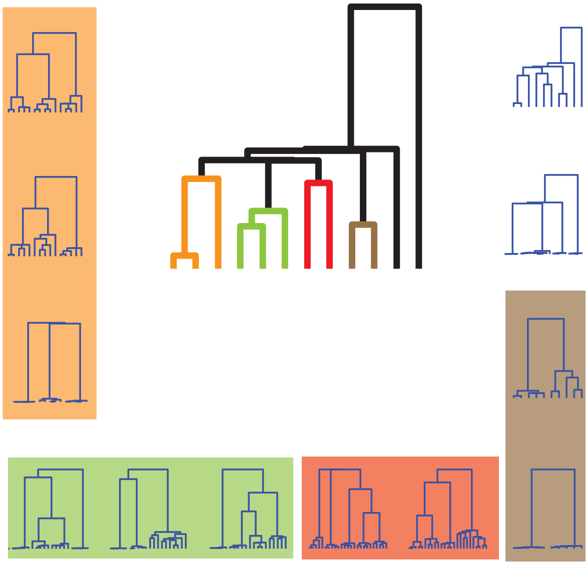

Fig. 9 (Color online) A clustering of hierarchical clusterings. The colors of the boxes enclosing groups of dendrograms correspond to the colors in the branches in the dendrogram *in the center*. Observe that the *brown box* contains dendrograms that show two clear clusters, the *red box* contains dendrograms that show five clusters, the *orange box* three clusters, the *green box* three or four clusters. The *small dendrograms*, *from left to right* correspond to the leaves of the *big dendrogram in the center*, *from left to right*. Notice that there are two *dendrograms on the right* that, judging from the dendrogram arising from **U**, do not seem to cluster with any other dendrogram. Visually, this makes sense given that the shapes of the other dendrograms are quite different

attached to the objects. This is in sharp contrast with the estimation to the Gromov–Hausdorff distance provided by the so called GMDS (generalized metric multidimensional scaling) [17, 18] which requires triangulated shapes and a metric smoothly varying on them, which is not the case for ultrametric spaces.

8.2 Matching of 3D Objects

We used the publicly available (triangulated) objects database [106]. This database comprises 72 objects from seven different classes: `camel`, `cat`, `elephant`, `faces`, `heads`, `horse` and `lion`.

Each class contains several different *poses* of the same object. These poses are richer than just rigid isometries, see Fig. 10 for an example of what these objects look like in the case of the camel models. The number of vertices in the models ranged from 7 K to 30 K. From each model X_k 4,000 points were selected using the Euclidean *farthest point sampling* procedure. Briefly, one first randomly chooses a point from the dataset. Then one chooses the second point as the one at maximal distance from the first one. Subsequent points are chosen always to maximize the minimal distance to the points already chosen. Let $\widehat{X}_k$ denote this reduced model. Then an intrinsic distance (or *graph distance*) was defined using Dijkstra's algorithm on the graph $G(X_k)$ with vertex set X_k, where each vertex is connected by an edge

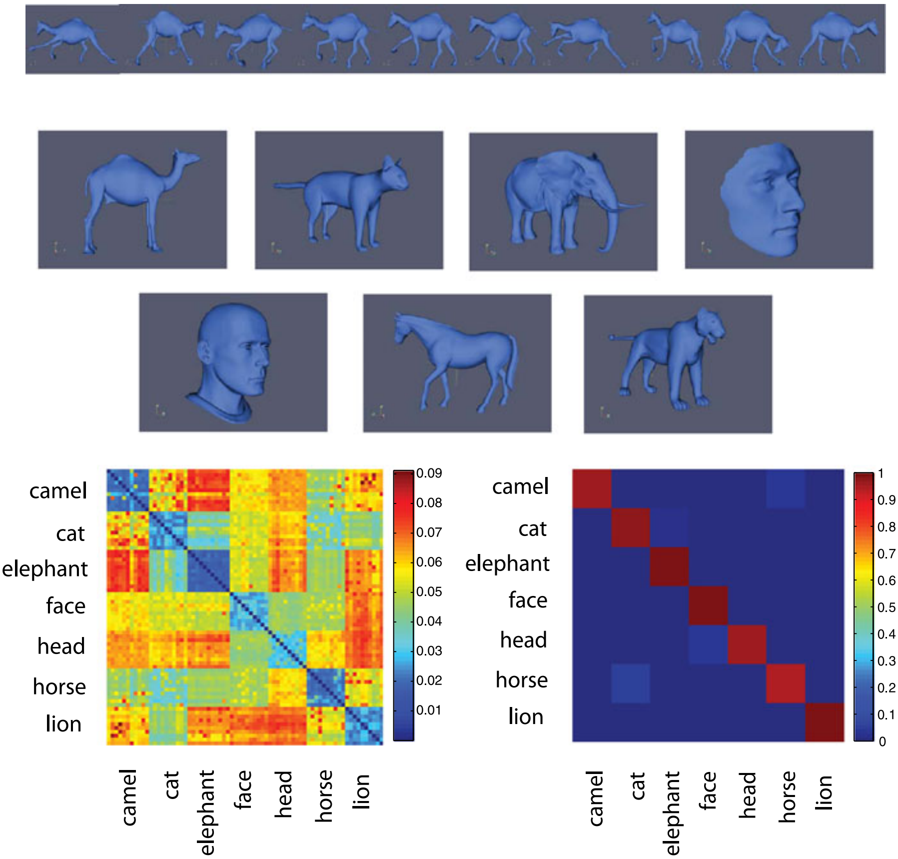

Fig. 10 *Top row*: As an example this figure show all the 10 poses of the camel model object. The *second* and *third row* show one pose of each of the objects in the database. *Last row*, *left*: This panel shows $((d_{ij}))$, see the text for more details. *Right*: Estimated confusion matrix for the 1-nearest neighbor classification problem described in the text (these figures are best appreciated in the online version of the paper)

to those vertices with which it shares a triangle. Since $\widehat{X}_k \subset X_k$, by restriction, one endows $\widehat{X}_k$ with this intrinsic distance as well. We further sub-sample $\widehat{X}_k$, again using the farthest point procedure (with the distance computed using $G(X_k)$), and we retain only 50 points. Denote the resulting set by $\mathbb{X}_k$. We then endowed $\mathbb{X}_k$ with the *normalized* distance metric inherited from the Dijkstra procedure described above, and a probability measure based on Voronoi partitions: the mass (measure) at point $x \in \mathbb{X}_k$ equals the proportion of points in $\widehat{X}_k$ which are closer to x than to any other point in $\mathbb{X}_k$.

From each model X_k one thus obtains a discrete mm-space $(\mathbb{X}_k, d^{(k)}, \nu^{(k)})$. A matrix $((d_{ij}))$ is then computed, such that $d_{ij} = \mathfrak{D}_1(\mathbb{X}_i, \mathbb{X}_j)$ (that is, the value $p = 1$ was fixed) where $1 \leq i < j \leq N$ and $N = 72$ (all models had between nine and 11 poses). See Fig. 10 for a graphical representation of $((d_{ij}))$.

In order to evaluate the discriminative power contained in $((d_{ij}))$ we considered a classification task as follows: one randomly selects one object from each class, form a *training set* T and use it for performing 1-nearest neighbor classification (where *nearest* is with respect to the metric $((d_{ij}))$ of the remaining objects. By simple comparison between the class predicted by the classifier and the actual class to which the object belongs, one thus obtains an estimate for the probability of misclassification $P_e(((d_{ij})))$. This procedure is then repeated for 10,000 random choices of the training set. Using the same randomized procedure one obtains an estimate of the *confusion matrix* C for this problem. That is, C_{ij} equals the probability that the classifier will assign class j to an object when the actual class was i. I also evaluated the performance of $\mathbf{FLB}_1$ using this method. The results that were obtained are $P_e(((d_{ij}))) = 0.025$ and $P_e(\mathbf{FLB}_1) = 0.141$. Refer to Fig. 10 for more details.

9 Discussion

This paper introduced a modification and expansion of the original Gromov–Hausdorff notion of distance between metric spaces which takes into account probability measures defined on measurable subsets of these metric spaces. The construction was explained as a relaxation of the concept of Gromov–Hausdorff distances and as an alternative to the relaxation proposed by Sturm. This paper also studied the theoretical properties of the proposed version of the Gromov–Wasserstein distance, including its topological equivalence to Sturm's proposal. Also, stronger types of equivalence were obtained in restricted classes of objects. Also, sufficient conditions were given for the pre-compactness of families of metric measure spaces in the topology induced by the metric that was defined in this work.

The new definition allows for a discretization which is more natural and more general than previous approaches, [17, 80, 81] in that the methods can be applied to datasets that do not enjoy smoothness.

In addition to this, several previous approaches to the problem of Object Matching/Comparison become interrelated when put into the framework presented in this paper. This interrelation should be understood in the sense that a number of the shape signatures or invariants used by previously reported methods

(1) can be expressed in the formalism of metric measure spaces, and
(2) they are quantitatively stable under perturbations in the Gromov–Wasserstein distance.

Computational experiments on

(1) a collection of dendrograms, and
(2) a database of objects, were presented to exemplify the applicability of the ideas.

