Appl. Comput. Harmon. Anal. 30 (2011) 363–401

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

A spectral notion of Gromov–Wasserstein distance and related methods

Facundo Mémoli *

Department of Mathematics, Stanford University, CA 94305, United States

ARTICLE INFO

Article history:
Received 12 February 2010
Revised 7 August 2010
Accepted 16 September 2010
Available online 29 September 2010
Communicated by Mauro Maggioni

Keywords:
Shape and data analysis
Spectral methods
Gromov–Wasserstein distance
Gromov–Hausdorff distance
Heat kernel
Spectral invariants
Diffusion distance
Laplace–Beltrami operator
Mass transport
Metric geometry

ABSTRACT

We introduce a spectral notion of distance between objects and study its theoretical properties. Our distance satisfies the properties of a metric on the class of isometric shapes, which means, in particular, that two shapes are at 0 distance if and only if they are isometric when endowed with geodesic distances. Our construction is similar to the Gromov–Wasserstein distance, but rather than viewing shapes merely as metric spaces, we define our distance via the comparison of heat kernels. This allows us to establish precise relationships of our distance to previously proposed spectral invariants used for data analysis and shape comparison, such as the spectrum of the Laplace–Beltrami operator, the diagonal of the heat kernel, and certain constructions based on diffusion distances. In addition, the heat kernel encodes a natural notion of scale, which is useful for multi-scale shape comparison. We prove a hierarchy of lower bounds for our distance, which provide increasing discriminative power at the cost of an increase in computational complexity. We also explore the definition of other spectral metrics on collections of shapes and study their theoretical properties.

1. Introduction

Due to the recent great advances in data acquisition and in shape acquisition and modelling, huge collections of datasets and digital models are becoming widely available. One of the major challenges in organizing these collections is to be able to define and compute meaningful notions of dissimilarity or distance between objects and classes of objects. In many scenarios that involve deformable objects, these notions of distance must exhibit invariance to deformations or poses of the objects. By "pose" we mean an arbitrary embedding of the shape in some ambient space.[1] Problems of this nature arise in areas such as molecular biology, metagenomics, face recognition and matching of articulated objects and pattern recognition in general.

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 [1], the *shape contexts* of Belongie et al. [2], the integral invariants of [3], the eccentricity functions of [4], the *shape distributions* of [5], the *canonical forms* of [6], and the *Shape-DNA* methods in [7]. The common idea of these methods is to compute certain *invariants*, or *signatures* of the shapes (we-will use these two terms inter-changeably). These signatures are then embedded into a common metric space to simplify comparison, and shapes whose signatures are at a small distance from each other are considered similar.

* Fax: +1 (650) 725 4066.
E-mail address: memoli@math.stanford.edu.
URL: http://math.stanford.edu/~memoli.

[1] Naturally, some restrictions need to be imposed on such embeddings.

1063-5203/$ – see front matter
doi:10.1016/j.acha.2010.09.005

The question of proving that a given family of signatures captures proximity or similarity of shapes in a reasonable way has hardly been addressed in the shape matching community. In particular, the degree to which two shapes with similar signatures are forced to be similar is in general not well understood. Conversely, one can ask the easier question of whether the similarity between two shapes forces their signatures to be similar.

These questions cannot be well formulated until one agrees on compatible notions of (1) *equality* and (2) *dis-similarity* between shapes. Compatibility here means that shapes which are 0-dissimilar are exactly those that are considered equal according to the notion of equality that one imposes.

A tentative but rather general framework for studying these problems would be to declare that we have a space $\mathfrak{S}$ comprising all shapes of interest, a notion of equality "=" between pairs of shapes, and a *metric* $D : \mathfrak{S} \times \mathfrak{S} \to \mathbb{R}^+$ s.t. $D(X, Y) = 0$ if and only if $X = Y$. In this context an $\mathbb{R}$-valued *invariant* is any map $I : \mathfrak{S} \to \mathbb{R}$ with the property that $I(X) = I(Y)$ whenever $X = Y$. In more generality one can consider V-valued invariants where V is some given metric space with metric d_V.

The question of whether similarity between two shapes forces the valuation of a given V-invariant I on the two shapes to be similar can now be phrased easily in terms of the concept of *stability* of the invariant I. We make precise what is the notion of stability of invariants that we are after.

Remark 1.1 *(Quantitative stability).* In order to say that the invariant $I : \mathfrak{S} \to V$ is quantitatively stable under D, we require that

$$D(X, Y) \geqslant d_V\big(I(X), I(Y)\big) \quad \text{for all shapes } X \text{ and } Y \text{ in } \mathfrak{S}.$$

See Theorems 8.1 and 8.2 for explicit results. This is in contrast with a different, weaker, type of stability that one might be able to prove which is of the form: if $(X_n) \subset \mathfrak{S}$ is a sequence of shapes which converges to a shape X in the metric D, then $I(X_n) \to I(X)$ as $n \uparrow \infty$ in the metric d_V.[2]

It is important to remark that, for a given metric D on $\mathfrak{S}$, knowing that a large number of invariants are stable according to our definition above is clearly desirable. Knowledge of stability of the invariants means that they can be used to discriminate shapes in a manner that is insensitive to perturbations of the shapes themselves (in the sense encoded by D). Furthermore, the precise form of the stability statement for a given invariant provides a lower bound for the distance D between two shapes. This has clear impact on the practical application of the ideas, since typically, the computation of $d_V(I(X), I(Y))$ will be less computationally demanding than the computation of $D(X, Y)$.

One instantiation of the type of framework alluded to above was suggested in [8,9], where the authors (1) consider shapes to be compact metric spaces, (2) define two shapes to be equal when they are *isometric*, and (3) use the *Gromov–Hausdorff distance* [10] as a measure of dissimilarity between shapes. Two metric spaces (X, d_X) and (Y, d_Y) are said to be isometric whenever there exists a bijective map $\Psi : X \to Y$ that preserves distances: $d_X(x, x') = d_Y(\Psi(x), \Psi(x'))$ for all $x, x' \in X$. Such a map Ψ is called an *isometry*. The Gromov–Hausdorff distance between isometry classes of compact metric spaces is defined as the infimal $\varepsilon > 0$ such that there exist a metric space Z and $X', Y' \subset Z$ s.t. X' is isometric to X and Y' is isometric to Y and the Hausdorff distance (in Z) between X' and Y' is bounded above by ε:

$$d_{\mathcal{H}}^Z\big(X', Y'\big) < \varepsilon.$$

By endowing shapes with different metrics, one obtains a great deal of flexibility in the different degrees of invariance that one can encode in the Gromov–Hausdorff measure of dissimilarity. For instance, endowing shapes embedded in $\mathbb{R}^d$ with the (restricted) Euclidean metric makes the Gromov–Hausdorff distance invariant under ambient rigid isometries [11]. In contrast, using intrinsic metrics within the shapes makes the Gromov–Hausdorff distance blind to intrinsic isometries (or *bends*). Here, the word *intrinsic* refers to the independence of the metric on the embedding, and an example of such a metric is the geodesic distance on a smooth surface.

Once this Gromov–Hausdorff framework is adopted, the natural question arises whether a given family of signatures is *quantitatively stable* under perturbations of the shape in the Gromov–Hausdorff (GH from now on) sense. A very simple example is given by the $\mathbb{R}$-valued invariant given by the *diameter*:

$$I_{\text{diam}} : \mathfrak{S} \to \mathbb{R} \quad \text{such that } X \mapsto \mathbf{diam}(X) = \max_{x,x' \in X} d_X\big(x, x'\big).$$

It follows easily from the definition of the Gromov–Hausdorff distance (see Proposition 2.2) that I_{diam} is quantitatively stable:

$$d_{\mathcal{GH}}(X, Y) \geqslant \frac{1}{2}\big|\mathbf{diam}(X) - \mathbf{diam}(Y)\big|.$$

Unfortunately, despite its generality, it has been difficult to use the Gromov–Hausdorff framework in a natural way for explaining most of the existing signature based shape matching procedures, with the exception of [12]. In addition, the

[2] This can be likened to the difference between continuity and Lipschitz or Hölder continuity in the context of real valued functions.

practical computation of the GH distance leads to NP-hard combinatorial optimization problems which are related to the bottleneck quadratic assignment problem [13].

These issues were addressed in [11,13,14], where the author defines the *Gromov–Wasserstein distance* (GW from now on) based on ideas from *mass transportation*. This family of distances:

- exhibits a number of desirable theoretical features,
- directly yields continuous variable quadratic optimization problems with linear constraints, and
- provides a dissimilarity measure under which a large number of shape signatures become quantitatively stable.

The framework of [13,14] assumes that in addition to a metric, shapes are also endowed with a certain notion of *weight* associated to each point of the shape. The signatures of [2–5] have all been shown to be quantitatively stable under perturbations in the Gromov–Wasserstein sense in [13]. More recently, certain persistence topology based signatures have also been shown to be GW stable [12].

1.1. Spectral methods in data and shape analysis

In this work, we aim to extend the ideas in [13] to a different class of techniques for shape or data classification and comparison, usually referred to as *spectral methods*. These methods generally rely on constructions that use the eigenvalues and eigenfunctions of the Laplace–Beltrami operator defined on shapes. Expositions of these methods from the points of view of geometry processing, pattern recognition, and applied harmonic analysis can be found in [15], in the PhD thesis of Qiu [16], and in the PhD thesis of Lafon [17], respectively.

Our work is inspired in large part by the recent development of spectral methods for geometry processing, data and shape analysis, and shape, graph and data matching [7,18–33].

The spectrum of the Laplace–Beltrami operator, which is perhaps the best known spectral invariant, was introduced to the shape matching community by Reuter et al. in the remarkable [7], where, in the context of shape retrieval and comparison, the authors propose to use a subset of the collection of all eigenvalues (spectrum) of the Laplace–Beltrami operator of a shape as its signature. The invariance of the spectrum of the Laplace–Beltrami operator to intrinsic isometries (deformations that leave the geodesic distance unchanged) ensures that this signature can be used to recognize the same shape in different poses. From a theoretical point of view, however, it is not possible to fully classify shapes using this signature, since there exist compact non-isometric shapes whose Laplace–Beltrami operators have the same spectra [34]. The work of Reuter et al. has provided us with the very attractive problem of proving lower bounds for a suitably modified notion of GH or GW distances that takes into account the spectra of the Laplace–Beltrami operators on shapes.

Problem 1. Define a metric on the collection of all shapes and find a suitable reinterpretation of the spectrum of the Laplace–Beltrami operator on shapes such that the spectrum becomes *quantitatively stable* under perturbations of the shape in this metric.

We mention in passing that an interesting theoretical problem would be to ascertain whether statement like “most shapes (in a certain, rich, class) can be classified using the spectrum” could be true (provided it is made rigorous). A sense in which such a statement could hold is that given by [35].

The work of Rustamov [20] is based on the observation that the eigenvalues of the Laplace–Beltrami operator together with the corresponding eigenspaces characterize the shape up to isometry. The author introduces the Global Point Signature (GPS) of a point on the shape, which encodes both the eigenvalues and the eigenfunctions of the Laplace–Beltrami operator evaluated at that point. The histogram of distances [5] between the signatures of all points is then used for shape comparison.

Problem 2. Find a reinterpretation of Rustamov’s proposal that becomes (quantitatively) stable under perturbations of the shape in this metric we construct.

More recently, Sun et al. [23] and Gebal et al. [36] simultaneously introduced a robust and multi-scale invariant of a shape based on the *heat kernel*. The heat kernel arises as the fundamental solution of the *heat equation* on a shape. This PDE involves the Laplace–Beltrami operator. Their signature, called the Heat Kernel Signature (or HKS) coincides with the diagonal of the heat kernel[3] and is defined for every point on the shape. By construction, this signature is also inherently multi-scale. Interestingly, Sun et al. proved that the set of all HKS on the shape “almost always” characterizes it up to isometry, see [23] for details and [23,36–39] for applications of this invariant on shape analysis tasks.

Problem 3. Define a metric on the collection of all shapes and find a suitable reinterpretation of the HKS that becomes (quantitatively) stable under perturbations of the shape in this metric.

[3] Scaled by the volume of the shape, see Definition 4.1.

Methods based on Heat Diffusion have also been used to analyze and match graphs (see [25,26,32] and references therein), where signatures similar to the HKS have proven useful in graph classification and comparison.

In the context of data analysis and matching, spectral ideas [40] or ideas based on heat diffusion have also been developed and successfully applied to several datasets [17,27–29,31,41].

A common idea of [28,41,42] and [27] is to embed a shape/dataset into a finite dimensional Hilbert space where the coordinates of each point are defined in some manner using the eigenfunctions and eigenvalues of the Laplace–Beltrami operator (or of a discrete version thereof). It has been observed that such embeddings can provide a great deal of insight into the underlying structure the datasets under consideration [17]. The comparison between two different shapes or datasets then proceeds by computing some notion of distance, e.g. the Hausdorff distance [28] or the Wasserstein distance (a.k.a. Earth Mover's distance) [27], between the images of the datasets via those embedding maps.

Jones et al. [43] have made a detailed study of the properties of the parametrizations provided by continuous constructions analogous to [40,44]. They prove that these satisfy a number of desirable properties with very lax assumptions on the underlying metric structures.

It has been observed in [20,33] that a potential problem with the type of shape matching procedure described above is that the spectral embedding of the shapes depends on the choice of the eigenfunctions one makes. Even in the case that the eigenvalues are non-repeating, there is a choice of sign for each of the eigenfunctions. One then would be led to considering which choice of signs for the eigenfunctions of the two shapes produces the smallest distance between the images through the embedding maps. This task could certainly be computationally demanding if done with a large number of embedding coordinates.

By directly conceiving their procedures using heat diffusions, the proposals of [17,28,41,44] already encode a certain notion of scale, which is given by the time parameter governing the diffusion. This has led Coifman and collaborators to employ the term *multiscale geometries* [41]. This is an extremely important feature of these methods which will also be present in our construction.

Problem 4. Relate the metric we construct to the shape/data matching idea underlying the proposals of [17,27,28,41].

Admittedly, the statement of this problem is rather vague. We provide precise formulations in Section 9.

As discussed above, the proposal of [8,9] leaves the door open for endowing shapes with any user specified notion of metric. This has been revisited recently by Bronstein et al. [24] who endow shapes with the spectrum-based *diffusion distance* (introduced to the applied literature by Lafon in [17]) and then estimate the Gromov–Hausdorff distance between the resulting metric spaces. The diffusion distance is an intrinsic notion of distance on Riemannian manifolds that can be related to diffusion processes taking place on such manifolds.

The motivation of [24] is to exploit the apparent stability of diffusion distances to local changes in the topology of the shape. As a result, their method can potentially be used to classify shapes undergoing such changes. One of the limitations of [24] is the difficulty in establishing the exact relationship between their distance and different notions of invariance. Indeed, it is unclear whether two objects whose diffusion distance based GH distance is 0 are necessarily isometric with respect to geodesic distances. In the same vein, it is unclear what is the precise notion of similarity encoded by the fact that two shapes (endowed with the diffusion distance) are at small GH distance from each other. These issues were not addressed by the authors of [24] but they provide somewhat encouraging experimental support for their claim that diffusion distances are more adequate for practical shape matching.

Problem 5. Prove that the proposal of [24] yields a lower bound for a true metric between shapes that we construct.

Also inspired by the useful properties of the diffusion distance, de Goes et al. [25] propose to use a certain function defined on a shape which they call *average diffusion distance*. This function is used for shape segmentation tasks and is in the same spirit of the *geodesic eccentricity function* that appears in the work of Hilaga et al. [45] and Hamza and Krim [4]. This diffusion distance based eccentricity function is an invariant of a shape which is based on spectral constructions. In this paper we also tackle:

Problem 6. Prove that the diffusion distance based eccentricity function of [25] is quantitatively stable in the sense of the metric we construct.

Despite this rich body of work, many of the spectral methods proposed so far for shape comparison and data matching often suffer from the lack of theoretical foundation that would (1) establish precisely what are the properties of the measures of dissimilarity between shapes they put forward, and (2) unify the different proposals.

In particular, it is not well understood what is precisely the notion of equality between shapes operating under the hood in these different proposals. In more generality, one can say that there exist a plethora of seemingly disjoint methods that have been proposed for tackling the shape and data matching problems, but however there exists no precise understanding of their interrelationship.

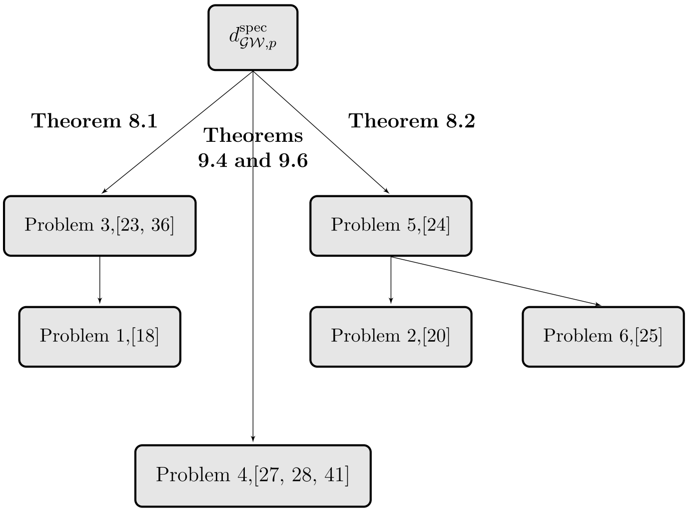


Fig. 1. Structure of the lower bounds for the spectral GW distance (which we denote by $d^{\mathrm{spec}}_{\mathcal{GW},p}$). The lower bounds are hierarchical in the manner depicted and they provide our answers to Problems 1 to 6, see Theorems 8.1 and 8.2.