Related approaches have appeared in [27, 76, 77, 79, 87]. Further developments such as the extension of the ideas here presented to *partial* object matching, the stability of more general invariants, and finer metric properties of $(\mathcal{G}_w, \mathfrak{D}_p)$ will be reported elsewhere.

10 Proofs

Please refer to Sect. 1.4 for the notation and background concepts and to the notation key in Sect. 11 as well.

Lemma 10.1 *Let X and Y be compact metric spaces and $S \subset X \times Y$ be non-empty. Assume that $\frac{1}{2}\|\Gamma_{X,Y}\|_{L^\infty(S\times S)} < \eta$. Define $d_S : X \times Y \to \mathbb{R}^+$ by*

$$(x, y) \mapsto \inf_{(x',y')\in S} \big(d_X(x, x') + d_Y(y', y)\big) + \eta.$$

Define d_S on $Y \times X$ as $d_S(y, x) = d_S(x, y)$ and on $X \times X$ and $Y \times Y$ by $d_S = d_X$ and $d_S = d_Y$. Then, $d_S \in \mathcal{D}(X, Y)$. Furthermore,

- *$d_S(x, y) \le \eta$ for all $(x, y) \in S$, and*
- *$d_{\mathcal{H}}^{(X\sqcup Y,d_S)}(X, Y) \le \eta + d_{\mathcal{H}}^X(\pi_1(S), X) + d_{\mathcal{H}}^Y(\pi_2(S), Y)$.*

Proof The proof of this lemma is standard, see [22, p. 258] or [61]. □

10.1 The Proof of Theorem 5.1

Lemma 10.2 *For fixed $X, Y \in \mathcal{G}_w$ and $\mu \in \mathcal{M}(\mu_X, \mu_Y)$,*

(1) *$\mathbf{J}_q(\mu) \le \mathbf{J}_p(\mu)$ for all $1 \le q \le p \le \infty$.*
(2) *$\mathbf{J}_p(\mu) \xrightarrow{p\uparrow\infty} \mathbf{J}_\infty(\mu)$.*

Proof The proof is standard and we omit it. □

Lemma 10.3 *Let (Z, d) be a compact metric space and $\mathcal{M}$ be subset of $\mathcal{P}(Z)$ which is sequentially compact for the weak convergence. Let $\phi : Z \times Z \to \mathbb{R}$ be Lipschitz for the L^1 metric on $Z \times Z$: $\widehat{d}((z_1, z_2), (z_1', z_2')) = d(z_1, z_1') + d(z_2, z_2')$ for all $(z_1, z_2), (z_1', z_2') \in Z \times Z$. Let $I(\mu) := \iint_{Z\times Z} \phi(z, z')\, \mu(dz)\mu(dz')$.*

(1) *Let $p_\mu(z) := \int_Z \phi(z, z')\, \mu(dz')$. If $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(Z)$ is such that $\mu_n \xrightarrow{w,n} \mu$, then $p_{\mu_n} \overset{n}{\to} p_\mu$ uniformly.*
(2) *If $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(Z)$ is such that $\mu_n \xrightarrow{w,n} \mu$, then $I(\mu_n) \to I(\mu)$ as $n \uparrow \infty$.*
(3) *There exist a minimizer of $I(\cdot)$ in $\mathcal{M}$.*

Proof

(1) Let K be the Lipschitz constant of ϕ. Let $z_1, z_2 \in Z$, then

$$\begin{aligned}\big|p_{\mu_n}(z_1) - p_{\mu_n}(z_2)\big| &\le \int_Z \big|\phi(z_1, z') - \phi(z_2, z')\big|\mu_n(z')\\ &\le \max_{z'} \big|\phi(z_1, z') - \phi(z_2, z')\big| \le K\, d(z_1, z_2).\end{aligned}$$

Also, $p_{\mu_n} \overset{n}{\to} p_\mu$ pointwise since $\mu_n \overset{n,w}{\to} \mu$. Fix $\varepsilon > 0$ and let $z_1, \ldots, z_N$ be an ε/K net of Z. Also, let $n_\varepsilon \in \mathbb{N}$ be such that $|p_{\mu_n}(z_i) - p_\mu(z_i)| \le \varepsilon$ for all

$i = 1, \dots, N$ and $n > n_\varepsilon$. Then by the triangle inequality, using the inequality above and taking limit as $n \uparrow \infty$ one finds $|p_{\mu_n}(z) - p_\mu(z)| < 3\varepsilon$ for all $z \in Z$ and $n > n_\varepsilon$ and hence the claim.

(2) Let $C_n := I(\mu) - I(\mu_n) = \int_Z p_\mu \, d\mu - \int_Z p_{\mu_n} \, d\mu_n$ where the last equality holds by Fubini's theorem. Also, define $A_n := \int_Z (p_\mu - p_{\mu_n}) \, d\mu_n$ and $B_n = \int_Z p_\mu \, d\mu - \int_Z p_\mu \, d\mu_n$. Note that $A_n \xrightarrow{n} 0$. Indeed, $|\int_Z (p_\mu - p_{\mu_n}) \, d\mu_n| \leq \max_z |p_\mu(z) - p_{\mu_n}(z)|$ and the right-hand side vanishes as $n \uparrow \infty$ by (1). Similarly, $B_n \xrightarrow{n} 0$ simply by weak convergence of μ_n to μ. Finally, since $C_n = A_n + B_n$, $|C_n| \leq |A_n| + |B_n| \xrightarrow{n} 0$.

(3) This follows immediately from the assumptions on $\mathcal{M}$ and (2). □

Corollary 10.1 *Fix each $p \in [1, \infty]$. Given (X, d_X, μ_X) and $(Y, d_Y, \mu_Y) \in \mathcal{G}_w$, there always exist a coupling $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ s.t. $\mathfrak{D}_p(X, Y) = \mathbf{J}_p(\mu)$.*

Proof

Case $1 \leq p < \infty$ It suffices to establish the sequential compactness of $\mathcal{M}(\mu_X, \mu_Y)$ and the continuity of $\mathbf{J}_p(\cdot)$ with respect to the weak convergence of measures. The former is standard, see [111, p. 49]. The latter follows from Lemma 10.3 for $Z = X \times Y$ and $\phi((x, y), (x', y')) = (\Gamma_{X,Y}(x, y, x', y'))^p$, together with the following choice of metric on Z: $d((x, y), (x', y')) = d_X(x, x') + d_Y(y, y')$. We check that ϕ is Lipschitz with respect to the metric $\widehat{d}$ on $X \times Y \times X \times Y$. Let $M = 2 \max(\operatorname{diam}(X), \operatorname{diam}(Y))$. Note first that $\phi_1((x, y), (x', y')) := \Gamma_{X,Y}(x, y, x', y')$ is Lipschitz with constant 1: indeed pick $((x_i, y_i), (x_i', y_i')) \in X \times Y \times X \times Y$ for $i = 1$ and 2. Then,

$$
\begin{aligned}
&\left|\phi_1\big((x_1, y_1), (x_1', y_1')\big) - \phi_1\big((x_2, y_2), (x_2', y_2')\big)\right| \\
&\quad = \left\| d_X\big(x_1, x_1'\big) - d_Y\big(y_1, y_1'\big)\right| - \left| d_X\big(x_2, x_2'\big) - d_Y\big(y_2, y_2'\big)\right\| \\
&\quad \leq \left| d_X\big(x_1, x_1'\big) - d_X\big(x_2, x_2'\big) + d_Y\big(y_1, y_1'\big) - d_Y\big(y_2, y_2'\big)\right| \\
&\quad \leq \left| d_X\big(x_1, x_1'\big) - d_X\big(x_2, x_2'\big)\right| + \left| d_Y\big(y_1, y_1'\big) - d_Y\big(y_2, y_2'\big)\right| \\
&\quad \overset{*}{\leq} \big(d_X(x_1, x_2) + d_X\big(x_1', x_2'\big)\big) + \big(d_Y(y_1, y_2) + d_Y\big(y_1', y_2'\big)\big) \\
&\quad = \big(d_X(x_1, x_2) + d_Y(y_1, y_2)\big) + \big(d_X\big(x_1', x_2'\big) + d_Y\big(y_1', y_2'\big)\big) \\
&\quad = d\big((x_1, y_1), (x_2, y_2)\big) + d\big(\big(x_1', y_1'\big), \big(x_2', y_2'\big)\big) \\
&\quad = \widehat{d}\big(\big((x_1, y_1), \big(x_1', y_1'\big)\big), \big((x_2, y_2), \big(x_2', y_2'\big)\big)\big),
\end{aligned}
$$

where inequality $\overset{*}{\leq}$ follows from the triangle inequalities for d_X and d_Y, respectively. Notice that by definition of ϕ_1 and the choice of M, $0 \leq \phi_1 \leq M$. Let $f : [0, M] \to \mathbb{R}^+$ be given by $t \mapsto t^p$. Then f is Lipschitz with constant bounded by pM^{p-1}. Hence, since $\phi = f \circ \phi_1$, ϕ is also Lipschitz with constant pM^{p-1}.

Case $p = \infty$ Pick $r, q \in [1, \infty)$ s.t. $r \geq q$. Let μ_r be a minimizer of $\mathbf{J}_r$ in $\mathcal{M}(\mu_X, \mu_Y)$ and μ any coupling between μ_X and μ_Y. Then,

$$
\mathbf{J}_q(\mu_r) \overset{(A)}{\leq} \mathbf{J}_r(\mu_r) \overset{(B)}{\leq} \mathbf{J}_r(\mu),
$$

where (A) holds because by Lemma 10.2, $\{J_\ell(\nu)\}_\ell$ is non-decreasing in ℓ for fixed $\nu \in \mathcal{M}(\mu_X, \mu_Y)$ and (B) holds because μ_r minimizes $\mathbf{J}_r$ and μ was arbitrary. Then we have

$$\mathbf{J}_q(\mu_r) \leq \mathbf{J}_r(\mu).$$

Since $\mathcal{M}(\mu_X, \mu_Y)$ is compact [111, p. 49], it follows that $\{\mu_r\}_r$ has a (weakly) converging subsequence that we still denote by $\{\mu_r\}_r$. Let μ_* denote the limit of $\{\mu_r\}_r$. Letting $r \uparrow \infty$ in the inequality above, we find (via Lemma 10.2) that $\mathbf{J}_q(\mu_*) \leq \mathbf{J}_\infty(\mu)$. Now, let $q \uparrow \infty$ and again by Lemma 10.2, we find

$$\mathbf{J}_\infty(\mu_*) \leq \mathbf{J}_\infty(\mu).$$

Since μ was arbitrary, μ_* is a minimizer of $\mathbf{J}_\infty$ and the claim follows. □

We will need the following lemma in the proof of Theorem 5.1.

Lemma 10.4 *Let X, Y be two compact metric spaces and $R \in \mathcal{R}(X, Y)$ s.t. $\|\Gamma_{X,Y}\|_{L^\infty(R\times R)} = 0$. Then, there exists an* isometry *$\Phi : X \to Y$ s.t. $R = \{(x, \Phi(x)), x \in X\}$.*

Proof Note that by hypothesis $d_X(x, x') = d_Y(y, y')$ for all $(x, y), (x', y') \in R$.

Fix $x \in X$ and let $y, y' \in Y$ be s.t. $(x, y), (x, y') \in R$. Then, $0 = d_X(x, x) = d_Y(y, y')$, hence $y = y'$. We then see that for all $x \in X$ there exists a unique $y \in Y$ s.t. $(x, y) \in R$. Similarly, we see that for all $y \in Y$ there exists a unique $x \in X$ s.t. $(x, y) \in R$. Hence, R is a left-total, right-total, functional, and injective binary relation and therefore, it is a bijection, [64, Ch. 1]. Let $\Phi : X \to Y$ bijective be s.t. $R = \{(x, \Phi(x)), x \in X\}$. Then, Φ is distance preserving and surjective, and therefore an isometry. □

Proof of Theorem 5.1

(a) *Symmetry* is obvious. We need to prove the triangle inequality plus the fact that $\mathfrak{D}_p(X, Y) = 0$ happens if and only if X and Y are isomorphic. The "if" part is trivial. For the other direction we proceed as follows. Note that by virtue of Corollary 10.1, $\mathfrak{D}_p(X, Y) = 0$ implies that there exist $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ such that

$$\int_{X\times Y}\int_{X\times Y} \left|d_X\left(x, x'\right) - d_Y\left(y, y'\right)\right|^p \mu(dx \times dy)\, \mu\left(dx' \times dy'\right) = 0.$$

Then, $d_X(x, x') = d_Y(y, y')$ for all $(x, y), (x', y') \in R(\mu)$. By Lemma 10.4, this forces $R(\mu)$ to describe an isometry between $X = \mathrm{supp}[\mu_X]$ and $Y = \mathrm{supp}[\mu_Y]$. Then, there exists an isometry $i : X \to Y$ such that $R(\mu) = \{(x, i(x)), x \in X\} = \{(i^{-1}(y), y), y \in Y\}$. Hence, for all measurable $A \subset X$, $\mu_X(A) = \mu_Y(i(A))$. In fact, one has the following two chains of equalities: $\mu_X(A) = \mu(A \times Y) = \mu(A \times Y \cap R(\mu)) = \mu(A \times i(A))$ and $\mu_Y(i(A)) = \mu(X \times i(A)) = \mu(X \times i(A) \cap R(\mu)) = \mu(A \times i(A))$. Hence, we conclude that actually X and Y are isomorphic.

For the *triangle inequality* fix $X, Y, Z \in \mathcal{G}_w$. Notice first that for all $x, x' \in X$, $y, y' \in Y$ and $z, z' \in Z$,

$$\Gamma_{X,Y}\big(x, y, x', y'\big) \leq \Gamma_{X,Z}\big(x, x', z, z'\big) + \Gamma_{Y,Z}\big(y, y', z, z'\big). \tag{10.1}$$

Let $\mu_1 \in \mathcal{M}(\mu_X, \mu_Z)$ and $\mu_2 \in \mathcal{M}(\mu_Z, \mu_Y)$ be s.t. $\mathbf{J}_p(\mu_1) = \mathfrak{D}_p(X, Z)$ and $\mathbf{J}_p(\mu_2) = \mathfrak{D}_p(Z, Y)$. This is possible by Corollary 10.1. By the Glueing Lemma [111, Lemma 7.6], there exist a probability measure $\mu \in \mathcal{P}(X \times Y \times Z)$ with marginals μ_1 on $X \times Z$ and μ_2 on $Z \times Y$. Let μ_3 be the marginal of μ on $X \times Y$. Using the fact that μ has marginal $\mu_Z \in \mathcal{P}(Z)$ on Z and the triangle inequality for the L^p norm (i.e., Minkowski's inequality) and (10.1), one obtains

$$\begin{aligned} 2\mathfrak{D}_p(X, Y) &\leq \|\Gamma_{X,Y}\|_{L^p(\mu_3\otimes\mu_3)} \\ &= \|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)} \\ &\leq \|\Gamma_{X,Z} + \Gamma_{Z,Y}\|_{L^p(\mu\otimes\mu)} \\ &\leq \|\Gamma_{X,Z}\|_{L^p(\mu\otimes\mu)} + \|\Gamma_{Z,Y}\|_{L^p(\mu\otimes\mu)} \\ &= \|\Gamma_{X,Z}\|_{L^p(\mu_1\otimes\mu_1)} + \|\Gamma_{Z,Y}\|_{L^p(\mu_2\otimes\mu_2)} \\ &= 2\mathfrak{D}_p(X, Z) + 2\mathfrak{D}_p(Z, Y) \end{aligned}$$

thus establishing the triangle inequality for $\mathfrak{D}_p$ for finite $p \geq 1$. The proof for the case $p = \infty$ follows the same steps as the proof of the triangle inequality for the Gromov–Hausdorff distance and we omit it (see [22, Proposition 7.3.16]).

(b) Pick $\mu \in \mathcal{M}(\mu_X, \mu_Y)$. Let $R(\mu) = \mathrm{supp}[\mu]$, then by Lemma 2.2, $R(\mu) \in \mathcal{R}(X, Y)$. Note first that according to (5.8), $\mathbf{J}_\infty(\mu) = \|\Gamma_{X,Y}\|_{L^\infty(R(\mu)\times R(\mu))}$. Since $\{R(\mu),\ \mu \in \mathcal{M}(\mu_X, \mu_Y)\} \subset \mathcal{R}(X, Y)$, invoking expression (4.5) for the Gromov–Hausdorff distance, one obtains

$$d_{\mathcal{GH}}(X, Y) = \frac{1}{2} \inf_{R\in\mathcal{R}(X,Y)} \|\Gamma_{X,Y}\|_{L^\infty(R\times R)} \leq \mathbf{J}_\infty(\mu).$$

The conclusion follows upon recalling Definition 5.7 and noting that $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ is arbitrary.

(c) Let $\varepsilon > 0$ and $\mu \in \mathcal{M}(\alpha, \beta)$ be s.t. $\|d\|_{L^p(\mu)} < \varepsilon$. Then, since $\Gamma_{X,Y}(x, y, x', y') = |d(x, y) - d(x', y')| \leq d(x, y) + d(x', y')$, from the triangle inequality for the norm $\|\cdot\|_{L^p(\mu\otimes\mu)}$ it follows that

$$\|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)} \leq 2\|d\|_{L^p(\mu)} < 2\varepsilon.$$

The claim follows since $\varepsilon > d^Z_{\mathcal{W},p}(\alpha, \beta)$ was arbitrary.

(d) Consider the coupling $\mu = (\mathrm{Id} \times \mathrm{Id})_\# \alpha$. Then,

$$\begin{aligned} &\int_{Z\times Z\times Z\times Z} \big|d\big(x, x'\big) - d'\big(y, y'\big)\big|^p \mu(dx \times dy)\, \mu\big(dx' \times dy'\big) \\ &\quad = \int_{Z\times Z} \big|d(x, y) - d'(x, y)\big|^p \alpha(dx)\alpha(dy) \end{aligned}$$

and hence, $\|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)} = \|d - d'\|_{L^p(\alpha\otimes\alpha)}$ and the claim follows.

(e) This follows from (g) and a similar statement for $\mathfrak{S}_p$ [105, Lemma 3.5].