Elucidating connections between different methods calls for understanding the relationships between the different invariant signatures that have been proposed. This in turn, asks for an understanding of which signatures are equivalent to or stronger than other signatures.

In particular, a weak area is that dealing with the choice of distance between the different shape invariants or signatures used to compare shapes. For example, the spectrum of the Laplace–Beltrami operator is a natural invariant of a non-rigid shape. However, comparing the spectra of two different shapes in a reasonable way is not obvious.

Based on the work of Reuter et al. [18] one could propose using the ℓ^2 metric between the two full sets of ordered eigenvalues. This metric, however, will tend to give more weight to larger eigenvalues, which correspond to "high frequency eigenfunctions". Even more dramatically, it is not guaranteed to converge when considering the full spectrum of the Laplace–Beltrami operator of a shape, and is therefore not defined in all cases, see Remark 4.4. Similarly, comparing the sets of eigenfunctions of the Laplace–Beltrami operators of two shapes is not easy (since there may be many *sign flips*, see [20]) and the histogram approach proposed by Rustamov [20] is only one of the many possible methods.

Problem 7. Interrelate (suitable reinterpretations of) the different spectral methods [7,20,23–25,27,28].

Our answer to Problem 7 lies in the panorama described in Fig. 1.

In this article we aim to propose solutions to Problems 1–6 above. We do this by constructing a notion of distance between two shapes and by proving that it satisfies the metric axioms on the collection of all isometry classes of Riemannian manifolds without boundary. Further, we demonstrate properties of this distance via a series of lower bounds relating our distance to different quantities (which are generally more tractable) and directly make use of several of the previously proposed spectral invariants mentioned above.

One particular aspect of the problem to which we pay close attention is the question of similarity between shapes based on the proximity of their signatures. In applications such as shape retrieval it is essential to be able to claim that shapes, whose signatures are significantly different are dissimilar, without incurring the typically higher computational cost of having to compute the full notion of dissimilarity. The hierarchies of lower bounds that we prove make a step in the direction of understanding the relation between the distances between various "spectral" signatures of the shapes and the degree of their similarity.

1.2. Metrics on shapes in the mathematical literature

Gromov [46] introduced the Gromov–Hausdorff distance, a metric on the collection of all compact metric spaces modulo isometries, and established a pre-compactness theorem on certain subclasses. As an application in the context of Riemannian manifolds, and as a consequence of the Bishop–Gromov volume comparison theorem, uniform upper bounds on diameters and uniform lower bounds on Ricci curvatures guarantee pre-compactness.

In a similar spirit, but using a more analytical construction based on the spectra of the Laplace–Beltrami operator defined on the manifolds, Berard, Besson and Gallot [47] defined a different notion of distance for which they identified pre-compact classes under hypotheses similar to those of Gromov. Kasue and Kumura [48] proposed another spectral notion of distance based on a certain modification of the Gromov–Hausdorff distance to which our construction is very similar.

More recently, other notions of distance have been proposed that allow to extend the original framework of ideas to richer classes of geometric structures. These structures are metric measure spaces: triples (X, d, ν) where (X, d) is a compact metric space and ν is a Borel probability measure on X. Gromov himself proposed a couple possible distances between measure metric spaces ans studied conditions that ensure compactness of families. By exploiting the connections of Ricci curvature with mass transportation concepts, the work of Lott and Villani [49], Sturm [50] and Ollivier [51] has expanded considerably the understanding of how curvature may be defined on metric measure spaces and how it may play a role in guaranteeing compactness for families of spaces.

1.3. Contributions and summary of our approach

Our contribution aims to be that of providing a set of ideas that unifies several previously related procedures for data/shape matching and analysis. From the point of view of shape and data matching, we provide solutions to Problems 1, 2, 3, 4, 5 and 6 above. From the more general shape analysis perspective, our results can be interpreted as establishing that many invariants that have been used for tasks other than matching (for segmentation, for example) are stable in a precise way.

The spectral notion of distance between shapes (which are regarded as compact Riemannian manifolds without boundary) that we define in this paper, is called the *Spectral Gromov–Wasserstein distance*. In Theorem 6.1 we formally show that our definition satisfies the properties of a metric on the collection of all isometry classes of shapes. This means, in particular, that two shapes at 0 distance from each other are necessarily isometric with respect to the geodesic distance.

As has been put forward in the data analysis community by Prof. Coifman and collaborators [17,28–30,41,44,52], and recently in the shape analysis community by Sun et al. [23], the heat kernel and derived invariants already contain a notion of *geometric scale* that is very interesting from the practical point of view. In agreement with this observation, we also encode this *scale parameter* into our definition and further argue how our notion of similarity/equality between shapes is *foliated* with respect to this parameter.

We also address the question of similarity between datasets/shapes by proving, in Theorems 8.1 and 8.2, a series of *lower bounds* on our metric that involve previously proposed spectral signatures. These lower bounds imply that two shapes such that a suitably chosen distance between their signatures is large, have to be far in terms of our spectral metric. In particular, in Theorem 8.1 we prove that two (interrelated) invariants: the HKS of [23,36] and the *heat trace*, are both stable with respect to the metric we construct.

One of the main observations is that the heat trace contains *exactly the same information as the spectrum* of [7],[4] and therefore the stability of the heat trace in the spectral GW sense can be reinterpreted as stability of the spectrum. This provides the formalization of Problem 1 that we posed.

A second, third and fourth sets of previously proposed ideas that we address with our construction and relate to are those of [20,27,28] and [24]. Again, we exhibit lower and upper bounds for our metric that establish explicit links (via Theorems 8.2, 9.4 and 9.6) to (suitably reinterpreted version of) those extremely interesting proposals.

At a high level, our construction is based on substituting the heat kernels in the definition of the Gromov–Wasserstein distance for the geodesic distances. This is motivated by classical the result by Varadhan (Lemma 4.1) which relates these two quantities:

$$-4t\ln k_X(t,x,x') \simeq d_X^2(x,x') \quad \text{for } t \simeq 0^+.$$

Using the heat kernel, however, has the advantage of directly encoding a scale parameter (t), which allows for scale dependent comparisons. We then use *measure couplings*, in the same way as done in the *mass transportation* inspired [13] to compare heat kernels on different shapes. Thus, our proposed dissimilarity measure (see Definition 6.1) is[5]

$$\inf_{\mu}\sup_{t>0} c^2(t)\cdot \left\| k_X(t,\cdot,\cdot) - k_Y(t,\cdot,\cdot)\right\|_{L^p(\mu\otimes\mu)} \tag{1.1}$$

where μ is a coupling between the normalized area measures on the shapes X and Y; $p \geqslant 1$; and $c(t)$ is a certain function that prevents the blow up of the heat kernels. Note that for a fixed coupling μ we are taking $\sup_{t>0}$ which is to be interpreted as choosing the *most discriminative scale* that tells X apart from Y.

1.4. Interpretation of our construction from the point of view of physics

Let's assume that a physicist is presented with two shapes X and Y (made of the same material) and that he is asked to establish how different these two shapes are, in some meaningful manner. Different "physical procedures" would be possible, but in this situation let's assume that the physicist has a predilection for thermodynamical concepts, over, say, electromagnetic concepts.

[4] This is certainly a classical observation.

[5] Here we are assuming for simplicity that the volumes of X and Y both equal 1. We give the general definition in Section 6.

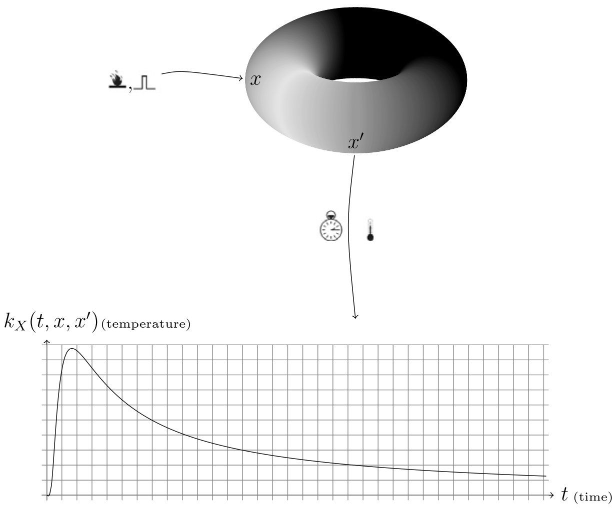


Fig. 2. A physics based way of characterizing/measuring a shape. For each pair of points x and x' on the shape X, one heats a tiny area around point x to a very high temperature in a very short interval of time around $t = 0$. Then, one measures the temperature at point x' for all later times and plots the resulting graph $k_X(t, x, x')$ as a function of t. The knowledge of these graphs for all $x, x' \in X$ and $t > 0$ translates into knowledge of the heat kernel of X (the plot in the figure corresponds to $x \neq x'$). In contrast, one can think that a geometer's way of characterizing the shape would be via the use of a geodesic ruler that can be used for measuring distances between all pairs of points on X. According to Varadhan's Lemma, both approaches are equivalent in the sense that they both capture the same information about X.

Recall that for a given shape X, the value $k_X(t, x, x')$ of the heat kernel can be thought of as the values of the temperature measured at x' at time t given that at time 0 a very small area around x was heated to a very high temperature.

We assume that armed with the correct tools, the physicist will measure the shapes in accordance with the description in Fig. 2.

By time invariance considerations, one sees that the variable t (time) is commensurable across both shapes[6] and therefore the physicist will believe that t will have to play special role in the construction of the measure of dissimilarity.

The physicist notices that the measured profiles for each pair of points x, x' on a shape X (see Fig. 2) may take very large values for $t \simeq 0$ (they do for $x = x'$, see (4.19)) and therefore he decides to affect the measured profiles by a dampening function $c(t)$ that he will use for "correcting" all his measurements of different shapes.[7] In other words, we can assume that the measurements corresponding to a shape X are the collection of functions

$$\left\{c^2(t) \cdot k_X\left(\cdot, x, x'\right) : \mathbb{R}^+ \to \mathbb{R}^+\right\}_{x,x' \in X}.$$

The physicist is then led to thinking about how to put in correspondence the spatial part of his measurements. A natural idea is that of finding the maps $f : X \to Y$ and $g : Y \to X$ that provide the best possible agreement between the measurements on X and the measurements on Y *for all* $t > 0$.

In other words, one would conceivably seek the infimal $\varepsilon > 0$ such that there exist maps f and g with the property that for all $t > 0$

- $|k_X(t, x, x') - k_Y(t, f(x), f(x'))| < \varepsilon / c^2(t)$ for all $x, x' \in X$, and
- $|k_X(t, g(y), g(y')) - k_Y(t, y, y')| < \varepsilon / c^2(t)$ for all $y, y' \in Y$.

In our actual construction (1.1), we find it more convenient not to use "hard maps" for establishing correspondences between shapes, but we prefer to use of the more flexible notion of measure couplings as discussed above. We also, opt for the more relaxed L^p norms as opposed to the choice of the L^∞ norm made by the physicist above.

The physics inspired construction that we just described corresponds to what a spectral Gromov–Hausdorff distance [48] would look like, see Section 9.1.

1.5. A zoo of metrics: How to choose?

There are many options for a metric on the collection of all shapes which will satisfy that two shapes are at distance 0 if and only they are isometric (w.r.t. the geodesic distance). In the present paper, shapes will always mean Riemannian manifolds without boundary.

[6] We assume (for now) that both shapes have the same total area.

[7] The precise form of this function should not concern us for the moment.

With regards to the choice of the metric there are two issues that we wish to address:

(1) What are the precise notions of similarity between shapes encoded by each of these metrics?
(2) How does the choice of the metric affect the claims about stability of invariants that one can make?

We now turn our attention to (1). A first example of a metric is the Gromov–Hausdorff distance discussed in Section 1. Other examples are the Gromov–Wasserstein distance [13,50], and the spectral distances that we discuss in this paper.

As we said at the beginning of this section, all these metrics agree on what it means for two shapes X and Y to be 0-dissimilar. They disagree however on the meaning of non-zero dissimilarity. For example, the fact that two shapes X and Y be at Gromov–Hausdorff distance smaller than $\varepsilon > 0$ means that one can find a map $f : X \to Y$ with the properties that

- $|d_X(x,x') - d_Y(f(x), f(x'))| < 2\varepsilon$ for all $x, x' \in X$, and
- $f(X)$ is a 2ε-net for Y.[8]

This can be understood as a notion of relaxed isometry and actually receives the name of an ε-isometry [53]. Also, there is a sense in which ε_n-isometries, for $\varepsilon_n \to 0$ as $n \uparrow \infty$, converge to isometries, therefore the concept of ε-isometry is a generalization of the concept of isometry and this provides a simple way of interpreting what it means to have two shapes at small Gromov–Hausdorff distance (there are many other ways of arguing about this, though).

With regards to the notion of dissimilarity encoded by the Gromov–Wasserstein distance [13,50], in this case, by a result of [11], having that X and Y are at Gromov–Wasserstein distance less than ε, implies that one can find $X_\varepsilon \subset X$ and $Y_\varepsilon \subset Y$ such that $d_{\mathcal{GH}}(X_\varepsilon, Y_\varepsilon) < \eta(\varepsilon)$ and X_ε and Y_ε are a large fraction of the points in X and Y, respectively. The precise nature of the claim (together with the behavior of $\eta(\cdot)$) can be consulted in [11]. The important point is that now, one can find an interpretation for the Gromov–Wasserstein distance based on the one for the Gromov–Hausdorff distance that highlights the fact that the former distance ensures the existence of $2\eta(\varepsilon)$ isometries between significant sub parts of the shapes X and Y.

The interpretation of the notion of dissimilarity encoded by the metric that we construct in this paper (the spectral Gromov–Wasserstein distance) is partially contained in the account of the physics inspired procedure described above (see Fig. 2). We will touch upon this in subsequent sections of the paper and try to relate this issue with that of understanding the parameter t as a notion of geometric scale that the user of the spectral Gromov–Wasserstein distance may want to profit from.

With respect to point (2), it is clear that different choices of the metric that we put on our collection of shapes will give the ability to control different sets of invariants. For instance, the Gromov–Hausdorff is purely metrical, and even when restricted to the class of Riemannian manifolds, it seems difficult to establish quantitative stability of invariants that encode elements that depend on the area measure.[9] Thus, by utilizing the measures directly into the notion of distance, as is the case for the Gromov–Wasserstein distances, one is able to obtain more control over a larger range of invariants which have been of interest to practitioners (see discussion in [12–14]).

The construction of the spectral version of the Gromov–Wasserstein distance is therefore an attempt at providing a metric which is rich enough to control many of the invariants of spectral nature that have been proposed in the literature. This might be achievable with the Gromov–Hausdorff distance directly, but at the present moment we do not know whether this is possible.

1.6. From a more theoretical point of view

As a consequence of our construction, and in particular of proving bound of the type modeled in Remark 1.1, we are able to compute lower bounds for the distances between certain standard Riemannian manifolds, this is typically more involved in the purely metric context of GH distances, see Examples 8.1 and 8.2 and [54].

In a related vein, we establish Proposition 6.1 which deals with Riemannian products $Z_\varepsilon := X \times \varepsilon \cdot Y$, for $\varepsilon > 0$ and smooth Riemannian manifolds X and Y, and proves that Z_ε converges to X as $\varepsilon \downarrow 0$ in the topology generated by the metric we construct. Note that the dimensions of the approaching and limit manifolds are not the same. See Section 2 for the meaning of $\varepsilon \cdot Y$.

In Section 9 we review the construction of the metrics of Berard et al. [47] and Kasue and Kumura [48] and show what is the relationship of these metrics with the spectral Gromov–Wasserstein distance that we define in this paper. We do this by establishing lower and upper bounds involving the different metrics. We also discuss the construction and establish the theoretical properties of other metrics that have been suggested in the applied literature.

[8] That is, the image of X via f is ε-dense in Y.
[9] On Riemannian manifolds, the measure is determined by the metric.