(f) When $Y = (\{z\}, 0, \delta_z^Z)$, $\Gamma_{X,Y}(x, z, x', z) = d_X(x, x')$ and according to Example 2.7, $\mathcal{M}(\mu_X, \mu_Y) = \{\mu_X\}$. Hence, $\|\Gamma_{X,Y}\|_{L^p(\mu_X \otimes \mu_X)} = \mathrm{diam}_p(X)$. The conclusion now follows.

(g) Assume first that $p \in [1, \infty)$. Let $\varepsilon > 0$, $d \in \mathcal{D}(d_X, d_Y)$ and $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ be s.t. $\|d\|_{L^p(\mu)} < \varepsilon$. Then, since $|d_X(x, x') - d_Y(y, y')| \le d(x, y) + d(x', y')$ for all $x, x' \in X$, $y, y' \in Y$, by the same argument used in (c),

$$\|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)} \le 2\|d\|_{L^p(\mu)} < 2\varepsilon.$$

The conclusion follows since $\varepsilon > \mathfrak{S}_p(X, Y)$ is arbitrary. The same reasoning applies to the case $p = \infty$. Indeed, let $\varepsilon > 0$, $d \in \mathcal{D}(d_X, d_Y)$ and $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ be s.t. $d(x, y) < \varepsilon$ for all $(x, y) \in R(\mu)$. Then $|d_X(x, x') - d_Y(y, y')| < 2\varepsilon$ for all (x, y), $(x', y') \in R(\mu)$ and thus $\mathfrak{D}_\infty(X, Y) < \varepsilon$.
For the reverse inequality, let $\varepsilon > 0$ and $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ be s.t. $\|\Gamma_{X,Y}\|_{L^\infty(\mu\otimes\mu)} < \varepsilon$. Then $|d_X(x, x') - d_Y(y, y')| < 2\varepsilon$ for all (x, y), $(x', y') \in R(\mu)$. Let $d \in \mathcal{D}(d_X, d_Y)$ be given by

$$d(x, y) := \inf_{(x', y') \in R(\mu)} \bigl(d_X\bigl(x, x'\bigr) + d_Y\bigl(y, y'\bigr)\bigr) + \varepsilon.$$

This construction is justified by Lemma 10.1. Obviously, $\|d\|_{L^\infty(R(\mu))} < \varepsilon$ and hence $\mathfrak{S}_\infty(X, Y) < \varepsilon$.

(h) The claim for $\mathfrak{D}_p$ follows from Lemma 10.2. The claim for $\mathfrak{S}_p$ follows from the observation that for fixed $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ and $d \in \mathcal{D}(d_X, d_Y)$, $p \mapsto \|d\|_{L^p(\mu)}$ is non-decreasing.

(i) Notice that since for all $x, x' \in X$ and $y, y' \in Y$, $\Gamma_{X,Y}(x, x', y, y') \le 2M$; then

$$\bigl(\Gamma_{X,Y}\bigl(x, y, x', y'\bigr)\bigr)^p \le (2M)^{p-q}\bigl(\Gamma_{X,Y}\bigl(x, y, x', y'\bigr)\bigr)^q.$$

Now, pick any $\mu \in \mathcal{M}(\mu_X, \mu_Y)$. Using the definition of $\mathfrak{D}_p$ and the above we find

$$\mathfrak{D}_p(X, Y) \le \frac{1}{2}\|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)} \le \frac{1}{2}(2M)^{1-q/p}\bigl(\|\Gamma_{X,Y}\|_{L^q(\mu\otimes\mu)}\bigr)^{q/p}.$$

But the right-hand side above equals $M^{1-q/p}(\frac{1}{2}\|\Gamma_{X,Y}\|_{L^q(\mu\otimes\mu)})^{q/p}$, and the proof is finished since μ is arbitrary.

□

10.2 Other Proofs

Lemma 10.5 [50] *Let $\delta > 0$, $\varepsilon \ge 0$ and $(X, d_X, \mu_X) \in \mathcal{G}_w$. If $v_\delta(X) \le \varepsilon$, then there exists $N < [1/\delta]$ and points $x_1, \ldots, x_N \in X$ such that the following hold*:

- $\mu_X(B_X(x_i, \varepsilon)) > \delta$ *for all* $i = 1, 2, \ldots, N$.
- $\mu_X(\bigcup_{i=1}^N B_X(x_i, 2\varepsilon)) > 1 - \varepsilon$.
- *For all* $i, j = 1, 2, \ldots, N$ *with* $i \ne j$, $d_X(x_i, x_j) > \varepsilon$.

Proof of Proposition 5.3 We adapt an argument of [50]. Let $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ be a coupling s.t. $\|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)} < 2\delta^5$. Set $\varepsilon = 4v_\delta(X)$.

Claim 10.1 *There exist points $x_1, \dots, x_N$ in X with $N \le [1/\delta]$, $\min_{i \neq j} d_X(x_i, x_j) \ge \varepsilon/2$, $\min_i \mu_X(B_X(x_i, \varepsilon)) > \delta$ and $\mu_X(\bigcup_{i=1}^N B_X(x_i, \varepsilon)) \ge 1 - \varepsilon$.*

Proof Apply Lemma 10.5 with $\varepsilon' = \varepsilon/2 = 2v_\delta(X) > v_\delta(X)$ and obtain points $x_1, \dots, x_N$ with $N \le [1/\delta]$, $\min_{i\neq j} d_X(x_i, x_j) > \varepsilon'$, $\mu_X(B_X(x_i, \varepsilon')) > \delta$ and $\mu_X(\bigcup_{i=1}^N B_X(x_i, 2\varepsilon')) \ge 1 - \varepsilon'$. But then

$$\mu_X\big(B_X(x_i, \varepsilon)\big) = \mu_X\big(B_X\big(x_i, 2\varepsilon'\big)\big) \ge \mu_X\big(B_X\big(x_i, \varepsilon'\big)\big) > \delta.$$

Also,

$$\mu_X\left(\bigcup_{i=1}^N B_X(x_i, \varepsilon)\right) = \mu_X\left(\bigcup_{i=1}^N B_X\big(x_i, 2\varepsilon'\big)\right) \ge 1 - \varepsilon' = 1 - \varepsilon/2 \ge 1 - \varepsilon;$$

whence the conclusion. □

Claim 10.2 *We claim that for every $i = 1, \dots, N$ there exists $y_i \in Y$ s.t.*

$$\mu\big(B_X(x_i, \varepsilon) \times B_Y\big(y_i, 2(\varepsilon + \delta)\big)\big) \ge \big(1 - \delta^2\big)\mu_X\big(B_X(x_i, \varepsilon)\big). \tag{10.2}$$

Proof Assume the assertion is false for some i and for each $y \in Y$ define $Q_i(y) = B_X(x_i, \varepsilon) \times (Y \backslash B_Y(y, 2(\varepsilon + \delta)))$. Then, since $\mu \in \mathcal{M}(\mu_X, \mu_Y)$,

$$\begin{aligned}\mu_X\big(B_X(x_i, \varepsilon)\big) &= \mu\big(B_X(x_i, \varepsilon) \times Y\big)\\ &= \mu\big(B_X(x_i, \varepsilon) \times B_Y\big(y, 2(\varepsilon + \delta)\big)\big)\\ &\quad + \mu\big(B_X(x_i, \varepsilon) \times Y \backslash B_Y\big(y, 2(\varepsilon + \delta)\big)\big),\end{aligned}$$

hence $\mu(Q_i(y)) \ge \delta^2 \mu_X(B_X(x_i, \varepsilon))$. Define as well

$$X \times Y \times X \times Y \supset \mathcal{Q}_i = \big\{\big(x, x', y, y'\big) | \, x, x' \in B_X(x_i, \varepsilon) \text{ and } d_Y\big(y, y'\big) \ge 2(\varepsilon + \delta)\big\}.$$

Notice that for $(x, y, x', y') \in \mathcal{Q}_i$, $\Gamma_{X,Y}(x, y, x', y') \ge 2\delta$. Furthermore, $\mu \otimes \mu(\mathcal{Q}_i) \ge \delta^4$. Indeed,

$$\begin{aligned}\mu \otimes \mu(\mathcal{Q}_i) &= \iint_{B_X(x_i,\varepsilon)\times Y} \iint_{Q_i(y)} 1 \, \mu\big(dx' \times dy'\big) \, \mu(dx \times dy)\\ &= \iint_{B_X(x_i,\varepsilon)\times Y} \mu\big(Q_i(y)\big) \, \mu(dx \times dy)\\ &= \mu_X\big(B_X(x_i, \varepsilon)\big) \int_Y \mu\big(Q_i(y)\big) \, \mu_Y(dy)\\ &\ge \big(\mu_X\big(B_X(x_i, \varepsilon)\big)\big)^2 \delta^2\\ &\ge \delta^4.\end{aligned}$$

But then

$$\|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)} \geq \|\Gamma_{X,Y}\|_{L^1(\mu\otimes\mu)} \geq \|\Gamma_{X,Y}\mathbb{1}_{Q_i}\|_{L^1(\mu\otimes\mu)} \geq 2\delta\cdot\mu\otimes\mu(Q_i) \geq 2\delta^5,$$

a contradiction. □

If for each k we define

$$S_k = B_X(x_k,\varepsilon)\times B_Y\big(y_k, 2(\varepsilon+\delta)\big),$$

then by (10.2) one has

$$\mu(S_k) \geq \delta\big(1-\delta^2\big), \quad \text{for all } k = 1,\dots,N. \tag{10.3}$$

Claim 10.3 $\Gamma_{X,Y}(x_i,y_i,x_j,y_j) \leq 6(\varepsilon+\delta)$ *for all* $i,j = 1,\dots,N$.