1.7. Outline of the paper

In Section 2 we review some basic theoretical background and terminology used throughout our presentation. In Section 2.1 we recall the main ideas and properties around the GH distance and Section 3 reviews the construction of the GW distance. In Section 4 we give an overview and a formal description of the main spectral methods that motivate our work. There we also recall the construction of the heat kernel and related invariants. Section 5 presents some basic estimates of the heat kernel that are needed both to guarantee that the invariants we construct are bounded and that the spectral GW distance we define is sound. We construct our proposed notion of distance and study its properties in Section 6. We touch upon the notion of geometric scale encoded by the spectral GW distance in Section 7 where we propose a point of view based on homogenization of partial differential equations. Two hierarchies of lower bounds for the spectral GW distance are presented in Section 8, where we also give a few examples on estimating the distance between standard Riemannian manifolds. We give details several other possible constructions of a spectral distances between shapes in Section 9. In order to maximize the readability of the paper, *all proofs are presented in Section 11*. We end the paper in Section 10 with a discussion of related areas that in our opinion deserve more attention.

An announcement of some of the results in this paper has appeared in [55].

2. Background

We review some standard concepts of metric geometry and measure theory that will be used in our presentation. A good reference of the former is [53]. A reference of the latter is [56]. We will use basic concepts from Riemannian geometry which can be consulted in [57,58].

Fix a compact metric space (Z,d) and let $\mathcal{C}(Z)$ denote the collection of all compact subsets of Z and by $\mathbf{diam}(Z)$ the *diameter* of Z, that is $\mathbf{diam}(Z)=\max_{z,z'}d(z,z')$. By $B(z,r)$ we denote the open ball of radius $r>0$ centered at the point $z\in Z$. Recall the definition of the Hausdorff distance on $\mathcal{C}(Z)$:

$$d_{\mathcal{H}}^{Z}(A,B):=\max\Big(\sup_{a\in A}\inf_{b\in B}d(a,b),\sup_{b\in B}\inf_{a\in A}d(a,b)\Big). \tag{2.2}$$

Definition 2.1. An *isometry* between metric spaces (X,d_X) and (Y,d_Y) is any surjective map $\psi:X\to Y$ s.t. $d_X(x,x')=d_Y(\psi(x),\psi(x'))$ for all $x,x'\in X$. We say that X and Y are *isometric* whenever there exists an isometry between these spaces.

Definition 2.2 *(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.

We now express the Hausdorff distance using the language of correspondences:

Proposition 2.1. *(See [13].) Let (Z,d) be a 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 $R\in\mathcal{R}(A,B)$.

For a metric space (Z,d) let $\mathcal{P}(Z)$ denote the collection of all Borel probability measures on Z. By δ_z^Z we denote the delta measure supported at $z\in Z$. Given a measurable map $f:X\to X'$ and $\nu\in\mathcal{P}(X)$, we denote by $(f)_\#\nu$ the *push-forward* measure on $\mathcal{P}(X')$ given by $(f)_\#\nu(A)=\nu(f^{-1}(A))$ for all Borel sets $A\subset X'$. For measures α and β on the measure spaces (X,Σ) and (X',Σ'), respectively, we denote by $\alpha\otimes\beta$ the *product measure* on $(X\times X',\Sigma\times\Sigma')$.

The *weak topology* on $\mathcal{P}(Z)$ is given as follows: one says that a sequence $(\mu_n)\subset\mathcal{P}(Z)$ converges weakly to some $\mu_0\in\mathcal{P}(Z)$ if and only if

$$\int_Z\varphi\,d\mu_n\to\int_Z\varphi\,d\mu_0,\quad\text{as }n\uparrow\infty\text{ for all }\varphi\in C_b(Z).$$

Here $C_b(Z)$ denotes the class of all $\phi:Z\to\mathbb{R}$ which are continuous and bounded.

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 w.r.t. the measure ν. By a slight abuse of notation, for a set $A\subset X$ we write $\|f\|_{L^\infty(A)}$ for $\sup_{x\in A}|f(x)|$.

Remark 2.1. When ν is a probability measure, i.e. $\nu(X)=1$, then $\|f\|_{L^p(\nu)} \geqslant \|f\|_{L^q(\nu)}$ for $p,q\in[1,\infty]$ and $p\geqslant q$.

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

$$\mu(A_0\times B)=\mu_A(A_0) \quad \text{and} \quad \mu(A\times B_0)=\mu_B(B_0) \tag{2.4}$$

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

Definition 2.4. The support of a Borel measure μ on a metric space (Z,d), denoted by supp$[u]$, is the minimal closed subset $Z_0\subset Z$ such that $\mu(Z\backslash Z_0)=0$.

Lemma 2.1. *(See [13].) Let μ_X and μ_Y be Borel probability measures on compact metric spaces X and Y. If $\mu\in\mathcal{M}(\mu_X,\mu_Y)$, then $R(\mu):=\operatorname{supp}[\mu]$ belongs to $\mathcal{R}(\operatorname{supp}[\mu_X],\operatorname{supp}[\mu_Y])$.*

Definition 2.5 *(Wasserstein distances).* (See Chapter 7 of [59].) For each $p\geqslant 1$ we consider the following family of distances on $\mathcal{P}(Z)$, where (Z,d) is a compact metric space:

$$d^Z_{\mathcal{W},p}(\mu_A,\mu_B):=\inf_{\mu\in\mathcal{M}(\mu_A,\mu_B)}\left(\int_{A\times B} d^p(a,b)\,d\mu(a\times b)\right)^{1/p} \tag{2.5}$$

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

$$d^Z_{\mathcal{W},\infty}(A,B):=\inf_{\mu\in\mathcal{M}(\mu_A,\mu_B)}\ \sup_{(a,b)\in R(\mu)} d(a,b). \tag{2.6}$$

These distances are known as Wasserstein–Kantorovich–Rubinstein distances between measures [56,59,60].

These distances do provide metrics on $\mathcal{P}(Z)$ and moreover, it is a standard fact that Wasserstein distances (for $p\in[1,\infty)$) metrize weak convergence [59, Theorem 7.12].

We will denote by $\mathfrak{R}$ the collection of all compact and connected Riemannian manifolds without boundary. For $(M,g)\in\mathfrak{R}$, where g is the metric tensor field on M, we will denote by T_xM the tangent space of M at x; by $\mathbf{vol}_M$ the Riemannian volume measure on M; its total volume by $\mathbf{Vol}(M)=\mathbf{vol}_M(M)$; by $d_M: M\times M\to\mathbb{R}^+$ the *geodesic distance* on M arising from the metric tensor g; and by $\mathbf{Ric}_X$ its Ricci curvature tensor. Given $(M,g)\in\mathfrak{R}$ and $a>0$ we use the notation aM for the element (M,a^2g) of $\mathfrak{R}$. Notice that $\mathbf{Vol}(aM)=a^d\,\mathbf{Vol}(M)$ for a d-dimensional Riemannian manifold. When $(M,g),(M',g')\in\mathfrak{R}$, the product Riemannian manifold is $(M\times M',g+g')\in\mathfrak{R}$. This is to be understood in the following way: if $(x,x')\in M\times M'$, then $T_{(x,x')}M\times M'\simeq T_xM\oplus T_{x'}M'$ and $(g+g')_{(x+x')}: T_{(x,x')}M\times M'\times M\times M'\to\mathbb{R}^+$ is given by $(v+v',w+w')\mapsto g_x(v,w)+g'_{x'}(v',w')$. It follows that $\mathbf{vol}_{M\times M'}=\mathbf{vol}_M\otimes\mathbf{vol}_{M'}$ and in particular, $\mathbf{Vol}(M\times M')=\mathbf{Vol}(M)\cdot\mathbf{Vol}(M')$.

On a Riemannian manifold (M,g), the volume measure $\mathbf{vol}_M$ is determined completely by the metric tensor g. Sometimes it is useful to consider the added flexibility of (partially) decoupling the metric tensor and the measure, and for this we will consider *weighted Riemannian manifolds*, which are triples (M,g,ν) where the measure $\nu=h^2\,\mathbf{vol}_M$ for a smooth positive function h on M, see [61].

In $\mathbb{R}^d$ we denote by $\|x-y\|$ or $\|x-y\|_{\mathbb{R}^d}$ the Euclidean distance between points $x,y\in\mathbb{R}^d$.

Finally, we let ℓ^2 denote the Hilbert space of all square summable sequences and given two such sequences $A=\{a_i\}$ and $B=\{b_i\}$ we denote by $A\bullet_{\ell^2}B=\sum_i a_ib_i$ their inner product.

2.1. The Gromov–Hausdorff distance

Following [10] we introduce the Gromov–Hausdorff distance between (compact) metric spaces X and Y:

$$d_{\mathcal{GH}}(X,Y):=\inf_{Z,f,g} d^Z_{\mathcal{H}}\big(f(X),g(Y)\big), \tag{2.7}$$

where $f:X\to Z$ and $g:Y\to Z$ are isometric embeddings (distance preserving) into a metric space Z. This expression seems daunting from the computational point of view since if we use this definition to compute $d_{\mathcal{GH}}(X,Y)$, we would have to optimize over huge spaces defining Z, f and g. We will recall an equivalent, tamer, expression in Proposition 2.2 below.

Let $\mathcal{G}$ denote the collection of all (isometry classes of) compact metric spaces. As we see below in Proposition 2.2, $\mathcal{G}$ can be made into a metric space in its own right by endowing it with the Gromov–Hausdorff metric.

Next, we state some well known properties of the Gromov–Hausdorff distance $d_{\mathcal{GH}}$ which will be essential for our 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

$$\Gamma_{X,Y}\big(x,y,x',y'\big):=\big|d_X\big(x,x'\big)-d_Y\big(y,y'\big)\big|. \tag{2.8}$$

Fig. 3. These two mm-spaces are isometric but they are not isomorphic.

Proposition 2.2. *(See [53].)*

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

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

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*

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

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

$$\frac{1}{2}\big|\mathbf{diam}(X) - \mathbf{diam}(Y)\big| \leqslant d_{\mathcal{GH}}(X, Y) \leqslant \frac{1}{2}\max\big(\mathbf{diam}(X), \mathbf{diam}(Y)\big). \tag{2.9}$$

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

$$d_{\mathcal{GH}}(X, Y) = \frac{1}{2}\inf_{R\in\mathcal{R}(X,Y)} \sup_{\substack{x_1,x_2\in X\\ y_1,y_2\in Y\\ \text{s.t. }(x_i,y_i)\in R}} \Gamma_{X,Y}(x_1, y_1, x_2, y_2) \quad \left(= \frac{1}{2}\inf_R \|\Gamma_{X,Y}\|_{L^\infty(R\times R)}\right), \tag{2.10}$$

where R *is a correspondence between* X *and* Y *as defined above.*

All these properties are desirable, see [8,9,13].

3. Gromov–Wasserstein distances

With the goal of obtaining a more computationally tractable alternative to the Gromov–Hausdorff distance, [13] proposes viewing shapes not just as a set of points with a metric on them, but also advocates in addition the use of a probability measure given on the (sets of) points. This probability measure can be thought of as signaling the importance of the different points in the dataset. The resulting structure is then a triple (X, d, ν) where X is a set, d a metric on X and ν a Borel probability measure on X. We now review the requisite background about such structures.

3.1. Measure metric spaces

Definition 3.1. (See [10].) A metric measure space (mm-space for short) is 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. $\mu_X(X) = 1$.
- μ_X has full support: $\text{supp}[\mu_X] = X$.

When it is clear from the context, we will denote the triple (X, d_X, μ_X) by only X.

Two mm-spaces (X, d_X, μ_X) and (Y, d_Y, μ_Y) are said to be *isomorphic* iff there exists an isometry $\psi : X \to Y$ such that $\mu_X(\psi^{-1}(B)) = \mu_Y(B)$ for all $B \subset Y$ measurable. Furthermore, we will denote by $\mathcal{G}_w$ the collection of all mm-spaces.[10]

Example 3.1. Consider the two mm-spaces given in Fig. 3. These are isometric (if we discard the weights attached to each point) but yet, they are non-isomorphic.

Example 3.2 *(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 $C \subset M$, $\mu_M(C) = \frac{\mathbf{vol}_M(C)}{\mathbf{Vol}(M)}$. 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 $\mathfrak{R} \subset \mathcal{G}_w$, *isomorphism reduces to isometry.*

[10] The subscript w stands for “weighted”, thus $\mathcal{G}_w$ is the collection of all weighted metric spaces.

Example 3.3. For a given $X \in \mathcal{G}_w$ and each $p \in [1, \infty)$, the p-diameter of X is defined to be $\mathbf{diam}_p(X) := \|d_X\|_{L^p(\mu_X \otimes \mu_X)}$. We define $\mathbf{diam}_\infty(X) := \mathbf{diam}(X)$. For example if $\mathbb{S}^n$ is a standard n-dimensional sphere then, $\mathbf{diam}_\infty(\mathbb{S}^n) = \pi$ for all $n \in \mathbb{N}$, $\mathbf{diam}_1(\mathbb{S}^n) = \pi/2$ for all $n \in \mathbb{N}$ but $\mathbf{diam}_2(\mathbb{S}^1) = \pi/\sqrt{3}$ and $\mathbf{diam}_2(\mathbb{S}^2) = \sqrt{\pi^2/2 - 1}$, [13]. So, one could loosely argue that taking into account the information contained in the volume measure provides more simple invariants that can be used to discriminate spaces.

3.2. Gromov–Wasserstein distances on $\mathcal{G}_w$

The Gromov–Hausdorff distance between metric spaces provides a natural measure of dissimilarity between shapes as was argued in [8,9]. In [13] it was pointed out that the GH distance leads to Quadratic Assignment Problems, which are difficult (NP-hard) combinatorial optimization problems that are typically solved using heuristics that are designed to work in restricted cases [9,62]. Further, despite the fact that the GH distance is a very powerful and general tool, it appears difficult to use in order to relate it to various existing approaches in the literature. In [13] it was pointed out that by considering shapes as mm-spaces it is possible to define a notion of distance between them, closely related to the GH distance, that (1) retains all the desirable theoretical features; (2) directly leads to continuous optimization problems; and (3) provides, via lower bounds, explicit links to a variety of shape matching techniques.

In this section we review the main features of this distance, which we call the *Gromov–Wasserstein* distance due to its relationship to *mass transportation problems*. There is not a unique notion of Gromov–Wasserstein distance [11,13]– different constructions lead to different properties. In this paper we stick to the construction proposed in [13] which we recall below.

Definition 3.2. Given two metric measure spaces (X, d_X, μ_X) and (Y, d_Y, μ_Y) we say that a measure μ on the product space $X \times Y$ is a *coupling* of μ_X and μ_Y iff

$$\mu(A \times Y) = \mu_X(A) \quad \text{and} \quad \mu\big(X \times A'\big) = \mu_Y\big(A'\big) \tag{3.11}$$

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

Definition 3.3. (See [13].) For $p \in [1, \infty]$ we define the *distance* $d_{\mathcal{GW},p}$ between two mm-spaces X and Y by

$$d_{\mathcal{GW},p}(X, Y) := \inf_{\mu \in \mathcal{M}(\mu_X, \mu_Y)} \frac{1}{2} \|\Gamma_{X,Y}\|_{L^p(\mu \otimes \mu)}. \tag{3.12}$$

Example 3.4. For finite mm-spaces X and Y, $\mathcal{M}(\mu_X, \mu_Y)$ can be regarded as the set of all matrices $M = ((m_{ij}))$ with non-negative elements such that $\sum_i m_{ij} = \mu_Y(j)$ and $\sum_j m_{ij} = \mu_X(i)$ for all $i = 1, \ldots, \#X$ and $j = 1, \ldots, \#Y$. For finite $p \geqslant 1$,

$$d_{\mathcal{GW},p}(X, Y) := \frac{1}{2}\left(\inf_M \sum_{i,j} \sum_{i',j'} \big(\Gamma_{X,Y}(x_i, y_j, x_{i'}, y_{j'})\big)^p m_{ij} m_{i'j'}\right)^{1/p},$$

which leads to solving a quadratic optimization problem with continuous variables and linear constraints [13].

Expression (3.12) defines a metric on the set of all isomorphism classes of mm-spaces. Below, we list the properties of $d_{\mathcal{GW},p}$ of similar spirit to those reported for $d_{\mathcal{GH}}$ in Proposition 2.2.

Theorem 3.1. *(See [13].) Let $p \in [1, \infty]$, then*

(a) *$d_{\mathcal{GW},p}$ defines a metric on the set of all (isomorphism classes of) mm-spaces.*
(b) *Let X and Y be two mm-spaces,*[11] *then we have*

$$d_{\mathcal{GH}}(X, Y) \leqslant d_{\mathcal{GW},\infty}(X, Y).$$

(c) *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*

$$d_{\mathcal{GW},p}(X, Y) \leqslant d^Z_{\mathcal{W},p}(\alpha, \beta).$$

(d) *Let (Z, α) be a measure space and α be a probability measure. For two measurable metrics $d, d' : Z \times Z \to \mathbb{R}^+$, consider the mm-spaces $X = (Z, d, \alpha)$ and $Y = (Z, d', \alpha)$. Then*

$$d_{\mathcal{GW},p}(X, Y) \leqslant \frac{1}{2}\big\|d - d'\big\|_{L^p(\alpha \otimes \alpha)}.$$

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

(e) *Let $\mathbb{X}_m \subset X$ be a set of m independent and identically distributed 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 $(\mathbb{X}_m, d_{X|_{\mathbb{X}_m\times\mathbb{X}_m}}, \mu_m)$, then, for μ_X-almost all $\omega \in \Omega$,*

$$(X, d_X, \mu_m) \xrightarrow{d_{\mathcal{GW},p}} (X, d_X, \mu_X) \quad \text{as } m \uparrow \infty.$$

(f) *Let $Y = \{y\}$, then*

$$d_{\mathcal{GW},p}(X, Y) = \frac{\mathbf{diam}_p(X)}{2}.$$

From this and property (a) (*triangle inequality*):

$$\frac{\mathbf{diam}_p(X) + \mathbf{diam}_p(Y)}{2} \geqslant d_{\mathcal{GW},p}(X, Y) \geqslant \left|\frac{\mathbf{diam}_p(X) - \mathbf{diam}_p(Y)}{2}\right|. \tag{3.13}$$

(h) *For any two mm-spaces X and Y, $d_{\mathcal{GW},p}(X, Y) \geqslant d_{\mathcal{GW},q}(X, Y)$ whenever $\infty \geqslant p \geqslant q \geqslant 1$.*

4. Spectral methods and the heat kernels on manifolds

On of our goals is to extend the Gromov–Wasserstein distance in order to be able to perform scale dependent comparison between shapes. An invariant which already encodes a natural notion of scale on a manifold is given by the fundamental solution of the heat equation, also known as the *heat kernel*. In this section we recall the construction of heat kernels on compact manifolds and list some of their key properties. For detailed exposition of the material presented here, we refer the reader to the excellent surveys [63–66].

Let (X, g) be any smooth connected n-dimensional Riemannian manifold. Then, in any local chart $x^1, \ldots, x^n$, the Laplace–Beltrami operator Δ_X on X is given by

$$\Delta_X := \frac{1}{\sqrt{|g|}} \sum_{i,j=1}^n \frac{\partial}{\partial x^i}\left(\sqrt{|g|}g^{i,j}\frac{\partial}{\partial x^j}\right) \tag{4.14}$$

where $|g| = \det(((g_{i,j})))$ and $((g^{i,j})) = ((g_{i,j}))^{-1}$. The *heat equation* on X is

$$\frac{\partial u}{\partial t} = \Delta_X u,$$

for $u : \mathbb{R}^+ \times X \to \mathbb{R}$. One says that a smooth function u is a *fundamental solution* of the heat equation at a point $x' \in X$ if the function u satisfies the heat equation and in addition the Dirac condition

$$u(t, \cdot) \to \delta\big(\cdot - x'\big) \quad \text{as } t \to 0^+.$$

This has to be understood in the distributional sense of course, see [63]. The *heat kernel* $k_X : \mathbb{R}^+ \times X \times X \to \mathbb{R}$ of X is defined for each $x' \in X$ as the (under our hypotheses) unique positive fundamental solution of the heat equation at x'. It turns out that the heat kernel is symmetric: for all $t > 0$, $k_X(t, x, x') = k_X(t, x', x)$, for all $x, x' \in X$.

Example 4.1. The heat kernel on $\mathbb{R}^d$ is given by

$$k_{\mathbb{R}^d}\big(t, x, x'\big) = \frac{1}{(4\pi t)^{d/2}} \exp\left(-\frac{\|x - x'\|^2}{4t}\right)$$

for all $x, x' \in \mathbb{R}^d$ and $t > 0$.

Example 4.2. (See [67].) The heat kernel on $\mathbb{H}_3$, the three-dimensional hyperbolic space is given by

$$k_{\mathbb{H}_3}\big(t, x, x'\big) = \frac{1}{(4\pi t)^{3/2}} \exp\left(-\frac{d^2_{\mathbb{H}_3}(x, x')}{4t} - t\right)\frac{d_{\mathbb{H}_3}(x, x')}{\sinh(d_{\mathbb{H}_3}(x, x'))}$$

for all $x, x' \in \mathbb{H}_3$ and $t > 0$, where $d_{\mathbb{H}_3}$ is the geodesic distance on $\mathbb{H}_3$.

Now let $X \in \mathfrak{R}$, that is X is a *compact connected Riemannian manifold without boundary*. The *eigenvalue problem* for the Laplace–Beltrami operator on X consists of finding all $\lambda \in \mathbb{R}$ for which there exists a non-trivial solution $\phi \in C^2(X)$ satisfying

$$\Delta_X \phi + \lambda \cdot \phi = 0. \tag{4.15}$$

The numbers λ sought after are referred to as *eigenvalues* of Δ_X and the vector space E_λ of solutions of the eigenvalue problem with value λ is called the *eigenspace* of λ. For each eigenspace E_λ we denote by $\mathcal{B}(E_\lambda)$ the set of all orthonormal bases of E_λ. The elements of the eigenspace are called *eigenfunctions*. All eigenfunctions are in $C^\infty(X)$. The collection of all the λ for which the eigenvalue problem admits a solution is called the *spectrum* of Δ_X. For each eigenvalue λ, the dimension of its associated eigenspace E_λ is called the *multiplicity* of λ. The standard result is that the spectrum consists of infinitely many eigenvalues

$$0 = \lambda_0 < \lambda_1 < \lambda_2 < \cdots < \lambda_k < \cdots \uparrow \infty \tag{4.16}$$

and each associated eigenspace is finite dimensional (i.e. the multiplicities of all eigenvalues are finite). Eigenspaces corresponding to different eigenvalues are orthogonal in $L^2(X)$ and $L^2(X)$ is the direct sum of all eigenspaces:

$$L^2(X) = \bigoplus_{i=0}^{\infty} E_{\lambda_i}.$$

We denote by $\mathcal{B}(X)$ the set of all corresponding orthonormal bases of $L^2(X)$:

$$\mathcal{B}(X) = \prod_{i=0}^{\infty} \mathcal{B}(E_{\lambda_i}).$$

It is a fact that for any $b = \{\zeta_i\}_{i\geqslant 0} \in \mathcal{B}(X)$, $\zeta_i(x) = \zeta_i(x')$ for all $i \geqslant 1$ implies that $x = x'$, that is, eigefunctions of the Laplace–Beltrami operator *separate* points.

Remark 4.1. Note that ζ_0 can be chosen to be equal to $(\mathbf{Vol}(X))^{-1/2}$, and since $\partial X = \emptyset$, $\int_X \zeta_i(x)\, \mathbf{vol}_X(dx) = 0$ for all $i \geqslant 1$.

The rate of growth of λ_i to infinity has been investigated by Weyl:

Remark 4.2 *(Weyl's formula).* One has the following asymptotic formula [63] for the eigenvalues of the Laplace–Beltrami operator on a closed compact and connected d-dimensional Riemannian manifold M:

$$\lambda_\ell \sim 4\pi^2 \left(\frac{\ell}{\omega_d\, \mathbf{Vol}(M)} \right)^{2/d} \quad \text{as } \ell \to \infty.$$

Here ω_d is the volume of the unit ball in $\mathbb{R}^d$.

For *compact* X, the heat kernel admits the following well known expansion:

$$k_X(t, x, x') = \sum_{i=0}^{\infty} e^{-\lambda_i t} \zeta_i(x) \zeta_i(x'), \tag{4.17}$$

where λ_i and ζ_i are the ith eigenvalue (each counted the number of times equal to its multiplicity) and the ith eigenfunction of the Laplace–Beltrami operator Δ_X, respectively, and the collection $\{\zeta_i\}_{i\in\mathbb{N}}$ is chosen to be *orthonormal*. The series converges uniformly on $[a, \infty) \times X \times X$, all $a > 0$ in the C^k topology for all $k \in \mathbb{N}$ [58, VI-§5.1].

Remark 4.3 *(Scaling).* Given a d-dimensional Riemannian manifold $(X, g) \in \mathfrak{R}$ with eigenvalues $\{\lambda_i\}$ and associated eigenfunctions $\{\zeta_i\} \in \mathcal{B}(X)$, and $a > 0$, the eigenvalues of $a \cdot X$ are $\{\lambda_i/a^2\}$ and $\{\zeta_i/a^{d/2}\} \in \mathcal{B}(a \cdot X)$ can be chosen as the associated eigenfunctions. It follows that

$$k_{aX}(t, x, x') = \frac{1}{a^d} \sum_{i=0}^{\infty} e^{-\frac{\lambda_i \cdot t}{a^2}} \zeta_i(x) \zeta_i(x') = \frac{1}{a^d} k_X(t/a^2, x, x').$$

Example 4.3. (See Circle and Tori, cf. §1.1.2 of [66].) An easy calculation shows that the eigenvalues and eigenfunctions (chosen so that they provide an orthonormal basis of $L^2(\mathbb{S}^1)$) of the Laplace–Beltrami operator on $\mathbb{S}^1$ are given by $\lambda_k = k^2$ and $\phi_k(\alpha) = \frac{1}{\sqrt{2\pi}} e^{ik\alpha}$ for $k \in \mathbb{Z}$. Then, using the expansion (4.17) we see that

$$k_{\mathbb{S}^1}(t, \alpha, \alpha') = \frac{1}{2\pi} Q(t, \alpha - \alpha'),$$

where $Q(t, z) = \sum_{k\in\mathbb{Z}} e^{ikz} e^{-k^2 t}$.

Let $a_1, a_2, \ldots, a_d > 0$ and consider the flat torus $\mathbb{T}^d = \times_{j=1}^{d}(a_k \cdot \mathbb{S}^1)$. Since in product Riemannian spaces the heat kernel can be factored [68, §1.6], using Remark 4.3 we find that

$$k_{\mathbb{T}^d}\big(t, (x_1, \ldots, x_d), (x'_1, \ldots, x'_d)\big) = \frac{1}{(2\pi)^d \prod_{j=1}^{d} a_j} Q\left(t/a_1^2, x_1 - x'_1\right) \cdot \cdots \cdot Q\left(t/a_d^2, x_d - x'_d\right)$$

for all $(x_1, \ldots, x_d), (x'_1, \ldots, x'_d) \in \mathbb{T}^d$ and $t > 0$.

Example 4.4 *(Sphere).* The heat kernel on $\mathbb{S}^2 \subset \mathbb{R}^3$ is given by [69]

$$k_{\mathbb{S}^2}\big(t, x, x'\big) = \sum_{k=0}^{\infty} \frac{2k+1}{4\pi} e^{-k(k+1)t} P_k\big(x \cdot x'\big) \quad \text{for all } x, x' \in \mathbb{S}^2 \text{ and } t > 0,$$

where P_k is the kth Legendre polynomial of order 0. For future reference, recall that $P_k(1) = 1$ for all $k \geqslant 0$.

Remark 4.4 *(On using the ℓ^2 norm for comparing spectra).* As mentioned in the introduction, based on Reuter et al. [7] one could consider the spectrum of the Laplace–Beltrami operator on a manifold as its signature and compare signatures of two manifolds by computing the ℓ^2 norm of the difference between the two sets of ordered eigenvalues. Although the practical associated signature (which uses cropped versions of the spectra) shows very good discriminative power, Remark 4.3 tells us that the resulting theoretical notion of distance would not be well defined when comparing, for example, (M, g) with $(M, a^{-2}g)$ for some $a \neq 1$. Indeed, in this case, $\|\Lambda - \Lambda'\|_{\ell^2} = |a^2 - 1| \cdot \|\Lambda\|_{\ell^2}$, where $\Lambda = \{\lambda_i\}_{i=0}^{\infty}$, $\Lambda' = \{a^2\lambda_i\}$, and λ_i is the ith eigenvalue of (M, g). Now, Weyl's expansion (Remark 4.2) guarantees that $\|\Lambda\|_{\ell^2} = \infty$ and hence the proposed distance between spectra is infinity as well. Therefore, as was observed in [7], in practice, one has to be careful to only consider a finite subset of the spectrum or by performing suitable scaling. However, from a theoretical point of view, this situation is not desirable, as it means that as one tries to use more and more eigenvalues to discriminate between two shapes, the dissimilarity measure diverges. This has to be interpreted as saying that increasing the amount of information used about the shapes actually hinders our ability to tell the shapes apart.

As a solution to Problem 1, in this paper we propose a slight change of perspective that removes these problems.

Intuitively, heat diffusion has to depend on both the geometry and the topology of the shape. A concrete way to see this is through the following property, that states that the heat kernel contains *all* the information about the shape:

Lemma 4.1. *(See Varadhan's Lemma [70].) For any $X \in \mathfrak{R}$,*

$$\lim_{t \downarrow 0}\big(-4t \ln k_X\big(t, x, x'\big)\big) = d_X^2\big(x, x'\big),$$

for all $x, x' \in X$. Here $d_X(x, x')$ is the geodesic distance between x and x' on X.

Remark 4.5. From Lemma 4.1 it is clear that two manifolds X and Y in $\mathfrak{R}$ are isometric if and only if there exists a surjective map $\phi : X \to Y$ such that $k_X(t, x, x') = k_Y(t, \phi(x), \phi(x'))$ for all $x, x' \in X$ and all $t \in \mathbb{R}^+$.

For compact manifolds, the long-term behavior of the heat-kernel is given explicitly:

$$\lim_{t \to \infty} k_X\big(t, x, x'\big) = \frac{1}{\mathbf{Vol}(X)}.$$

In other words, as t goes to infinity, the heat distribution on X converges to a constant. Alternatively, one can say that $k_X(t, x, x')$ converges to the uniform probability measure.

4.1. Heat kernel signature and heat trace

As discussed above, the heat-kernel provides a natural isometry invariant of a Riemannian manifold that encodes a certain notion of scale. Furthermore, by Varadhan's Lemma 4.1 and Remark 4.5, the heat kernel contains all the information that is necessary in order to discriminate between two different shapes $X, Y \in \mathfrak{R}$. The heat kernel will serve as the basis of the spectral version of the Gromov–Wasserstein distance that we construct in this work.

Heat kernels can be simplified into simpler, more manageable invariants that are still very informative. An example of such simplification is the diagonal of the heat kernel. This invariant is a standard tool in the study of the heat kernel itself and appears in different arguments dealing with bounds for the heat kernel [67].

Example 4.5. Note that from Examples 4.1 and 4.2, for all $t > 0$,

$$k_{\mathbb{H}_3}(x, x, t) = \frac{e^{-t}}{(4\pi t)^{3/2}} \quad \text{for all } x \in \mathbb{H}_3,$$

whereas

$$k_{\mathbb{R}^3}(x,x,t)=\frac{1}{(4\pi t)^{3/2}}\quad \text{for all } x\in\mathbb{R}^3.$$

Thus, a crude invariant such as the diagonal of the heat kernel still retains information that permits discriminating these two different geometries.

In the arenas of shape analysis and computer graphics, this invariant has recently been introduced simultaneously by Sun et al. [23] and Gebal et al. [36] who recognized it as a powerful tool for shape analysis.

Definition 4.1. We define the *Heat Kernel Signature* (HKS) hks_X of a shape $X\in\mathfrak{R}$ as the restriction of the heat kernel $k_X(t,\cdot,\cdot)$ to the diagonal times the volume of the shape. More concretely:

$$\mathrm{hks}_X : X\times\mathbb{R}^+\to\mathbb{R}^+,\qquad (x,t)\mapsto \mathbf{Vol}(X)\cdot k_X(t,x,x). \tag{4.18}$$

Using the thermodynamical interpretation of the heat kernel, then $\mathrm{hks}_X(t,x)$ measures the amount of heat that remains in x after time t of having placed a unit of heat at x. For this reason Gebal et al. [36] refer to the HKS as the *auto diffusion function*.

Sun et al. proved that two manifolds X and Y whose Laplace–Beltrami operators have no repeated eigenvalues are isometric if and only if there exists an homeomorphism between them that preserves the Heat Kernel Signatures at all points. In addition, Sun et al. point out that Bando and Urakawa [35] have shown that the set of metrics on a given compact manifold X without boundary that induce Laplace–Beltrami operators with non-repeating eigenvalues is a countable intersection of open, dense subsets of the space of all C^∞ metrics on X. This means that for generic Riemannian metrics, all the eigenvalues of the Laplacian have multiplicity one, which in turn implies that the set of Heat Kernel Signatures characterizes a generic manifold up to isometry. One of the practical advantages of the HKS over the heat kernel is that it is possible to more easily compare the signatures of two points defined on different manifolds X and Y, see Remark 8.5.

As Sun et al. observe, for small values of t, the Heat Kernel Signature of a point x is also related to the scalar curvature $s_X(x)$ of X at x [71][12]:

$$\mathrm{hks}_X(x,t)\simeq(4\pi t)^{-d/2}\sum_{i=0}^{\infty}a_i(x)t^i\quad \text{as } t\to 0^+, \tag{4.19}$$

where d is the dimension of X, and for example $a_0=\mathbf{Vol}(X)$ and $a_1(x)=\frac{\mathbf{Vol}(X)}{6}s_X(x)$. This expansion corresponds to a well-known property of the heat diffusion process, which states that heat tends to diffuse more slowly at points with positive curvature, and faster at points with negative curvature. Note that this expansion also illustrates our previous observation that the parameter t defines a natural notion of scale on the manifold: for small values of t, $\mathrm{hks}_X(x,t)$ captures differential information around x, whereas the collection

$$\{\mathrm{hks}_X(\cdot,t): X\to\mathbb{R}^+,\ t\in\mathbb{R}^+\}$$

of HKSs for all t provides (almost always) a good characterization of the manifold up to isometry (as per [23]).