Proof Assume that this condition fails for some (i_0,j_0); then for $(x,y)\in S_{i_0}$ and $(x',y')\in S_{j_0}$, $\Gamma_{X,Y}(x,y,x',y')\geq 2\delta$. This follows from the inequality

$$\begin{aligned}\Gamma_{X,Y}\big(x,y,x',y'\big) &\geq \Gamma_{X,Y}(x_{i_0},y_{i_0},x_{j_0},y_{j_0}) - \big|d_X(x_{i_0},x_{j_0}) - d_X\big(x,x'\big)\big| \\ &\quad - \big|d_Y(y_{i_0},y_{j_0}) - d_Y\big(y,y'\big)\big| \\ &\geq 6(\varepsilon+\delta) - 2\varepsilon - 4(\varepsilon+\delta) \\ &= 2\delta.\end{aligned}$$

Now,

$$\begin{aligned}\|\Gamma_{X,Y}\|_{L^p(\mu\otimes\mu)} &\geq \|\Gamma_{X,Y}\|_{L^1(\mu\otimes\mu)} \geq \|\Gamma_{X,Y}\mathbb{1}_{S_{i_0}}\mathbb{1}_{S_{j_0}}\|_{L^1(\mu\otimes\mu)} \geq 2\delta\cdot\mu(S_{i_0})\mu(S_{j_0}) \\ &\geq 2\delta\big(\delta\big(1-\delta^2\big)\big)^2.\end{aligned}$$

But since $\delta\leq 1/2$, the right-hand side is bounded below by $2\delta^5$, which leads to a contradiction. □

Finally, consider $S\subset X\times Y$ given by $S = \{(x_i,y_i),\ i = 1,2,\dots,N\}$ and let d_S be the metric on $X\sqcup Y$ given by Lemma 10.1. Clearly, $d_S(x_i,y_i)\leq 3(\varepsilon+\delta)$ for $i = 1,\dots,N$. Furthermore, one has the simple bound

$$d_S(x,y) \leq \operatorname{diam}(X) + \operatorname{diam}(Y) + 3(\varepsilon+\delta) \leq \operatorname{diam}(X) + \operatorname{diam}(Y) + 15 =: M'.$$

Claim 10.4 *Fix* $i\in\{1,\dots,N\}$. *Then, for all* $(x,y)\in S_i$, $d_S(x,y)\leq 6(\varepsilon+\delta)$.

Proof Assume $(x,y)\in S_i$. Then, $d_X(x,x_i) < \varepsilon$ and $d_Y(y,y_i) < 2(\varepsilon+\delta)$. Then, by the triangle inequality for d_S:

$$\begin{aligned}d_S(x,y) &\leq d_X(x,x_i) + d_Y(y,y_i) + d_S(x_i,y_i) \\ &\leq \varepsilon + 2(\varepsilon+\delta) + 3(\varepsilon+\delta) \leq 6(\varepsilon+\delta).\end{aligned}$$ □

Let $L := \bigcup_{i=1}^{N} S_i$. Then, $\{(x, y) \in X \times Y \,|\, d_S(x, y) < 6(\varepsilon+\delta)\} \subseteq L$. The next step is to estimate the mass in the complement of L.

Claim 10.5 *One has* $\mu(X \times Y \backslash L) \leq \varepsilon + \delta$.

Proof For each i let $A_i := B_X(x_i, \varepsilon) \times (Y \backslash B_Y(y_i, 2(\varepsilon+\delta)))$. Then,

$$A_i = \big(B_X(x_i, \varepsilon) \times Y\big) \backslash B_X(x_i, \varepsilon) \times B_Y\big(y_i, 2(\varepsilon+\delta)\big) = \big(B_X(x_i, \varepsilon) \times Y\big) \backslash S_i.$$

Hence,

$$\mu(A_i) = \mu\big(B_X(x_i, \varepsilon) \times Y\big) - \mu(S_i) = \mu_X\big(B_X(x_i, \varepsilon)\big) - \mu(S_i),$$

where the last equality follows form the fact that $\mu \in \mathcal{M}(\mu_X, \mu_Y)$. Now, since by construction $\mu(S_i) \geq \mu_X(B_X(x_i, \varepsilon))(1-\delta^2)$, one finds

$$\mu(A_i) \leq \mu_X\big(B_X(x_i, \varepsilon)\big) \cdot \delta^2.$$

Notice that one can write

$$X \times Y \backslash L \subseteq \left(X \Big\backslash \bigcup_{i=1}^{N} B_X(x_i, \varepsilon)\right) \times Y \cup \left(\bigcup_{i=1} A_i\right),$$

and hence, using for the first term that $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ and using the union bound for the second, one finds

$$\mu(X \times Y \backslash L) \leq \mu_X\left(X \Big\backslash \bigcup_{i=1}^{N} B_X(x_i, \varepsilon)\right) + \sum_{i=1}^{N} \mu(A_i).$$

This concludes the proof of the claim since by construction the fist term above is bounded above by ε whereas the second term, by previous computations, is bounded by $N \cdot \delta^2 \leq \delta$. □

Now write

$$\begin{aligned}\iint_{X\times Y} d_S^p(x, y)\mu(dx \times dy) &= \iint_L d_S^p(x, y)\mu(dx \times dy) \\ &\quad + \iint_{X\times Y\backslash L} d_S^p(x, y)\mu(dx \times dy) \\ &\leq \big(6(\varepsilon+\delta)\big)^p + M'^p \cdot (\varepsilon+\delta).\end{aligned}$$

Invoking the inequality $a^{1/p} + b^{1/p} \geq (a+b)^{1/p}$ valid for all $p \geq 1$ and $a, b \geq 0$ one finds

$$\mathfrak{S}_p(X, Y) \leq (\varepsilon+\delta)^{1/p}\big(6(\varepsilon+\delta)^{1-1/p} + M'\big),$$

from which the result follows, since $\varepsilon = 4v_\delta(X) \leq 4$ and $\delta \leq 1$. □

Proof of Proposition 5.2 For $\alpha \in [0, \pi/\sqrt{K}]$ let $V_{K,n}(\alpha) := c_{n-1}\int_0^\alpha (\frac{1}{\sqrt{K}} \times \sin(\sqrt{K}t))^{n-1}\, dt$, where $c_{n-1} = n\pi^{n/2}\Gamma(n/2+1)$. Notice that for any $X \in \mathcal{F}(n, K, V, I, D)$, by the Bishop–Günther theorem [26, Theorem III-4.2], for $x \in X$

$$\mu_X\big(B_X(x,\varepsilon)\big) \geq \frac{V_{K,n}(\varepsilon)}{V} =: \rho_{\mathcal{F}}(\varepsilon)$$

for all $0 \leq \varepsilon \leq \min(I, \pi/\sqrt{K})$. Hence, the conditions of Remark 5.20 hold. Also, note that since $V_{K,n}$ is increasing and continuous,

$$\sup_{X\in\mathcal{F}} v_\delta(X) \leq \inf\big\{\varepsilon > 0 | \rho_{\mathcal{F}}(\varepsilon) > \delta\big\} = V_{K,n}^{-1}(V\delta) \text{ for } \delta \leq \tfrac{V_{K,n}(\min(I,\pi/\sqrt{K}))}{V}. \quad (10.4)$$

Notice that for all $\alpha \in [0, \sqrt{K}\pi/2]$

$$V_{K,n}(\alpha) \geq c_{n-1}\int_0^\alpha \frac{1}{K^{(n-1)/2}}\left(\frac{2t}{\pi\sqrt{K}}\right)^{n-1} dt = q_n(K)\alpha^n,$$

where $q_n(K) := \frac{c_{n-1}}{n}(\frac{\pi K}{2})^{n-1}$. Hence,

$$V_{K,n}^{-1}(V\delta) \leq \big(V\delta/q_n(K)\big)^{1/n} \quad \text{as long as} \quad \delta \leq 2/V. \quad (10.5)$$

Define

$$\delta_{\mathcal{F}} := \left(\frac{1}{V}\min\big(1/2, V_{K,n}\big(\min(I,\pi/\sqrt{K})\big)\big)\right)^5$$

and assume $X, Y \in \mathcal{F}$ are such that $\mathfrak{D}_p(X,Y) \leq \delta^5 < \delta_{\mathcal{F}}$. Then by Proposition 5.3,

$$\begin{aligned}
\mathfrak{S}_p(X,Y) &\leq \Big(4 \cdot \sup_{Z\in\mathcal{F}} v_\delta(Z) + \delta\Big)^{1/p} M \\
&\leq \big(4 \cdot V_{K,n}^{-1}(V\delta) + \delta\big)^{1/p} M \quad \text{(by (10.4))} \\
&\leq \left(\frac{4}{q_n(K)}\delta^{1/n} + \delta\right)^{1/p} M \quad \text{(by (10.5))} \\
&\leq \big(4/q_n(K) + \delta^{\frac{n-1}{n}}\big)^{1/p} M \cdot \delta^{1/np} \\
&\leq \big(4/q_n(K) + 1\big)^{1/p} M \cdot \delta^{1/np}.
\end{aligned}$$

Now $M \leq 2D + 45$. Let $C_{\mathcal{F}} := (4/q_n(K)+1)^{1/p}(2D+45)$. One concludes from the above since $\delta \geq (\mathfrak{D}_p(X,Y))^{1/5}$ was arbitrary. □

Proof of Theorem 5.3 Recall the notion of the Gromov–Prokhorov distance introduced in [50]. For $X, Y \in \mathcal{G}_w$ define

$$d_{\mathcal{GP}}(X,Y) := \inf\Big\{\varepsilon > 0 \,|\, \exists \mu \in \mathcal{M}(\mu_X, \mu_Y), d \in \mathcal{D}(d_X, d_Y) \\ \text{s.t. } \mu\big\{(x,y)|d(x,y) \geq \varepsilon\big\} \leq \varepsilon\Big\}.$$

Note that $d_{\mathcal{GP}}(X,Y) \leq 1$ for all $X, Y \in \mathcal{G}_w$.