We now define two additional related invariants.

Definition 4.2. For a given $X\in\mathfrak{R}$ define the *heat kernel signature distribution* (or HKSD) $\mathcal{H}_X:\mathbb{R}^+\times\mathbb{R}^+\to[0,1]$ by

$$\mathcal{H}_X(t,s)=\mu_X\{x\in X \mid \mathrm{hks}_X(x,t)\leqslant s\}. \tag{4.20}$$

In other words, for a fixed $t>0$ and $s>0$, $\mathcal{H}_X(t,s)$ gives the normalized area of the set of points on X whose HKS at scale t is below the threshold s (recall Example 3.2).

Remark 4.6. Notice that in practice the HKSDs of shapes X and Y can potentially be compared very easily by for example by computing the L^1 norm of their difference for a fixed $t>0$, since they are *both defined on the same domain* ($\mathbb{R}^+$). This feature is extremely important in practical applications since it dispenses with having to match the points in X to the points in Y. Theorem 8.1 proves a lower bound for the spectral GW distance along these lines.

To the best of our knowledge this invariant has not been previously reported in the literature.

One can further aggregate (average, actually) Heat Kernel Signatures at all points on the manifold to obtain

[12] Recall that in the case of surfaces, the scalar curvature agrees with the Gaussian curvature.

Definition 4.3. Define the *heat trace* K_X of $X \in \mathfrak{R}$ by

$$K_X(t) := \frac{1}{\mathbf{Vol}(X)} \int_X \mathrm{hks}_X(t,x)\,\mathbf{vol}_X(dx) = \sum_{i=0}^{\infty} e^{-\lambda_i t}, \tag{4.21}$$

where λ_i is, again, the ith eigenvalue of Δ_X (as in (4.16)).

That is, for each $t > 0$, the value of heat trace $K_X(t)$ is the average of the HKS $\mathrm{hks}_X(\cdot, t)$ over all of X. Notice that by our convention that we view elements of $\mathfrak{R}$ as mm-spaces with normalized volume measures (see Example 3.2), we may also write

$$K_X(t) = \int_X \mathrm{hks}_X(t,x)\mu_X(dx).$$

Example 4.6. From Examples 4.3 and 4.4 one sees that

$$\mathrm{hks}_{\mathbb{S}^1}(t,x) = K_{\mathbb{S}^1}(t) = \sum_{k\in\mathbb{Z}} e^{-k^2 t} \quad \text{for all } t>0 \text{ and } x \in \mathbb{S}^1,$$

and

$$\mathrm{hks}_{\mathbb{S}^2}(t,x) = K_{\mathbb{S}^2}(t) = \sum_{k\geqslant 0} (2k+1)e^{-k(k+1)t} \quad \text{for all } t>0 \text{ and } x \in \mathbb{S}^2.$$

Similarly to the Heat Kernel Signature, the heat trace contains a great deal of geometric information about the manifold, as can be seen, from the following well known expansion:

$$K_X(t) \simeq (4\pi t)^{-d/2} \sum_{i=0}^{\infty} u_i t^i \quad \text{as } t \to 0,$$

where $u_0 = \mathbf{Vol}(X)$, $u_1 = \frac{1}{6}\int_X s_X(x)\,\mathbf{vol}_X(dx)$. Note that for two-dimensional manifolds (surfaces) without boundary, $u_1 = \frac{1}{3}\pi\chi(X)$ by the Gauss–Bonnet theorem, where $\chi(X)$ is the Euler characteristic of X. See [7] for an application of this expansion to shape analysis.

Remark 4.7 *(Heat kernel signature distribution and heat trace).* Notice that for each $t > 0$, $\mathcal{H}_X(t,\cdot)$ induces a probability measure on the positive real line, which we denote by $d\mathcal{H}_X(t)$. Then one sees that $K_X(t) = \int_0^\infty s\,d\mathcal{H}_X(t)(ds) = \int_0^\infty (1 - \mathcal{H}_X(t,s))\,ds$. In words, for each $t > 0$, the heat trace $K_X(t)$ is the first moment of the probability measure $d\mathcal{H}_X(t)$.

We now make an elementary but important observation.

Remark 4.8 *(Heat trace and spectrum are equivalent).* By definition, the heat trace is completely specified by the spectrum. Conversely, note that each of the eigenvalues λ_i can be deduced from the heat trace $K_X(t)$. Indeed,

$$\lambda_0 = \inf\Big\{a > 0 \text{ s.t. } \lim_{t\to\infty} e^{at} K_X(t) \neq 0\Big\}$$

and the multiplicity of λ_0, $N_0 = \lim_{t\to\infty} e^{\lambda_0 t} K_X(t)$. Both λ_1 and its multiplicity can be obtained similarly by defining: $K'_X(t) = K_X(t) - N_0 e^{-t\lambda_0}$, and this process can be iterated to obtain the whole spectrum. Thus, knowledge of the Heat Trace is *equivalent* to knowledge of the spectrum, which suggests a way to formally analyze the Shape DNA of [7] from the point of view of $K_X(t)$.

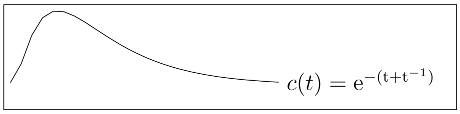
Another useful property that the heat trace shares with the HKSD, which is useful in practical applications, is that the heat traces $K_X(\cdot)$ and $K_Y(\cdot)$ of two different manifolds X and Y can be compared naturally, because they are defined over a common temporal domain, and in order to compare K_X to K_Y no spatial variables need to be matched. Having in mind the equivalence of the spectrum with the heat trace mentioned in Remark 4.8, this suggests that in order to compare the spectra of two shapes X and Y, one could compute[13] for example a quantity like $\|c(K_X - K_Y)\|_{L^\infty(\mathbb{R}^+)}$, where c is function that provides some suitable normalization and avoids blow-ups. This property will become prominent when we show that the L^∞ norm of the difference between heat traces is a lower bound of the spectral version of the Gromov–Wasserstein distance that we define in the following section, see Theorem 8.1 ahead.

[13] Instead of the ℓ^2 norm of the difference of the ordered sequences of spectra.

4.2. Diffusion distances and Rustamov's invariant

In 1994 Berard et al. [47] introduced the idea of embedding a Riemannian manifold into a Banach space via a spectral type of embedding that makes use of the heat kernel. This idea is deeply connected to the proposal of *diffusion distances* introduced in the applied literature in [17].

For each $t > 0$, let $c(t) := e^{-(t+t^{-1})}$:



Fix $X \in \mathfrak{R}$ and an orthonormal basis $b = \{\zeta_i\}_{i\in\mathbb{N}} \in \mathcal{B}(X)$ of $L^2(X)$ composed by eigenfunctions of the Laplace–Beltrami operator on X, with associated eigenfunctions $\{\lambda_i\}_{i\in\mathbb{N}}$. Consider the map

$$I_X^b : X \times \mathbb{R}^+ \to \ell^2,$$

given by

$$(x,t) \mapsto I_X^b[x](t) = \left\{c(t) \cdot \left(\mathbf{Vol}(X)\right)^{1/2} \cdot e^{-\frac{\lambda_i}{2}t}\zeta_i(x)\right\}_{i\geqslant 1}. \tag{4.22}$$

It can be seen that this map provides a continuous embedding [47, Theorem 5] and it is clear that for all $b \in \mathcal{B}(X)$, $x, x' \in X$ and $t > 0$

$$I_X^b[x](t) \bullet_{\ell^2} I_X^b[x'](t) = \left(c(t)\right)^2 \cdot \left(\mathbf{Vol}(X) \cdot k_X(t,x,x') - 1\right). \tag{4.23}$$

Hence,

$$\left\| I_X^b[x](t) - I_X^b[x'](t) \right\|_{\ell^2} = c(t) \cdot d_{X;t}^{\text{spec}}(x,x') \quad \text{for all } x, x' \in X \text{ and } t > 0, \tag{4.24}$$

where

$$d_{X;t}^{\text{spec}}(x,x') := \left(\mathbf{Vol}(X)\right)^{1/2} \cdot \left(k_X(t,x,x) + k_X(t,x',x') - 2k_X(t,x,x')\right)^{1/2}. \tag{4.25}$$

The function $d_{X;t}^{\text{spec}} : X \times X \to \mathbb{R}^+$ will be called the *diffusion distance on X at scale t* (which coincides with the definition of the diffusion distance of [17] when X is Riemannian manifold and $\mathbf{Vol}(X) = 1$).

The use of $c(t)$ in (4.24) avoids the blow up as $t \downarrow 0$ of the usual definition of the diffusion distance (as introduced in [17]). Roughly speaking, for a d-dimensional Riemannian manifold X, for $t \simeq 0^+$, the heat kernel behaves like $O(t^{-d/2})$ (and therefore it blows up). Hence $c(t) = e^{-(t^{-1}+t)}$ is "strong enough" to keep $c(t) \cdot k_X(t,\cdot,\cdot)$ bounded as t approaches 0 for any $d \in \mathbb{N}$, see Proposition 5.1 for a precise statement.

The fact that, as defined by (4.25), the diffusion distance defines a strict metric on X is not totally obvious.

Proposition 4.1 *(Properties of the diffusion distance). Given $X \in \mathfrak{R}$ with* $\dim(X) = n$, *for any $t > 0$, the diffusion distance $d_{X;t}^{\text{spec}}$ given by* (4.25) *defines a strict metric on X and for all $x, x' \in X$ one has*

$$\max_{x,x'} d_{X;t}^{\text{spec}}(x,x') \leqslant C_X \cdot t^{-n/4},$$

where C_X depends only on the dimension, the volume, the diameter, and a lower bound on the Ricci curvature of X.

We remind the reader that in Section 11, a proof will be given for any lemma, proposition, theorem or corollary for which a reference is not provided in its statement.

Remark 4.9. In [24], the authors propose to compute the GH distance between shapes X, Y which are endowed with the diffusion metric at a *fixed* scale t. In Theorem 8.2 and Remark 8.7 we show that this provides a lower bound for the spectral notion of metric between shapes that we construct in this paper.

Definition 4.4. For $X \in \mathfrak{R}$ and $p \geqslant 1$, define the *p-spectral eccentricity function* by

$$\text{ecc}_{X;p}^{\text{spec}} : \mathbb{R}^+ \times X \to \mathbb{R}^+, \qquad (t,x) \mapsto \left\| d_{X;t}^{\text{spec}}(x,\cdot) \right\|_{L^p(\mu_X)}.$$

This invariant (for $p=2$) has been used in shape analysis applications (in segmentation of shapes, to be precise) by de Goes et al. in [25]. This invariant can be derived from the HKS. For instance, it is easy to see that [23]

$$\mathrm{ecc}^{\mathrm{spec}}_{X;2}(x,t)=\big(\mathrm{hks}_X(x,t)+K_X(t)-2\big)^{1/2}.$$

Eccentricities seem to have been first introduced in shape analysis by Hilaga et al. [45], also used by [4] for shape matching.

We now define another related spectral invariant associated to any $X\in\mathfrak{R}$:

Definition 4.5. For each $t>0$ define the *distribution of diffusion distances*

$$\mathcal{G}_X(t,\cdot):[0,\infty)\to[0,1]$$

to be total (normalized) mass of pairs of points (x,x') in $X\times X$ s.t. $d^{\mathrm{spec}}_{X;t}(x,x')\leqslant s$. Precisely,

$$\mathcal{G}_X(t,s):=(\mu_X\otimes\mu_X)\big(\big\{(x,x')\;\big|\;d^{\mathrm{spec}}_{X;t}(x,x')\leqslant s\big\}\big). \tag{4.26}$$

Remark 4.10 *(The GPS embedding of Rustamov).* Similarly to the spectral embedding of Berard et al., for a given $b\in\mathcal{B}(X)$, Rustamov [20] proposes embedding $X\in\mathfrak{R}$ into ℓ^2 via the map $R^b_X:X\to\ell^2$ defined by

$$X\ni x\mapsto\left\{\frac{1}{\sqrt{\lambda_i}}\zeta_i(x)\right\}_{i\geqslant1}$$

where we denoted $b=\{\zeta_i\}_{i\in\mathbb{N}}$ and the eigenvalue corresponding to ζ_i is λ_i, for each $i\in\mathbb{N}$. His shape matching proposal is to consider a certain version of the shape distributions idea of [5] in the embedded space. Namely, for a given shape X, he proposes to compute the histogram of all distances

$$\big\{\big\|R^b_X[x]-R^b_X[x']\big\|_{\ell^2},\ x,x'\in X\big\}.$$

We will show below that a certain reformulation of the GPS+Shape distributions procedure of Rustamov can be expressed using the invariant $\mathcal{G}_X$ defined above, and that this yields a lower bound for the spectral notion of distance we construct in this paper. The proposed reinterpretation is to look at $I^b_X[x]$ instead of $R^b_X[x]$ and to use $\mathcal{G}_X$ as a proxy for the shape distributions invariant that Rustamov proposed to use. Note that this is correct since by definition of I^b_X and $d^{\mathrm{spec}}_{X;t}$,

$$\big\|I^b_X[x](t)-I^b_X[x'](t)\big\|_{\ell^2}=c(t)\cdot d^{\mathrm{spec}}_{X;t}(x,x')\quad\text{for all }x,x'\in X\text{ and }t>0.$$

Also, recall that for all $t>0$ and $x,x'\in X$, by (4.17) we have

$$I^b_X[x](t)\bullet_{\ell^2}I^b_X[x'](t)=\big(c(t)\big)^2\cdot\big(\mathbf{Vol}(X)\cdot k_X(t,x,x')-1\big)$$

which can be compared to Rustamov's motivation for the embedding he proposed: namely, the fact that for all $b\in\mathcal{B}(X)$

$$R^b_X[x]\bullet_{\ell^2}R^b_X[x']=G_X(x,x')\quad\text{for all }x,x'\in X,$$

where G_X is the *Green function* on X, see [20, Section 4]. The metric on X induced by the ℓ^2 distance between the images of the points via the embedding R^b_X is known as the *commute time distance* [72].

5. Heat kernel estimates

We need to make sure that the heat kernel invariants we have defined remain suitably bounded. We recall some standard results due to Davies [73, p. 31]: for the heat kernel on any $X\in\mathfrak{R}$:

$$k_X(t,x,x')\leqslant C\cdot\big(\mathbf{vol}_X\big(B\big(x,t^{1/2}\big)\big)\big)^{-1/2}\big(\mathbf{vol}_X\big(B\big(x',t^{1/2}\big)\big)\big)^{-1/2}\cdot e^{t/2}\cdot e^{-\frac{2}{9}d^2_X(x,x')/t},$$

for all $t>0$, $x,x'\in X$ and for some constant $C>0$. Now, assuming that X is n-dimensional and that the Ricci curvature of X is bounded below by $-\kappa^2(n-1)$ for some $\kappa>0$, by virtue of the Bishop–Gromov comparison theorem [58, p. 156] (and simple estimates) one obtains that for all $x,x'\in X$

$$k_X(t,x,x')\leqslant C\cdot e^t\cdot e^{-d^2_X(x,x')/8t}\cdot\begin{cases}1 & \text{for }t\geqslant D^2,\\ \frac{V_{n,\kappa}(D)}{V_{n,\kappa}(t^{1/2})} & \text{for }t\in(0,D^2],\end{cases} \tag{5.27}$$

where $D=\mathbf{diam}(X)$ and $V_{n,\kappa}(s)=\kappa^{1-n}\int_0^s(\sinh(\kappa u))^{n-1}\,du$.

Proposition 5.1. *For all $X\in\mathfrak{R}$ of dimension n satisfying $\mathbf{diam}(X)\leqslant D$, and $\mathbf{Ric}_X\geqslant-(n-1)\kappa^2$, and $\beta\geqslant0$ one has*

$$\sup_{t>0}c^2(t)\cdot t^{-\beta}\big\|k_X(t\cdot,\cdot)\big\|_{L^\infty(X\times X)}\leqslant C(\kappa,n)\cdot\max\big(1,\gamma(D,\kappa,n),(n/2+\beta)^{n/2+\beta}e^{-(n/2+\beta)}\big),$$

where $C(\kappa,n)>0$ depends only on κ and n, and $\gamma(D,\kappa,n)>0$ depends only on D, κ and n.

Definition 5.1. For each $p \in [1,\infty]$ we define the *p-spectral variance* of $X \in \mathfrak{R}$ as

$$\mathbf{var}_p^{\mathrm{spec}}(X) := \sup_{t>0} c^2(t) \cdot \big\| \mathbf{Vol}(X) \cdot k_X(t,\cdot,\cdot) - 1 \big\|_{L^p(\mu_X \otimes \mu_X)}.$$

Naming the quantity above spectral variance is justified by the interpretation that it measures the distance from a manifold to a point in a sense made precise by Proposition 6.1 and Remark 6.3.