Claim 10.6 *For* $p \geq 1$,

$$\mathfrak{D}_p(X,Y) \leq \big(d_{\mathcal{GP}}(X,Y)\big)^{1/p}\big(D^p + 1\big)^{1/p}$$

for all $X, Y \in \mathcal{G}_w$ *s.t.* $\max(\mathrm{diam}(X), \mathrm{diam}(Y)) \leq D$.

Proof Pick $\varepsilon > 0$ s.t. $d_{\mathcal{GP}}(X,Y) < \varepsilon$. Let $d \in \mathcal{D}(d_X, d_Y)$ and $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ be s.t.

$$\mu\big\{(x,y)|\, d(x,y) \geq \varepsilon\big\} \leq \varepsilon.$$

Then,

$$\begin{aligned}
\big(\mathfrak{S}_p(X,Y)\big)^p &\leq \iint_{X\times Y} d^p(x,y)\mu(dx\times dy) \\
&= \iint_{\{(x,y)|d(x,y)\geq\varepsilon\}} d^p(x,y)\mu(dx\times dy) \\
&\quad + \iint_{\{(x,y)|d(x,y)<\varepsilon\}} d^p(x,y)\mu(dx\times dy) \\
&\leq D^p\varepsilon + \varepsilon^p.
\end{aligned}$$

The conclusion follows since $\varepsilon > d_{\mathcal{GP}}(X,Y)$ was arbitrary and by Theorem 5.1, $\mathfrak{D}_p \leq \mathfrak{S}_p$. □

Condition (1) of the theorem trivially implies that the collection $\{df_X,\ X \in \mathcal{F}\}$ is tight. On the other hand, condition (2), via Remark 5.20, implies that

$$\lim_{\delta\downarrow 0} \sup_{X\in\mathcal{F}} v_\delta(X) = 0.$$

Then, by Proposition 7.1 of [50] the family $\mathcal{F}$ is pre-compact in the Gromov–Prokhorov topology.

Let $\{X_n\}_n \subset \mathcal{F}$ be any sequence in the family $\mathcal{F}$. Then, there exist $\{n_k\}_k \subset \mathbb{N}$ and $X \in \mathcal{G}_w$ s.t. $\lim_k d_{\mathcal{GP}}(X_{n_k}, X) = 0$. Finally, since by hypothesis $\max(\mathrm{diam}(X), \sup_k \mathrm{diam}(X_{n_k})) \leq D$, by Claim 10.6,

$$\mathfrak{D}_p(X_{n_k}, X) \to 0 \quad \text{as } k \uparrow \infty.$$

One concludes that $\mathcal{F}$ is pre-compact in the $\mathfrak{D}_p$-topology as well. □

11 Notations

Symbol	**Meaning**
$\mathcal{L}$	One-dimensional Lebesgue measure
$B_X(x,r)$	An open ball in the metric space X, centered at x and with radius r
(X,d_X)	Metric space X with metric d_X
$\mathcal{B}(X)$	Borel σ-algebra of the topological space X
$\mathcal{P}(X)$	Collection of all Borel probability measures on the metric space X
δ_z^Z	Delta measure on Z concentrated at point z
$f_\# \nu$	Push-forward measure of ν through the map f
dF	Probability measure on $\mathbb{R}$ associated to the distribution F
(X,d_X,μ_X)	Compact metric measure space X with metric d_X and Borel probability measure μ_X
$\mathrm{supp}[\mu]$	Support of the probability measure μ, Sect. 2.2
$\mathcal{G}$	Collection of all compact metric spaces
$\mathcal{G}_w$	Collection of all metric measure spaces
Riem	Collection of all Riemannian manifolds without boundary
$\mathcal{C}(Z)$	Collection of all compact subsets of the metric space Z
$\mathcal{C}_w(Z)$	Collection of all compactly supported probability measures on Z
$d_{\mathcal{H}}^Z$	Hausdorff distance between subsets of the metric space Z, Sect. 2.1
$d_{\mathcal{W},p}^Z$	p-Wasserstein distance between (Borel) probability measures on the metric space Z, Sect. 2.7
$d_{\mathcal{GH}}$	Gromov–Hausdorff distance between metric spaces, Sect. 4.1
$\mathfrak{S}_p$ and $\mathfrak{D}_p$	Gromov–Wasserstein distances between metric measure spaces, Sect. 5.10
$\mathrm{diam}(X)$	Diameter of the metric space X, Sect. 4
$\mathrm{diam}_p(X)$	p-Diameter of the metric measure space X, Sect. 5.2
Π_n	All $n!$ permutations of elements of the set $\{1,\dots,n\}$
$\mathbb{P}(\mathcal{E})$	Probability of the event $\mathcal{E}$
$\mathbb{E}(x)$	Expected value of the random variable x
$\mathbb{E}_\nu(\cdot)$, $\mathbb{P}_\nu(\cdot)$	Expected and probability value according to the law ν
$\mathbb{1}_A : Z \to \mathbb{R}$	Indicator function of $A \subset Z$: $\mathbb{1}_A(z)=1$ if $z \in A$ and $\mathbb{1}_A(z)=0$ otherwise
$\mathcal{R}(A,B)$	A correspondence: a subset of $A \times B$, satisfying certain properties, Sect. 2.1
$\mathcal{M}(\mu_A,\mu_B)$	Set of measure couplings: measures on $A \times B$, satisfying certain properties, Sect. 2.3
$\mathcal{D}(d_X,d_Y)$	Set of metric couplings: metrics on $X \times Y$, satisfying certain properties, Sect. 4.1
$\mathcal{M}_1^+(d)$	Collection of all $(v_1,\dots,v_d)$ in $\mathbb{R}^d$ s.t. $v_i \geq 0$, $i=1,\dots,d$, and $\sum_i v_i = 1$
δ_{ij}	Kronecker delta: for $i,j \in \mathbb{N}$, $\delta_{i,j}=1$ whenever $i=j$ and $\delta_{i,j}=0$ otherwise
A^T	Transposed matrix of A

Acknowledgements The author gratefully acknowledges Gunnar Carlsson for his patient mentorship, support, and for providing an environment where this research was possible. Thanks to Leonidas Guibas, Robert Gulliver, Tigran Ishkhanov, Guillermo Sapiro, and Ofer Zeitouni.

This work was supported by DARPA grant number HR0011-05-1-0007 and ONR grant number N00014-09-1-0783.