Remark 5.1. Note that for all $p \in [1,\infty]$,

$$\mathbf{var}_p^{\mathrm{spec}}(X) \leqslant \mathbf{var}_\infty^{\mathrm{spec}}(X) \leqslant 1 + \mathbf{Vol}(X) \cdot \sup_{t>0} c^2(t) \cdot \big\| k_X(t,\cdot,\cdot) \big\|_{L^\infty(X \times X)}.$$

Corollary 5.1. *For any $X \in \mathfrak{R}$, $\mathbf{var}_p^{\mathrm{spec}}(X)$ remains bounded for all $p \in [1,\infty]$.*

The following proposition will also be useful for our presentation.

Proposition 5.2. *Given any $X \in \mathfrak{R}$, there exist a constant $C_X > 0$ s.t. for all $t > 0$ and $x, x' \in X$*

$$\sum_{i=1}^{\infty} e^{-\lambda_i t} \big|\zeta_i(x)\big| \big|\zeta_i(x')\big| \leqslant C_X t^{-n/2},$$

where $n = \dim(X)$, $\{\lambda_i\}_{i\geqslant 0}$ are the eigenvalues of the Laplace–Beltrami operator Δ_X on X given in increasing order, and repeated according to their multiplicity, and $\{\zeta_i\}_{i\geqslant 0}$ is an orthonormal basis of $L^2(X)$ composed of eigenfunctions of Δ_X s.t. ζ_i is associated to the eigenvalue λ_i.

The constant C_X can be shown to depend on $\dim(X)$, the volume of X, a lower bound on the Ricci curvatures of X and the diameter of X; see the proof in Section 11.

6. A spectral notion of the Gromov–Wasserstein distance

In this section we carry out an adaptation of the Gromov–Wasserstein distance to the class of compact Riemannian manifolds without boundary. A similar construction for the Gromov–Hausdorff distance, essentially due to Kasue and Kumura [48], is possible. We delay presenting those details to Section 9.1.

Notice that we could interpret Varadhan's Lemma (Lemma 4.1) as asserting that the heat kernel provides a naturally multi-scale decomposition of the geometry of a Riemannian manifold. We want to make this appear explicitly in the definition of a metric on the collection of isometry classes of $\mathfrak{R}$. For $X, Y \in \mathfrak{R}$ and $t \in \mathbb{R}^+$ define

$$\Gamma_{X,Y,t}^{\mathrm{spec}} : X \times Y \times X \times Y \to \mathbb{R}^+$$

by

$$\big(x, y, x', y'\big) \mapsto \big|\mathbf{Vol}(X) \cdot k_X\big(t, x, x'\big) - \mathbf{Vol}(Y) \cdot k_Y\big(t, y, y'\big)\big|. \tag{6.28}$$

Definition 6.1 *(Spectral Gromov–Wasserstein distance).* For $X, Y \in \mathfrak{R}$ and $p \in [1,\infty]$ let

$$d_{\mathcal{GW},p}^{\mathrm{spec}}(X,Y) := \inf_{\mu \in \mathcal{M}(\mu_X,\mu_Y)} \sup_{t>0} c^2(t) \cdot \big\| \Gamma_{X,Y,t}^{\mathrm{spec}} \big\|_{L^p(\mu\otimes\mu)},$$

where $c(t) = e^{-(t^{-1}+t)}$.[14]

Just for the sake of clarity, we write down the expanded form of the definition (for $p \in [1,\infty)$):

$$d_{\mathcal{GW},p}^{\mathrm{spec}}(X,Y) := \inf_{\mu \in \mathcal{M}(\mu_X,\mu_Y)} \sup_{t>0} c^2(t) \cdot \left(\iint_{X\times Y} \iint_{X\times Y} \big(\Gamma_{X,Y,t}^{\mathrm{spec}}\big(x,x',y,y'\big)\big)^p \mu(dx \times dy)\mu\big(dx' \times dy'\big) \right)^{1/p}.$$

Remark 6.1 *(Multi-scale aspect of the definition).* According to our definition two Riemannian manifolds will be considered to be similar in the spectral GW sense if and only if they are similar *at all scales t*. This is encoded in the definition of our metric by first taking the supremum over all $t > 0$, and then choosing the best coupling. The use of the parameter t in this fashion provides a natural *foliation* of the notion of approximate isometry between Riemannian manifolds.

[14] Compare with Definition 3.3.

In more detail, assume for example that $d^{\text{spec}}_{\mathcal{GW},\infty}(X,Y) < \eta$ for some constant $\eta > 0$. Then, there exists $\mu \in \mathcal{M}(\mu_X,\mu_Y)$ s.t.

$$\left|\mathbf{Vol}(X)\cdot k_X\big(t,x,x'\big) - \mathbf{Vol}(Y)\cdot k_Y\big(t,y,y'\big)\right| < \frac{\eta}{c^2(t)} \quad \text{for all } (x,y), \big(x',y'\big) \in \operatorname{supp}[\mu] \text{ and } \textit{for all } t > 0.$$

This should be interpreted as stating that the shapes are close to each other *at all scales*. We provide more arguments in favor of interpreting t as a sort of geometric scale in Section 7.

Theorem 6.1. *For all $p \in [1,\infty]$, $d^{\text{spec}}_{\mathcal{GW},p}$ defines a metric on the collection of all isometry classes of $\mathfrak{R}$. Moreover,*

- *for any $X,Y \in \mathfrak{R}$, $d^{\text{spec}}_{\mathcal{GW},p}(X,Y) \geqslant d^{\text{spec}}_{\mathcal{GW},q}(X,Y)$ for all $1 \leqslant q \leqslant p \leqslant \infty$.*
- *for all $X,Y \in \mathfrak{R}$ and $p \in [1,\infty]$,*

$$\left|\mathbf{var}^{\text{spec}}_p(X) - \mathbf{var}^{\text{spec}}_p(Y)\right| \leqslant d^{\text{spec}}_{\mathcal{GW},p}(X,Y) \leqslant \mathbf{var}^{\text{spec}}_p(X) + \mathbf{var}^{\text{spec}}_p(Y). \tag{6.29}$$

Compare (6.29) with (2.9) and (3.13).

Remark 6.2 *(Boundedness).* The boundedness of the definition of the spectral GW distance above follows from the boundedness of $\mathbf{var}^{\text{spec}}_p(X)$ for $X \in \mathfrak{R}$ which was proved in Section 5.

We now give an example of collapsing a "thin" torus onto $\mathbb{S}^1$ in the spectral GW sense. To provide intuition about our construction, here we shall provide all the details of the proof. Similar technical ingredients are used in the proof of the more general statement of Proposition 6.1, see Section 11.

Example 6.1 *(Collapsing).* Fix $\varepsilon > 0$ and consider the two-dimensional flat torus $\mathbb{T}_\varepsilon := \mathbb{S}^1 \times (\varepsilon \cdot \mathbb{S}^1)$. Here, as before, $\mathbb{T}_\varepsilon$ and $\mathbb{S}^1$ are endowed with uniform probability measures. We prove below that

$$d^{\text{spec}}_{\mathcal{GW},\infty}\big(\mathbb{T}_\varepsilon,\mathbb{S}^1\big) \leqslant C\cdot\varepsilon^2$$

for some constant $C > 0$ independent of ε and hence

$$d^{\text{spec}}_{\mathcal{GW},\infty}\big(\mathbb{T}_\varepsilon,\mathbb{S}^1\big) \to 0 \quad \text{as } \varepsilon \to 0.$$

Proof. By Example 4.3, we have

$$k_{\mathbb{T}_\varepsilon}\big(t,(\alpha,\beta),\big(\alpha',\beta'\big)\big) = \frac{1}{4\pi^2\varepsilon} Q\big(t,\alpha-\alpha'\big)\cdot Q\big(t/\varepsilon^2,\beta-\beta'\big),$$

for all $(\alpha,\beta),(\alpha,\beta) \in \mathbb{T}_\varepsilon$. Note that since $\mathbf{Vol}(\mathbb{T}_\varepsilon) = (2\pi)^2\cdot\varepsilon$, then

$$\Gamma^{\text{spec}}_{\mathbb{T}_\varepsilon,\mathbb{S}^1,t}\big(t,(\alpha,\beta),\theta,\big(\alpha',\beta'\big),\theta'\big) = \left|Q\big(t,\alpha-\alpha'\big)Q\big(t/\varepsilon^2,\beta-\beta'\big) - Q\big(t,\theta-\theta'\big)\right|,$$

for all $(\alpha,\beta),(\alpha',\beta') \in \mathbb{T}_\varepsilon$ and $\theta,\theta' \in \mathbb{S}^1$. In particular, for $\theta = \alpha$ and $\theta' = \alpha'$, one has:

$$\Gamma^{\text{spec}}_{\mathbb{T}_\varepsilon,\mathbb{S}^1,t}\big(t,(\alpha,\beta),\alpha,\big(\alpha',\beta'\big),\alpha'\big) = Q\big(t,\alpha-\alpha'\big)\cdot\left|Q\big(t/\varepsilon^2,\beta-\beta'\big) - 1\right|. \tag{6.30}$$

Notice that for all $t > 0$ and β,

$$\left|Q\big(t/\varepsilon^2,\beta\big) - 1\right| \leqslant 2\sum_{k\geqslant 1} e^{-k^2t/\varepsilon^2} \leqslant 2\sum_{k\geqslant 1} e^{-kt/\varepsilon^2} = \frac{2}{e^{t/\varepsilon^2}-1}.$$

Similarly, for all $t > 0$ and α,

$$Q(t,\alpha) \leqslant \frac{e^t+1}{e^t-1}.$$

Hence, for any $R \in \mathcal{R}(\mathbb{T}_\varepsilon,\mathbb{S}^1)$ s.t. $((\alpha,\beta),\gamma) \in R$ if and only if $\alpha \simeq \gamma$, taking into account that $0 \leqslant c(t) \leqslant 1$, we see from (6.30) that

$$c^2(t)\cdot\left\|\Gamma^{\text{spec}}_{\mathbb{T}_\varepsilon,\mathbb{S}^1,t}\right\|_{L^\infty(R\times R)} \leqslant c(t)\cdot\frac{e^t+1}{(e^t-1)(e^{t/\varepsilon^2}-1)} =: f(t).$$

Claim 6.1. *For all $t > 0$, $c(t)/t^2 \leqslant e^{-2}$.*

Proof. Indeed, $c(t)/t^2 \leqslant e^{-1/t}/t^2 =: h(t)$ for all $t > 0$. Simple calculus applied to h yields the claim. □

Now, since for all $t > 0$, $e^t - 1 \geqslant t$, $e^{t/\varepsilon^2} - 1 \geqslant t/\varepsilon^2$ and $\frac{c(t)}{t^2} \leqslant e^{-2}$, we have that for any fixed $\delta > 0$

$$f(t) \leqslant \begin{cases} 4(e^\delta + 1)e^{-2} \cdot \varepsilon^2 & \text{for } t \in (0, \delta), \\ \frac{e^\delta+1}{e^\delta-1}\frac{1}{e^{t/\varepsilon^2}-1} & \text{for } t \geqslant \delta. \end{cases}$$

Thus,

$$\sup_{t>0} f(t) \leqslant \max\left(4\left(e^\delta + 1\right)e^{-2} \cdot \varepsilon^2, \frac{e^\delta + 1}{e^\delta - 1} \cdot \frac{1}{e^{\delta/\varepsilon^2} - 1}\right) \leqslant C(\delta) \cdot \varepsilon^2$$

where $C(\delta) = (e^\delta + 1)\max(4e^{-2}, \delta^{-2})$. Hence,

$$\sup_{t>0} c^2(t) \cdot \left\| \Gamma^{\mathrm{spec}}_{\mathbb{T}_\varepsilon, \mathbb{S}^1, t} \right\|_{L^\infty(R \times R)} \leqslant C(\delta) \cdot \varepsilon^2$$

for any choice of $R \in \mathcal{R}(\mathbb{T}_\varepsilon, \mathbb{S}^1)$ s.t. $((\alpha, \beta), \gamma) \in R$ if and only if $\alpha = \gamma$. Finally, in order to conclude we need to construct $\mu \in \mathcal{M}(\mu_{\mathbb{T}_\varepsilon}, \mu_{\mathbb{S}^1})$ with $\mathrm{supp}[\mu] \subseteq R$. In order to do this define the map $\phi : \mathbb{T}_\varepsilon \to \mathbb{T}_\varepsilon \times \mathbb{S}^1$ by $(x, y) \mapsto (x, y, x)$ and let $\mu = \phi_\# \mu_{\mathbb{T}_\varepsilon}$. □

This generalizes in the following manner

Proposition 6.1 *(Collapsing). For all $X, Y \in \mathfrak{R}$ and $\varepsilon > 0$, the Riemannian product $X \times (\varepsilon \cdot Y)$ converges to X in the L^∞-spectral GW sense as $\varepsilon \to 0$.*

Proposition 6.2. *For all $X, Z \in \mathfrak{R}$ and $p \in [1, \infty]$,*

$$\lim_{\varepsilon \downarrow 0} d^{\mathrm{spec}}_{\mathcal{GW}, p}(X, \varepsilon \cdot Z) = \mathbf{var}^{\mathrm{spec}}_p(X).$$

Remark 6.3. Since from a metric sense, $\varepsilon \cdot Z$ approaches a point as $\varepsilon \downarrow 0$,[15] one can then interpret $\mathbf{var}^{\mathrm{spec}}_p(X)$ as the spectral GW distance from X to a point. Recall that in the context of GH distances $d_{\mathcal{GH}}(X, \{q\}) = \mathbf{diam}(X)/2$ and for GW distances $d_{\mathcal{GW},p}(X, \{q\}) = \mathbf{diam}_p(X)/2$.

7. The notion of scale

One of the theses of this work is that the time parameter t in the heat kernel can be naturally interpreted as a certain notion of scale. For example, as (4.19) shows, for small values of t, the HKS of a point x reflects differential properties of the surface at x (such as curvature), whereas $\lim_{t\to\infty} \mathrm{hks}_X(x, t) = 1$ independently of x, and thus we would say that for large scales, all points look the same.

Here we provide a point of view based on *homogenization* of partial differential equations [74] which allows to argue for the interpretation of t as a geometric notion of scale. Our argument is merely suitable for providing some intuition. Our belief is that there is still a lot of work to be done in order to elucidate good definitions and prove useful results about the role of t as a geometric scale.

Consider the real line with metric given by the periodic function $g : \mathbb{R} \to \mathbb{R}^+$ with period 1 and measure given by the one-dimensional Lebesgue measure. One can regard $M_g = (\mathbb{R}, g, \lambda)$ as a weighted Riemannian manifold [61].

Assumption 7.1. We consider metrics $g \in C^2(\mathbb{R}, \mathbb{R}^+)$ with the property that there exist $\gamma > 0$ s.t. $\gamma^{-1} \leqslant g(x) \leqslant \gamma$ for all $x \in \mathbb{R}$.

For any metric g satisfying the assumption above one can consider the resulting Laplace–Beltrami operator on M_g to be

$$\Delta_g = \frac{d}{dx}\left(\frac{1}{g(x)}\frac{d}{dx}\right).$$

Let k_g denote the heat kernel associated to M_g. The geodesic distance on M_g admits an explicit expression:

$$d_g\left(x, x'\right) = \int_{x'}^{x} g^{1/2}(s)\, ds, \quad x \geqslant x'.$$

[15] One way of being precise about this is by saying that $d_{\mathcal{GH}}(\varepsilon \cdot Z, \{q\}) \to 0$ as $\varepsilon \downarrow 0$.

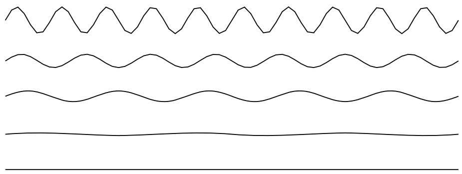
Note in particular that when $g = g_0$ constant,

$$d_{g_0}(x,x') = g_0^{1/2}|x - x'|,$$

and the heat kernel is

$$k_{g_0}(t,x,x') = \sqrt{\frac{g_0}{4\pi t}} \cdot e^{-g_0 \frac{(x-x')^2}{4t}} \quad \text{for all } x, x' \in \mathbb{R}.$$

One would expect that for $t \to \infty$, the spatial behavior of the heat kernel $k_g(t,\cdot,\cdot)$ looks like that of the heat kernel $k_{\bar{g}}(t,\cdot,\cdot)$ corresponding to a certain *constant* metric $\bar{g}$. This would be in agreement with the intuition that large values of t offer a *coarse scale view* of the underlying metric structure.



This intuition can be made precise by recalling a result due to Tetsuo Tsuchida:

Theorem 7.1. *(See [75].) Let $g : \mathbb{R} \to \mathbb{R}^+$ satisfying Assumption* 7.1 *be in addition periodic with period* 1*. Then, there exists a positive constant C such that*

$$\sup_{x,x' \in \mathbb{R}} \left|k_g(t,x,x') - k_{\bar{g}}(t,x,x')\right| \leqslant \frac{C}{t} \quad \textit{for all } t > 0,$$

where $\bar{g} = \int_0^1 g(x)\,dx$.

Remark 7.1. Note that according to [75, Corollary 1.3], heat kernels on $\mathbb{R}$ arising from metrics g satisfying Assumption 7.1 are bounded as follows: for each $\delta > 0$ there exists $A_g(\delta) > 0$ s.t. for $t \uparrow \infty$,

$$A_g(\delta) \cdot t^{-\frac{1}{2}} \leqslant k_g(t,x,x') \leqslant B_g \cdot t^{-\frac{1}{2}}, \quad \text{for all } x, x' \in \mathbb{R} \text{ s.t. } |x - x'| < \delta,$$

for some constant $B_g > 0$. Therefore, the decay rate given by Theorem 7.1 is meaningful.

Example 7.1. Pick $0 < \varepsilon \ll 1$ and $m \in \mathbb{N}$ and let $g(x) := 1 + \varepsilon \cdot \sin(2\pi m \cdot x)$. Then $\bar{g} = 1$ and Tsuchida's theorem guarantees that as t approaches infinity,

$$\left|k_g(t,x,x') - \frac{1}{4\pi t}\exp\left\{-(x-x')^2/4t\right\}\right| \to 0,$$

for all $x, x' \in \mathbb{R}$, with rate $1/t$. That is, as t grows $k_g(t,\cdot,\cdot)$ starts resembling the heat kernel corresponding to a flat one-dimensional profile.

A related intuition is that the heat kernels corresponding to two similar metrics on $\mathbb{R}$ will become very similar as t approaches infinity. In order to make this precise we appeal to a result of Davies:

Theorem 7.2. *(See [76].) Let g_1 and g_2 be two metrics on $\mathbb{R}$ satisfying the conditions given in Assumption* 7.1 *above. Then, there exists $C' > 0$ such that*

$$\sup_{x,x' \in \mathbb{R}} \left|k_{g_1}(t,x,x') - k_{g_2}(t,x,x')\right| \leqslant \frac{C'}{t} \sup_{z \in \mathbb{R}} \left| \int_0^z \big(g_1(s) - g_2(s)\big)\,ds \right|$$

for all $t > 0$.

Remark 7.2. Remark 7.1 also applies. Notice that in contrast to Tsuchida's result, Davies' does not assume that the metrics are periodic. Also, for $g_2 = g_1 + c$ for some constant c, the RHS above is infinite unless $c = 0$.

Remark 7.3. Similar bounds in the case of general Riemannian manifolds of dimension > 1 do not seem to have been established in the literature so far, with exception of [77] where the authors study the asymptotic behavior of the heat kernel corresponding to a non-flat metric on $\mathbb{R}^n$.

Using this intuition, we can argue that the parameter t can be interpreted as a notion of *geometric scale*. Much remains to be done in order to obtain precise definitions and prove general results in this direction.

8. Lower bounds for the spectral Gromov–Wasserstein distance

In the two theorems below we establish two different *hierarchies of different lower bounds* for the spectral GW distance that involve the spectral invariants that we defined in Section 4.

Theorem 8.1 *(First hierarchy). For all $X, Y \in \mathfrak{R}$ and $p \geqslant 1$,*

$$\begin{aligned} d^{\mathrm{spec}}_{\mathcal{GW},\infty}(X,Y) &\overset{(A)}{\geqslant} \sup_{t>0} c^2(t) \cdot \inf_{\mu \in \mathcal{M}(\mu_X,\mu_Y)} \left\| \Gamma^{\mathrm{spec}}_{X,Y,t} \right\|_{L^\infty(\mu)} \\ &\overset{(B)}{\geqslant} \sup_{t>0} c^2(t) \cdot \inf_{\mu \in \mathcal{M}(\mu_X,\mu_Y)} \left\| \mathrm{hks}_X(\cdot,t) - \mathrm{hks}_Y(\cdot,t) \right\|_{L^p(\mu)} \\ &\overset{(C)}{\geqslant} \sup_{t>0} c^2(t) \cdot \int_0^\infty \left| \mathcal{H}_X(t,s) - \mathcal{H}_Y(t,s) \right| ds \\ &\overset{(D)}{\geqslant} \sup_{t>0} c^2(t) \cdot \left| K_X(t) - K_Y(t) \right|. \end{aligned}$$

Example 8.1 *(Distance between $\mathbb{S}^2$ and $\mathbb{S}^1$).* As an example, we estimate the spectral GW distance between $\mathbb{S}^2$ and $\mathbb{S}^1$ by $\frac{e^{-5}}{3}$ using the lower bounds given by Theorem 8.1 (D).

From Example 4.6 we know that

$$K_{\mathbb{S}^1}(t) = 1 + 2\sum_{k \geqslant 1} e^{-k^2 t} \quad \text{and} \quad K_{\mathbb{S}^2}(t) = 1 + \sum_{k \geqslant 1} (2k+1) e^{-k(k+1)t}.$$

Hence,

$$\left| K_{\mathbb{S}^1}(t) - K_{\mathbb{S}^2}(t) \right| = \left| \sum_{k \geqslant 1} e^{-k^2 t} a_k \right|,$$

where $a_k = 2 - (2k+1)e^{-k}$. Note that $a_k \geqslant 0.8964 > 1/3$ for all $k \geqslant 1$ and hence $|K_{\mathbb{S}^1}(t) - K_{\mathbb{S}^2}(t)| \geqslant \frac{1}{3}\sum_{k \geqslant 1} e^{-k^2 t} \geqslant \frac{e^{-t}}{3}$ for $t > 0$. Then, from Theorem 8.1 (D),

$$d^{\mathrm{spec}}_{\mathcal{GW},\infty}\left(\mathbb{S}^1, \mathbb{S}^2\right) \geqslant c^2(1) \cdot \left| K_{\mathbb{S}^1}(1) - K_{\mathbb{S}^2}(1) \right| \geqslant \frac{e^{-5}}{3}.$$

Similar lower bounds for the standard Gromov–Hausdorff distance between spheres seem more difficult to establish [54].

Example 8.2 *(Distance between $\mathbb{T}^2$ and $\mathbb{S}^1$).* Similarly to the previous example we can prove that $d^{\mathrm{spec}}_{\mathcal{GW},\infty}(\mathbb{T}^2, \mathbb{S}^1) \geqslant 2e^{-5}$.

Now we make a series of remarks, some of which have practical consequences and others that propose answers to the problems posed in Section 1.

Remark 8.1. Note that in the lower bounds above the order of the $\sup_{t>0}$ and the $\inf_\mu$ are inverted with respect to the order that appears in the definition of $d^{\mathrm{spec}}_{\mathcal{GW},p}$ (Definition 6.1). This will allow us to obtain lower bounds at different scales and then consider the most discriminative scale. It is obvious that we can take the sup over a smaller, possibly finite, collection $T \subseteq \mathbb{R}^+$ of interesting and/or computable scales and we will still obtain a lower bound for $d^{\mathrm{spec}}_{\mathcal{GW},p}$.

Remark 8.2. Observe that in the context of standard GH/GW distances, a bound of the same structure as (B) would be trivial since the restriction of $\Gamma_{X,Y}$ to $\{(x,y,x,y), x \in X, y \in Y\}$ is 0.

Remark 8.3 *(A solution to Problem 1: about bound (D) and the Shape DNA).* Since knowledge of the heat trace is equivalent to knowledge of the spectrum (cf. Remark 4.8), lower bound (D) can be interpreted as a version of the Shape-DNA signature of Reuter et al. [7] that is compatible with the spectral GW distance. Furthermore, (D) establishes the quantitative stability (in the sense of Remark 1.1) of the heat trace. This was one of the goals of our project and constitutes our answer to Problem 1.

Remark 8.4 *(Quality of the lower bounds).* The question of the quality of the discrimination provided by the heat trace is of course very important from both the theoretical and the practical points of view. It is known that there exist isospectral Riemannian manifolds that are not isometric. Examples of these constructions are the tori of Milnor [34] and the spheres of

Szabo [78]. An interesting theoretical problem is that of finding non-isometric $X, Y \in \mathfrak{R}$ s.t. they have (1) the same HKS, (2) they have the same distribution of HKSs (but different HKSs), and (3) have the same heat traces (but different distributions of HKSs).

Remark 8.5 *(Computational implications).* Notice that in practical applications,[16] for each t in a given range of scales T, computing the HKS based lower bound given by Theorem 8.1 (B) involves solving a Linear Optimization Problem (a mass transportation problem actually) with $n_X \times n_Y$ variables (and $n_X + n_Y$ constraints) where n_X (resp. n_Y) is the number of vertexes in a mesh representing X (resp. Y). This may be expensive for large models. Therefore, lower bounds (C) and (D) which are based on the distribution functions associated to the heat kernel signature and the heat kernel, respectively, seem more suitable in practice.

Remark 8.6 *(A solution to Problem 3: stability of HKSs).* Lower bound (B) proves that comparing the HKSs of two shapes X and Y via solving a certain mass transportation problem provides a lower bound to the spectral GW distance. In a dual way, we see that HKS is a quantitatively stable invariant, in the spectral GW sense.

The following theorem presents another hierarchy of lower bounds for the spectral GW distance based on invariants derived from the diffusion distance, see Section 4.2.

Theorem 8.2 *(Second hierarchy). For all $X, Y \in \mathfrak{R}$,*

$$\left(d^{\mathrm{spec}}_{\mathcal{GW},\infty}(X,Y)\right)^{1/2} \overset{(A')}{\geqslant} \sup_{t>0} c(t)\cdot d_{\mathcal{GW},\infty}(X_t,Y_t) \begin{cases} \overset{(B')}{\geqslant} \sup_{t>0} c(t)\cdot \frac{1}{2}\int_0^\infty |\mathcal{G}_X(t,s)-\mathcal{G}_Y(t,s)|\,ds, \\ \overset{(C')}{\geqslant} \sup_{t>0} c(t)\cdot \frac{1}{2}\inf_\mu \|\mathrm{ecc}^{\mathrm{spec}}_{X;p}(t,\cdot)-\mathrm{ecc}^{\mathrm{spec}}_{Y;p}(t,\cdot)\|_{L^p(\mu)}, \end{cases}$$

where $X_t = (X, d^{\mathrm{spec}}_{X;t}, \mu_X)$ and $Y_t = (Y, d^{\mathrm{spec}}_{Y;t}, \mu_Y)$.

Remark 8.7 *(A solution to Problem 5: connection to using diffusion distances in the Gromov–Hausdorff/Gromov–Wasserstein distance).* Note that the lower bound (A') establishes a link with the proposal of [24]. Indeed, (A') embodies the idea of endowing X and Y with their diffusion distances at the same fixed scale $t > 0$ and then computing the Gromov–Wasserstein distance between these spaces.

In particular, it follows that for any $t > 0$,

$$\frac{1}{c^2(t)} d^{\mathrm{spec}}_{\mathcal{GW},\infty}(X,Y) \geqslant \left(d_{\mathcal{GW},\infty}(X_t,Y_t)\right)^2.$$

Now, by Theorem 3.1 (b), we obtain from the above that, for all $t > 0$,

$$\frac{1}{c(t)}\left(d^{\mathrm{spec}}_{\mathcal{GW},\infty}(X,Y)\right)^{1/2} \geqslant d_{\mathcal{GH}}(X_t,Y_t), \tag{8.31}$$

whose RHS is exactly what was (approximately) computed by Bronstein et al. in [24].

It is not known whether for a fixed $t > 0$, $d_{\mathcal{GW},\infty}(X_t,Y_t) = 0$ or $d_{\mathcal{GH}}(X_t,Y_t) = 0$ imply that X and Y are isometric (in the sense that the geodesic distances agree). Varadhan's lemma seems to suggest that this is not possible in general.

Furthermore, it would be interesting to investigate the properties of the RHS of (A') as to whether it provides a (pseudo) metric on $\mathfrak{R}$ and whether this metric would be topologically equivalent to $d^{\mathrm{spec}}_{\mathcal{GW},\infty}$.

Finally, if one accepts that X_t and Y_t provide representations of shapes X and Y at scale t, then (A') can be interpreted as a statement which reinforces the idea that in order for two shapes to be close in the spectral Gromov–Wasserstein sense, they must be close to each other at all scales.

Remark 8.8 *(A solution to Problem 2: connection of (B') to Rustamov's proposal).* Lower bound (B') in Theorem 8.2 embodies the computation of a procedure which is essentially the same as the one proposed by Rustamov, see Remark 4.10. Indeed, since $\mathcal{G}_X(t,\cdot)$ and $\mathcal{G}_Y(t,\cdot)$ are distributions of probability measures on $\mathbb{R}$, results on mass transportation on the real line [59] imply that

$$\int_0^\infty \left|\mathcal{G}_X(t,s)-\mathcal{G}_Y(t,s)\right| ds$$

[16] Provided one is able to estimate the value of the heat kernel from finitely many samples of a manifold, see Section 10.

equals the 1-Wasserstein distance between the probability measures on $\mathbb{R}$ induced by the push-forward measures $(d^{\text{spec}}_{X;t})_{\#}\mu_X \otimes \mu_X$ and $(d^{\text{spec}}_{Y,t})_{\#}\mu_Y \otimes \mu_Y$ which is one way of formalizing the notion of "comparing histograms" of the diffusion distances, see [13] as well.

Finally, (B') establishes the quantitative stability of the invariants $\mathcal{G}_X$.

Remark 8.9 *(A solution to Problem 6: (C') and stability of the diffusion distance eccentricity).* The lower bound (C') above plays the same role as the eccentricity lower bound in [13] and establishes the quantitative stability of the diffusion distance eccentricities used by [25]. Of course the lower bound (C') could be used for shape matching. The computation of the RHS leads to solving, for each $t > 0$, a continuous variable linear optimization problem with linear constraints.

9. Other spectral metrics on $\mathfrak{R}$

In this section we analyze other spectral metrics on $\mathfrak{R}$ that are possible. Some of these have arisen in theoretical works [47,48] and others are suggested by practical algorithms used to compare shapes and analyze data [27,28]. We also relate our spectral GW metric to the metrics that we analyze in this section.

9.1. A spectral Gromov–Hausdorff distance

It is also possible to define a GH type of distance, but this distance will be less useful from computational point of view. These ideas are essentially due to Kasue and Kumura [48]. One may define the spectral Gromov–Hausdorff distance $d^{\text{spec}}_{\mathcal{GH}}(X,Y)$ as the infimum over all $\varepsilon > 0$ s.t. there exists a correspondence $R \in \mathcal{R}(X,Y)$ for which

$$c^2(t) \cdot \left\| \Gamma^{\text{spec}}_{X,Y,t} \right\|_{L^\infty(R\times R)} < \varepsilon \quad \text{for all } t > 0.$$

Kasue and Kumura initiate a deep study of the topological properties of such distance and in particular, in a spirit similar to Gromov's compactness theorem [10], they identify pre-compact subclasses of $\mathfrak{R}$. The spectral GH distance, however, seems less suitable for practical considerations such as ours due the lack of explicit control on measure dependent invariants. In fact, by using measure couplings in the construction of the spectral GW distance instead of just correspondences, we gain the ability to explicitly/quantitatively control the variation of spectral invariants, cf. Remark 1.1. In the context of non-spectral distances, this observation led to proposing Gromov–Wasserstein distances as an alternative to the Gromov–Hausdorff distance [11,13].

From results in [48] one obtains that

Theorem 9.1. *$d^{\text{spec}}_{\mathcal{GH}}$ defines a metric on isometry classes of $\mathfrak{R}$.*

Finally, we should point out that from Lemma 2.1 it easily follows that

Proposition 9.1. *For all $X, Y \in \mathfrak{R}$,*

$$d^{\text{spec}}_{\mathcal{GW},\infty}(X,Y) \geqslant d^{\text{spec}}_{\mathcal{GH}}(X,Y). \tag{9.32}$$

9.2. Constructing other spectral metrics

We describe a construction of a metric on $\mathfrak{R}$ that is related to the metric of Berard et al. [47]. Let a non-empty $\mathrm{T} \subset \mathbb{R}^+$ be given. Consider the (pseudo) metric space $\mathcal{F}_{\mathrm{T}} := (C((0,\infty),\ell^2), \|\cdot\|_{\mathrm{T}})$ where $\|\alpha-\beta\|_{\mathrm{T}} := \sup_{t\in\mathrm{T}} \|\alpha(t)-\beta(t)\|_{\ell^2}$ for $\alpha, \beta \in \mathcal{F}_{\mathrm{T}}$. For $X \in \mathfrak{R}$ and $a \in \mathcal{B}(X)$ let

$$\mathcal{X}(a) := \left\{ I^a_X[x] : (0,\infty) \to \ell^2, x \in X \right\} \subset \mathcal{F}_{\mathrm{T}}.$$

Recall the definition of $I^a_X[x]$ given in Section 4.2. In words, $\mathcal{X}(a)$ contains all curves with the real parameter t mapping points in X to traces of curves in ℓ^2.