References

1. P. Alestalo, D.A. Trotsenko, J. Väisälä, Isometric approximation, *Isr. J. Math.* **125**, 61–82 (2001).
2. L. Ambrosio, P. Tilli, *Topics on Analysis in Metric Spaces*. Oxford Lecture Series in Mathematics and Its Applications, vol. 25 (Oxford University Press, Oxford, 2004).
3. M. Ankerst, G. Kastenmüller, H.-P. Kriegel, T. Seidl, 3D shape histograms for similarity search and classification in spatial databases, in *SSD '99: Proceedings of the 6th International Symposium on Advances in Spatial Databases*, London, UK (Springer, Berlin, 1999), pp. 207–226.
4. S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(4), 509–522 (2002).
5. S. Berchtold, *Geometry-based search of similar parts*. Ph.D. thesis, University of Munich, Germany (1998).
6. P. Billingsley, *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics, 3rd edn. (Wiley, New York, 1995).
7. G.S. Bloom, A counterexample to a theorem of S. Piccard, *J. Comb. Theory, Ser. A* **22**(3), 378–379 (1977).
8. F.L. Bookstein, The study of shape transformation after D'Arcy Thompson, *Math. Biosci.* **34**, 177–219 (1977).
9. F.L. Bookstein, *The Measurement of Biological Shape and Shape Change* (Springer, Berlin, 1978).
10. F.L. Bookstein, *Morphometric Tools for Landmark Data* (Cambridge University Press, Cambridge, 1997). Geometry and biology, reprint of the 1991 original.
11. M. Boutin, G. Kemper, On reconstructing n-point configurations from the distribution of distances or areas, *Adv. Appl. Math.* **32**(4), 709–735 (2004).
12. D. Brinkman, P.J. Olver, Invariant histograms. *Am. Math. Monthly* (2011). doi:10.1007/s11263-009-0301-6
13. A. Bronstein, M. Bronstein, R. Kimmel, Topology-invariant similarity of nonrigid shapes, *Int. J. Comput. Vis.* **81**(3), 281–301 (2009).
14. A. Bronstein, M. Bronstein, R. Kimmel, M. Mahmoudi, G. Sapiro, A Gromov–Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching (submitted).
15. A. Bronstein, M. Bronstein, A. Bruckstein, R. Kimmel, Partial similarity of objects, or how to compare a centaur to a horse, *Int. J. Comput. Vis.* **84**(2), 163–183 (2009).
16. A. Bronstein, M. Bronstein, R. Kimmel, Three-dimensional face recognition, *Int. J. Comput. Vis.* **64**(1), 5–30 (2005).
17. A. Bronstein, M. Bronstein, R. Kimmel, Efficient computation of isometry-invariant distances between surfaces, *SIAM J. Sci. Comput.* **28**(5), 1812–1836 (2006).
18. A.M. Bronstein, M.M. Bronstein, R. Kimmel, Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching, *Proc. Natl. Acad. Sci. USA* **103**(5), 1168–1172 (2006).
19. A. Bronstein, M. Bronstein, R. Kimmel, Calculus of nonrigid surfaces for geometry and texture manipulation, *IEEE Trans. Vis. Comput. Graph.* **13**(5), 902–913 (2007).
20. A. Bronstein, M. Bronstein, R. Kimmel, Expression-invariant representations of faces, *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(1), 1042–1053 (2007).
21. A. Bronstein, M. Bronstein, R. Kimmel, On isometric embedding of facial surfaces into S^3, in *Scale Space* (Springer, Berlin, 2004).
22. D. Burago, Y. Burago, S. Ivanov, *A Course in Metric Geometry*. AMS Graduate Studies in Math., vol. 33 (American Mathematical Society, Providence, 2001).
23. B. Bustos, D.A. Keim, D. Saupe, T. Schreck, D.V. Vranić, Feature-based similarity search in 3D object databases, *ACM Comput. Surv.* **37**(4), 345–387 (2005).
24. G. Carlsson, F. Mémoli, Characterization, stability and convergence of hierarchical clustering methods, *J. Mach. Learn. Res.* **11**(Apr), 1425–1470 (2010).

25. T.K. Carne, The geometry of shape spaces, *Proc. Lond. Math. Soc. (3)* **61**(2), 407–432 (1990).
26. I. Chavel, *Riemannian Geometry: A Modern Introduction* (Cambridge University Press, Cambridge, 1997).
27. F. Chazal, D. Cohen-Steiner, L. Guibas, F. Mémoli, S. Oudot, Gromov–Hausdorff stable signatures for shapes using persistence, in *Proc. of SGP* (2009).
28. G.E. Christensen, R.D. Rabbitt, M.I. Miller, 3D brain mapping using a deformable neuroanatomy, *Phys. Med. Biol.* **39**(3), 609–618 (1994).
29. G.E. Christensen, R.D. Rabbitt, M.I. Miller, Deformable templates using large deformation kinematics, *IEEE Trans. Image Process.* **5**(10), 1435–1447 (1996).
30. U. Clarenz, M. Rumpf, A. Telea, Robust feature detection and local classification for surfaces based on moment analysis, *IEEE Trans. Vis. Comput. Graph.* **10**(5), 516–524 (2004).
31. S.D. Cohen, L.J. Guibas, The earth mover's distance under transformation sets, in *ICCV (2)* (1999), pp. 1076–1083.
32. T.F. Cox, M.A.A. Cox, *Multidimensional Scaling*. Monographs on Statistics and Applied Probability, vol. 59 (Chapman & Hall, London, 1994).
33. M. d'Amico, P. Frosini, C. Landi, Natural pseudo-distance and optimal matching between reduced size functions. Technical report 66, DISMI, Univ. degli Studi di Modena e Reggio Emilia, Italy (2005).
34. M. d'Amico, P. Frosini, C. Landi, Using matching distance in size theory: a survey, *Int. J. Imaging Syst. Technol.* **16**(5), 154–161 (2006).
35. D.B. Rusch, A.L. Halpern, G. Sutton, K.B. Heidelberg, S. Williamson, et al., The sorcerer ii global ocean sampling expedition: Northwest Atlantic through Eastern Tropical Pacific, *PLoS Biol.* **5**(3) (2007) cover.
36. R.M. Dudley, *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics, vol. 74 (Cambridge University Press, Cambridge, 2002). Revised reprint of the 1989 original.
37. A. Elad (Elbaz), R. Kimmel, On bending invariant signatures for surfaces, *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(10), 1285–1295 (2003).
38. P. Frosini, A distance for similarity classes of submanifolds of Euclidean space, *Bull. Aust. Math. Soc.* **42**(3), 407–416 (1990).
39. P. Frosini, *Omotopie e invarianti metrici per sottovarieta di spazi euclidei (teoria della taglia)*. PhD thesis, University of Florence, Italy (1990).
40. P. Frosini, M. Mulazzani, Size homotopy groups for computation of natural size distances, *Bull. Belg. Math. Soc. Simon Stevin* **6**(3), 455–464 (1999).
41. W. Gangbo, R.J. McCann, Shape recognition via Wasserstein distance, *Q. Appl. Math.* **58**(4), 705–737 (2000).
42. N. Gelfand, N.J. Mitra, L. Guibas, H. Pottmann, Robust global registration, in *SGP '05: Proceedings of the Third Eurographics Symposium on Geometry Processing* (2005), pp. 197–206.
43. N. Giorgetti, Glpkmex: A matlab mex interface for the glpk library. http://www.dii.unisi.it/cohes/tools/mex/downloads/glpkmex/index.html.
44. C.R. Givens, R.M. Shortt, A class of Wasserstein metrics for probability distributions, *Mich. Math. J.* **31**(2), 231–240 (1984).
45. J. Glaunès, M. Vaillant, M.I. Miller, Landmark matching via large deformation diffeomorphisms on the sphere, *J. Math. Imaging Vis.* **20**(1–2), 179–200 (2004).
46. A. Gray, *Tubes*. Progress in Mathematics, vol. 221, 2nd edn. (Birkhäuser, Basel, 2004). With a preface by Vicente Miquel.
47. U. Grenander, *Pattern Synthesis.* Lectures in Pattern Theory, vol. 1. Applied Mathematical Sciences, vol. 18 (Springer, New York, 1976).
48. U. Grenander, *General pattern theory*. Oxford Mathematical Monographs (Clarendon Press, New York, 1993). A Mathematical Study of Regular Structures, Oxford Science Publications.
49. U. Grenander, M.I. Miller, Computational anatomy: an emerging discipline, *Q. Appl. Math.* **LVI**(4), 617–694 (1998).
50. A. Greven, P. Pfaffelhuber, A. Winter, Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees), *Probab. Theory Relat. Fields* **145**(1–2), 285–322 (2009).
51. C. Grigorescu, N. Petkov, Distance sets for shape filters and shape recognition, *IEEE Trans. Image Process.* **12**(10), 1274–1286 (2003).
52. M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*. Progress in Mathematics, vol. 152 (Birkhäuser, Boston, 1999).
53. A.B. Hamza, H. Krim, Geodesic object representation and recognition, in *Lecture Notes in Computer Science*, vol. 2886 (Springer, Berlin, 2003), pp. 378–387.

54. M. Hilaga, Y. Shinagawa, T. Kohmura, T.L. Kunii, Topology matching for fully automatic similarity estimation of 3D shapes, in *SIGGRAPH '01: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques* (ACM, New York, 2001), pp. 203–212.
55. J.R. Hoffman, R.P.S. Mahler, Multitarget miss distance via optimal assignment, *IEEE Trans. Syst. Man Cybern., Part A* **34**(3), 327–336 (2004).
56. L. Holm, C. Sander, Protein structure comparison by alignment of distance matrices, *J. Mol. Biol.* **233**(1), 123–138 (1993).
57. D.P. Huttenlocher, G.A. Klanderman, W.J. Rucklidge, Comparing images using the Hausdorff distance, *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(9), 850–863 (1993).
58. M. Jin, J. Kim, F. Luo, X. Gu, Discrete surface Ricci flow, *IEEE Trans. Vis. Comput. Graph.* **14**(5), 1030–1043 (2008).
59. M. Jin, W. Zeng, F. Luo, X. Gu, Computing Teichmüller shape space, *IEEE Trans. Vis. Comput. Graph.* **15**(3), 504–517 (2008).
60. A. Johnson, *Spin-images: a representation for 3D surface matching*. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, August 1997.
61. N.J. Kalton, M.I. Ostrovskii, Distances between Banach spaces, *Forum Math.* **11**(1), 17–48 (1999).
62. G. Kastenmüller, H.P. Kriegel, T. Seidl, Similarity search in 3D protein databases, in *Proc. GCB* (1998).
63. D.G. Kendall, Shape manifolds, procrustean metrics, and complex projective spaces, *Bull. Lond. Math. Soc.* **16**(2), 81–121 (1984).
64. M. Kilp, U. Knauer, A.V. Mikhalev, *Monoids, Acts and Categories*. de Gruyter Expositions in Mathematics, vol. 29 (Walter de Gruyter & Co., Berlin, 2000).
65. O. Klein, R.C. Veltkamp, Approximation algorithms for computing the earth mover's distance under transformations, in *ISAAC* (2005), pp. 1019–1028.
66. R. Kolodny, N. Linial, Approximate protein structural alignment in polynomial time, *Proc. Natl. Acad. Sci. USA* **101**, 12201–12206 (2004).
67. W.A. Koppensteiner, P. Lackner, M. Wiederstein, M.J. Sippl, Characterization of novel proteins based on known protein structures, *J. Mol. Biol.* **296**(4), 1139–1152 (2000).
68. H.L. Le, D.G. Kendall, The Riemannian structure of Euclidean shape spaces: a novel environment for statistics, *Ann. Stat.* **21**(3), 1225–1271 (1993).
69. M. Leordeanu, M. Hebert, A spectral technique for correspondence problems using pairwise constraints, in *International Conference of Computer Vision (ICCV)*, October, vol. 2 (2005), pp. 1482–1489.
70. H. Ling, D.W. Jacobs, Using the inner-distance for classification of articulated shapes, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2 (2005), pp. 719–726.
71. J. Löfberg, YALMIP: A toolbox for modeling and optimization in MATLAB, in *Proceedings of the CACSD Conference*, Taipei, Taiwan (2004).
72. L. Lovász, M.D. Plummer, *Matching Theory*. North-Holland Mathematics Studies, vol. 121, Annals of Discrete Mathematics, vol. 29 (North-Holland Publishing Co., Amsterdam, 1986).
73. D.G. Luenberger, *Linear and Nonlinear Programming*, 2nd edn. (Kluwer Academic, Boston, 2003).
74. S. Manay, D. Cremers, B.W. Hong, A.J. Yezzi, S. Soatto, Integral invariants for shape matching, *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(10), 1602–1618 (2006).
75. F. Mémoli, On the use of Gromov-Hausdorff distances for shape comparison, in *Proceedings of Point Based Graphics*, Prague, Czech Republic (2007).
76. F. Mémoli, Gromov–Hausdorff distances in Euclidean spaces, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June (2008), pp. 1–8.
77. F. Mémoli, Spectral Gromov–Wasserstein distances for shape matching, in *Workshop on Non-rigid Shape Analysis and Deformable Image Alignment (ICCV Workshop, NORDIA'09)*, October 2009.
78. F. Mémoli, *Estimation of distance functions and geodesics and its use for shape comparison and alignment: theoretical and computational results*. PhD thesis, Electrical and Computer Engineering Department, University of Minnesota, May 2005.
79. F. Mémoli, Spectral Gromov–Wasserstein distances and related approaches, *Appl. Comput. Harmon. Anal.* (2010). doi:10.1016/j.acha.2010.09.005.
80. F. Mémoli, G. Sapiro, Comparing point clouds, in *SGP '04: Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing* (2004), pp. 32–40.
81. F. Mémoli, G. Sapiro, A theoretical and computational framework for isometry invariant recognition of point cloud data, *Found. Comput. Math.* **5**(3), 313–347 (2005).

82. P.W. Michor, D. Mumford, Riemannian geometries on spaces of plane curves, *J. Eur. Math. Soc.* **8**(1), 1–48 (2004).
83. M.I. Miller, L. Younes, Group actions, homeomorphisms, and matching: a general framework, *Int. J. Comput. Vis.* **41**(1–2), 61–84 (2001).
84. M.I. Miller, A. Trouvé, L. Younes, On the metrics and Euler-Lagrange equations of computational anatomy, *Annu. Rev. Biomed. Eng.* **4**(1), 375–405 (2002).
85. D. Mumford, Mathematical theories of shape: do they model perception? *Proc. SPIE* **1570**, 2–10 (1991).
86. R. Osada, T. Funkhouser, B. Chazelle, D. Dobkin, Shape distributions, *ACM Trans. Graph.* **21**(4), 807–832 (2002).
87. M. Ovsjanikov, Q. Mérigot, F. Mémoli, L. Guibas, in *One Point Isometric Matching with the Heat Kernel*, Lyon, France (2010), pp. 1555–1564.
88. C.H. Papadimitriou, K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity* (Dover, Mineola, 1998). Corrected reprint of the 1982 original.
89. P.M. Pardalos, H. Wolkowicz (eds.), *Quadratic Assignment and Related Problems*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 16 (American Mathematical Society, Providence, 1994). Papers from the workshop held at Rutgers University, New Brunswick, New Jersey, May 20–21, 1993.
90. Protein Data Bank. RCSB, Protein Data Bank. Rutgers University and UCSD. http://www.rcsb.org/pdb/home/home.do.
91. H. Pottmann, J. Wallner, Q. Huang, Y.-L. Yang, Integral invariants for robust geometry processing, *Comput. Aided Geom. Des.* **26**(1), 37–60 (2008).
92. D. Raviv, A. Bronstein, M. Bronstein, R. Kimmel, Symmetries of non-rigid shapes, in *IEEE 11th International Conference on Computer Vision*, October 2007 (2007), pp. 1–7.
93. G. Reeb, Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique, *C. R. Math. Acad. Sci. Paris* **222**, 847–849 (1946).
94. M. Reuter, F.E. Wolter, N. Peinecke, Laplace-spectra as fingerprints for shape matching, in *SPM '05: Proceedings of the 2005 ACM Symposium on Solid and Physical Modeling*, New York, NY, USA (ACM Press, New York, 2005), pp. 101–106.
95. M. Riser, Protein docking using local shape distributions. Master's thesis, ETH Zürich, Switzerland, September 2004.
96. Y. Rubner, C. Tomasi, L. Guibas, The Earth mover's distance as a metric for image retrieval, *Int. J. Comput. Vis.* **40**(2), 99–121 (2000).
97. M. Ruggeri, D. Saupe, Isometry-invariant matching of point set surfaces, in *Proceedings Eurographics 2008 Workshop on 3D Object Retrieval* (2008).
98. M. Rumpf, B. Wirth, An elasticity-based covariance analysis of shapes, *Int. J. Comput. Vis.* **92**(3), 281–295 (2009).
99. S. Rusinkiewicz, M. Levoy, Efficient variants of the icp algorithm, in *3DIM* (IEEE Comput. Soc., Los Alamitos, 2001), pp. 145–152.
100. R.M. Rustamov, Laplace–Beltrami eigenfunctions for deformation invariant shape representation, in *Symposium on Geometry Processing* (2007), pp. 225–233.
101. T. Sakai, *Riemannian geometry*. Translations of Mathematical Monographs, vol. 149 (American Mathematical Society, Providence, 1996).
102. Y. Shi, P.M. Thompson, G.I. de Zubicaray, S.E. Rose, Z. Tu, I. Dinov, A.W. Toga, Direct mapping of hippocampal surfaces with intrinsic shape context, *NeuroImage* **37**(3), 792–807 (2007).
103. C.G. Small, *The Statistical Theory of Shape*. Springer Series in Statistics (Springer, New York, 1996).
104. V. Strassen, The existence of probability measures with given marginals, *Ann. Math. Stat.* **36**, 423–439 (1965).
105. K.T. Sturm, On the geometry of metric measure spaces. I, *Acta Math.* **196**(1), 65–131 (2006).
106. R.W. Sumner, J. Popovic, Mesh data from deformation transfer for triangle meshes. http://people.csail.mit.edu/sumner/research/deftransfer/data.html.
107. D. Thompson, *On Growth and Form* (Cambridge University Press, Cambridge, 1917).
108. A. Trouvé, Diffeomorphisms groups and pattern matching in image analysis, *Int. J. Comput. Vis.* **28**(3), 213–221 (1998).
109. R.C. Veltkamp, Shape matching: similarity measures and algorithms, in *Shape Modeling International*, doi:10.1109/SMA.2001.923389.

110. R.C. Veltkamp, L.J. Latecki, Properties and performances of shape similarity measures, in *Content-Based Retrieval*, ed. by T. Crawford, R.C. Veltkamp. Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, vol. 06171 (2006).
111. C. Villani, *Topics in Optimal Transportation*. Graduate Studies in Mathematics, vol. 58 (American Mathematical Society, Providence, 2003).
112. L. Younes, *Shapes and Diffeomorphisms* (Springer, Berlin, 2010).
113. L. Younes, Computable elastic distances between shapes, *SIAM J. Appl. Math.* **58**, 565–586 (1995).
114. L. Younes, Optimal matching between shapes via elastic deformations, *Image Vis. Comput.* **17**(5–6), 381–389 (1999).
115. L. Younes, P.W. Michor, J. Shah, D. Mumford, A metric on shape space with explicit geodesics, *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.* **19**(1), 25–57 (2008).
116. W. Zeng, X. Yin, Y. Zeng, Y. Lai, X. Gu, D. Samaras, 3D face matching and registration based on hyperbolic Ricci flow, in *CVPR Workshop on 3D Face Processing*, Anchorage, Alaska, June 2008 (2008), pp. 1–8.
117. W. Zeng, Y. Zeng, Y. Wang, X. Yin, X. Gu, D. Samaras, 3D non-rigid surface matching and registration based on holomorphic differentials, in *The 10th European Conference on Computer Vision (ECCV)*, Marseille, France, October 2008
118. Z. Zhang, Iterative point matching for registration of free-form curves and surfaces, *Int. J. Comput. Vis.* **13**(2), 119–152 (1994).