Given now $Y \in \mathfrak{R}$ and $b \in \mathcal{B}(Y)$ we similarly construct $\mathcal{Y}(b) \subset \mathcal{F}_{\mathrm{T}}$. One may conceive of computing the Hausdorff distance between $\mathcal{X}(a)$ and $\mathcal{Y}(b)$ as subsets of $\mathcal{F}_{\mathrm{T}}$ and then "matching" a to b. This is implemented by the following definition:

$$\mathcal{H}d_{\mathrm{T}}(X,Y) := \inf_{C\in\mathcal{R}(\mathcal{B}(X),\mathcal{B}(Y))} \sup_{(a,b)\in C} d^{\mathcal{F}_{\mathrm{T}}}_{\mathcal{H}}\big(\mathcal{X}(a),\mathcal{Y}(b)\big). \tag{9.33}$$

Note that when $t \in (0,\infty)$ is fixed and $\mathrm{T} = \{t\}$, $\mathcal{H}d_{\mathrm{T}}(X,Y) < \delta$ means that:

(1) for each $a \in \mathcal{B}(X)$ there exists $b \in \mathcal{B}(Y)$ s.t. the Hausdorff distance (as subsets of ℓ^2) between $I^a_X[X](t)$ and $I^b_Y[Y](t)$ is bounded above by δ, and

(2) symmetrically, for each $b \in \mathcal{B}(Y)$ there exists $a \in \mathcal{B}(X)$ s.t. the Hausdorff distance (as subsets of ℓ^2) between $I_Y^b[Y](t)$ and $I_X^a[X](t)$ is bounded above by δ.

Remarkably, in [47, Theorem 10] the authors prove that when $\mathrm{T} = \{t\}$ and *for any fixed* $t > 0$ $\mathcal{H}d_{\mathrm{T}}$ defines a metric on isometry classes of $\mathfrak{R}$.

As an immediate corollary we obtain

Theorem 9.2. *For any* $\mathbb{R}^+ \supset \mathrm{T} \neq \emptyset$ $\mathcal{H}d_{\mathrm{T}}$ *defines a metric on the isometry classes of* $\mathfrak{R}$.

We will be interested mainly in the case $\mathrm{T} = (0, \infty)$.

9.3. Variants of the metric of Berard, Besson and Gallot

A closely related definition that embodies the idea of finding the bases a in $\mathcal{B}(X)$ and b in $\mathcal{B}(Y)$ s.t. $\mathcal{X}(a)$ and $\mathcal{Y}(b)$ are closest to each other is

$$\mathcal{H}\hat{d}_{\mathrm{T}}(X,Y) := \inf_{a \in \mathcal{B}(X), b \in \mathcal{B}(Y)} d_{\mathcal{H}}^{\mathcal{F}_{\mathrm{T}}}\big(\mathcal{X}(a), \mathcal{Y}(b)\big). \tag{9.34}$$

Obviously,

$$\mathcal{H}d_{\mathrm{T}}(X,Y) \geqslant \mathcal{H}\hat{d}_{\mathrm{T}}(X,Y) \quad \text{for all } X, Y \in \mathfrak{R}. \tag{9.35}$$

Less obvious is the following result that we prove in Section 11:

Theorem 9.3. *For any* $\mathbb{R}^+ \supset \mathrm{T} \neq \emptyset$ $\mathcal{H}\hat{d}_{\mathrm{T}}$ *given by* (9.34) *defines a metric on the collection of isometry classes of* $\mathfrak{R}$.

For a given isometry $R \in \mathrm{iso}(\ell^2)$ of ℓ^2 let the action of R on $\mathcal{X}(a)$ be given by

$$R^*\mathcal{X}(a) = \big\{R I_X^a[x]\colon (0,\infty) \to \ell^2, x \in X\big\}.$$

We define

$$\mathcal{H}D_{\mathrm{T}}^{\ell^2}(X,Y) := \inf_{R \in \mathrm{iso}(\ell^2)} d_{\mathcal{H}}^{\mathcal{F}_{\mathrm{T}}}\big(\mathcal{X}(a), R^*\mathcal{Y}(b)\big). \tag{9.36}$$

It is clear that since changes of basis are isometries, (1) the definition of $\mathcal{H}D_{\mathrm{T}}^{\ell^2}$ does not depend on the choices of a and b, and (2):

$$\mathcal{H}\hat{d}_{\mathrm{T}}(X,Y) \geqslant \mathcal{H}D_{\mathrm{T}}^{\ell^2}(X,Y), \quad \text{for all } X, Y \in \mathfrak{R}. \tag{9.37}$$

For $t > 0$ we will henceforth use the notation $\mathcal{H}D_t^{\ell^2}$ for $\mathcal{H}D_{\mathrm{T}}^{\ell^2}$ when $\mathrm{T} = \{t\}$.

Remark 9.1. We do not know whether for fixed $t > 0$ $\mathcal{H}D_t^{\ell^2}$ defines a strict metric on (the isometry classes of) $\mathfrak{R}$. This is similar to the current state of knowledge about whether $d_{\mathcal{GH}}(X_t, Y_t) = 0$ for a fixed t implies that X and Y are isometric (w.r.t. geodesic distances), recall Remark 8.7. Remark 9.2 below explains that $c(t) \cdot d_{\mathcal{GH}}(X_t, Y_t) \leqslant \mathcal{H}D_t^{\ell^2}(X,Y)$.

We now let $\mathrm{T} = \mathbb{R}^+$ and relate the metrics $\mathcal{H}d_{\mathbb{R}^+}$, $\mathcal{H}\hat{d}_{\mathbb{R}^+}$ and $\mathcal{H}D_{\mathbb{R}^+}^{\ell^2}$ to $d_{\mathcal{GH}}^{\mathrm{spec}}$.

Theorem 9.4. *For all* $X, Y \in \mathfrak{R}$

$$M \cdot \mathcal{H}D_{\mathbb{R}^+}^{\ell^2}(X,Y) \geqslant d_{\mathcal{GH}}^{\mathrm{spec}}(X,Y),$$

where $M = 2(\max(\mathbf{var}_\infty^{\mathrm{spec}}(X), \mathbf{var}_\infty^{\mathrm{spec}}(Y)))^{1/2}$.

Recall (9.35) and (9.37). Since by Theorem 9.1 $d_{\mathcal{GH}}^{\mathrm{spec}}$ defines a metric on isometry classes of $\mathfrak{R}$ we obtain

Corollary 9.1. $\mathcal{H}d_{\mathbb{R}^+}$, $\mathcal{H}\hat{d}_{\mathbb{R}^+}$ *and* $\mathcal{H}D_{\mathbb{R}^+}^{\ell^2}$ *define metrics on isometry classes of* $\mathfrak{R}$.

We treat the proof of the triangle inequalities for these metrics in Section 11.

9.4. Some computationally motivated definitions

From the computational point of view, it makes sense to define for fixed $t > 0$ and each $m \in \mathbb{N}$:

$$\mathcal{H}D_t^{\mathbb{R}^m}(X,Y) := \inf_{R \in \mathrm{iso}(\mathbb{R}^m)} d_{\mathcal{H}}^{\mathbb{R}^m}\big(\Pi_m I_X^a[X](t), R\big(\Pi_m I_Y^b[Y](t)\big)\big). \tag{9.38}$$

Here, $\Pi_m : \ell^2 \to \mathbb{R}^m$ is the canonical projection $(x_1, x_2, \ldots, x_m, x_{m+1}, \ldots) \mapsto (x_1, x_2, \ldots, x_m)$.

The definition of $\mathcal{H}D_t^{\mathbb{R}^m}$ is in the same spirit as what was proposed and used by Coifman and Lafon in [41].

For each $m \in \mathbb{N}$ define the following version of the diffusion distance (cf. (4.24), (4.25)), which we call *truncated diffusion distance*:

$$d_{X;t;m}^{\mathrm{spec}}(x,x') := \frac{1}{c(t)} \big\| \Pi_m I_X^a[x](t) - \Pi_m I_X^a[x'](t) \big\|_{\mathbb{R}^m}. \tag{9.39}$$

It is not difficult to verify that $d_{X;t;m}^{\mathrm{spec}} \to d_{X;t}^{\mathrm{spec}}$ uniformly over $X \times X$ as $m \uparrow \infty$. In fact one has

Lemma 9.1. *For $X \in \mathfrak{R}$, $t > 0$ and $m \geqslant 1$,*

$$\big\| d_{X;t}^{\mathrm{spec}} - d_{X;t;m}^{\mathrm{spec}} \big\|_{L^\infty(X\times X)} \leqslant \alpha_X(t;m),$$

where $\alpha_X(t;m) \to 0$ as $m \uparrow \infty$.

One can actually prove that α_X above can be made to depend on a lower bound for the Ricci curvature, an upper bound for the diameter, the dimension and the volume of X.

Similarly to the notation used in Theorem 8.2, given $X \in \mathfrak{R}$, $t > 0$ and $m \in \mathbb{N}$ we use the notation $X_{t,m}$ for the mm-space $(X, d_{X;t;m}^{\mathrm{spec}}, \mu_X)$ where as usual μ_X is the normalized volume measure on X.

From [11, Theorem 2] we immediately have that computing $\mathcal{H}D_t^{\mathbb{R}^m}$ between shapes X and Y is comparable to computing $d_{\mathcal{GH}}(X_{t,m}, Y_{t,m})$:

Theorem 9.5. *Given $m \in \mathbb{N}$ there exists a constant c_m s.t. for all $X, Y \in \mathfrak{R}$ and $t > 0$*

$$d_{\mathcal{GH}}(X_{t,m}, Y_{t,m}) \leqslant \frac{1}{c(t)} \cdot \mathcal{H}D_t^{\mathbb{R}^m}(X,Y) \leqslant c_m\big(d_{\mathcal{GH}}(X_{t,m}, Y_{t,m})\big)^{1/2} \cdot M_t^{1/2},$$

where $M_t \leqslant \max(C_X, C_Y) \cdot t^{-\min(\dim(X),\dim(Y))/4}$ and $C_\bullet$ is the constant given by Proposition 4.1.

This theorem asserts that the results of comparing the extrinsic spectral geometry of X and Y that is perceived via the spectral embeddings I_X and I_Y (restricted to the first m coordinates) is equivalent to comparing the intrinsic spectral geometries given by the (truncated) diffusion distance. We should point out that the $1/2$ exponent for $d_{\mathcal{GH}}(X_{t,m}, Y_{t,m})$ on the RHS is necessary in general, see [11].

Remark 9.2. It is easy to verify that letting $m \uparrow \infty$ in the leftmost inequality produces $d_{\mathcal{GH}}(X_t, Y_t) \leqslant \frac{1}{c(t)} \cdot \mathcal{H}D_t^{\ell^2}(X,Y)$. However, the same trick applied to the rightmost inequality does not seem to work a priori unless one has precise knowledge of the behavior of c_m as $m \uparrow \infty$, see [11].

9.5. Related metrics derived from the Wasserstein distance

Much in the same way as we defined $\mathcal{H}d_{\mathrm{T}}$, one can define a mass transportation based distance as follows: for $\emptyset \neq \mathrm{T} \subset \mathbb{R}^+$ and each $p \in [1,\infty]$ let

$$\mathcal{W}_p d_{\mathrm{T}}(X,Y) := \inf_{C \in \mathcal{R}(\mathcal{B}(X),\mathcal{B}(Y))} \sup_{(a,b)\in C} \inf_{\mu} \big\| \big\| I_X^a[\cdot](\cdot) - I_Y^b[\cdot](\cdot) \big\|_{\mathrm{T}} \big\|_{L^p(\mu)}, \tag{9.40}$$

where $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ and we have used the shorter notation

$$\big\| \big\| I_X^a[\cdot](\cdot) - I_Y^b[\cdot](\cdot) \big\|_{\mathrm{T}} \big\|_{L^p(\mu)} = \left(\iint_{X\times Y} \sup_{t\in\mathrm{T}} \big\| I_X^a[x](t) - I_Y^b[y](t) \big\|_{\ell^2}^p \mu(dx \times dy) \right)^{1/p}.$$

One can argue that this definition is in the same spirit as the proposal of [27].

In analogy with the definitions of $\mathcal{H}\hat{d}_t$ and $\mathcal{H}\hat{D}_t^{\ell^2}$, we define for $p \in [1,\infty]$

$$\mathcal{W}_p \hat{d}_t(X,Y) := \inf_{a\in\mathcal{B}(X), b\in\mathcal{B}(Y)} \inf_{\mu} \big\| \big\| I_X^a[\cdot](\cdot) - I_Y^b[\cdot](\cdot) \big\|_{\mathrm{T}} \big\|_{L^p(\mu)}, \tag{9.41}$$

and

$$\mathcal{W}_p D_t^{\ell^2}(X,Y) := \inf_{R\in\mathrm{iso}(\ell^2)} \inf_{\mu} \big\| \big\| I_X^a[\cdot](\cdot) - R I_Y^b[\cdot](\cdot) \big\|_{\mathrm{T}} \big\|_{L^p(\mu)}. \tag{9.42}$$

By the same remarks that we used to obtain (9.35) and (9.37) we have:

$$\mathcal{W}_p d_{\mathrm{T}} \geqslant \mathcal{W}_p \hat{d}_{\mathrm{T}} \geqslant \mathcal{W}_p D_{\mathrm{T}}^{\ell^2}. \tag{9.43}$$

Similarly to Theorem 9.4, we now let $\mathrm{T} = \mathbb{R}^+$ and relate the metrics $\mathcal{W}_p d_{\mathbb{R}^+}$, $\mathcal{W}_p \hat{d}_{\mathbb{R}^+}$ and $\mathcal{W}_p D_{\mathbb{R}^+}^{\ell^2}$ to $d_{\mathcal{GW},p}^{\mathrm{spec}}$.

Theorem 9.6. *For all $X, Y \in \mathfrak{R}$ and $p \in [1,\infty]$,*

$$M \cdot \mathcal{W}_p D_{\mathbb{R}^+}^{\ell^2}(X,Y) \geqslant d_{\mathcal{GW},p}^{\mathrm{spec}}(X,Y),$$

where $M = 2(\max(\mathbf{var}_\infty^{\mathrm{spec}}(X), \mathbf{var}_\infty^{\mathrm{spec}}(Y)))^{1/2}$.

Recall (9.43). Since by Theorem 6.1 $d_{\mathcal{GW},p}^{\mathrm{spec}}$ defines a metric on isometry classes of $\mathfrak{R}$ we obtain:

Corollary 9.2. *For all $p \in [1,\infty]$, $\mathcal{W}_p d_{\mathbb{R}^+}$, $\mathcal{W}_p \hat{d}_{\mathbb{R}^+}$ and $\mathcal{W}_p D_{\mathbb{R}^+}^{\ell^2}$ define metrics on isometry classes of $\mathfrak{R}$.*

We treat the proof of the triangle inequalities for these metrics in Section 11.

Remark 9.3 *(A solution to Problem 4).* The definitions of $\mathcal{H}_p d_{\mathbb{R}^+}$, $\mathcal{H}_p \hat{d}_{\mathbb{R}^+}$ and $\mathcal{H}_p D_{\mathbb{R}^+}^{\ell^2}$, and $\mathcal{W}_p d_{\mathbb{R}^+}$, $\mathcal{W}_p \hat{d}_{\mathbb{R}^+}$ and $\mathcal{W}_p D_{\mathbb{R}^+}^{\ell^2}$, together with Theorems 9.4 and 9.6, and Corollaries 9.1 and 9.2 provide an answer to Problem 4, cf. [17,27,28,41].

10. Discussion

Our proposal hinges on a specialization of the original Gromov–Wasserstein notion of distance between mm-spaces which we call *spectral Gromov–Wasserstein distance*. This distance incorporates spectral information directly via the use of heat kernels. We proved that our proposed metric satisfies all the properties of a metric on the collection $\mathfrak{R}$ of compact Riemannian manifolds without boundary. The extension to more general classes of shapes, starting with an extension to manifolds with boundary and to manifolds satisfying less stringent regularity assumptions (as in [43]), probably deserves further investigation.

The fact that the GW-spectral metric encodes scale in a natural way makes it suitable for multi-scale matching of datasets and shapes. Besides the lower bounds we have presented, others may be possible.

An important topic seems to be that of obtaining a much deeper understanding of the notion of geometric scale provided by the heat kernel parameter t already present in the works of [17,23,25,28,79].

Also of interest is the understanding of the relationship between the two hierarchies of lower bounds we have established in Section 8 and to deepen the understanding of the inter-relation between the different spectral metrics on $\mathfrak{R}$ that we discussed in Section 9. A related question is that made explicit by Remark 8.4, namely, finding different counterexamples that help understand the relative strength of the different invariants.

A topic that we did not explore in this note but deserves investigation is the equivalence of the Gromov–Wasserstein and spectral Gromov–Wasserstein metrics on subclasses of $\mathfrak{R}$, cf. Section 1.5.

10.1. Discrete approximations of heat kernels

In our opinion, one of the most important questions that need to be answered in connection with our work is that of constructing provably convergent end efficient approximations for the heat kernel on triangulated or point clouds surfaces. In most applications, the underlying manifold is not known, and instead, one is often given an approximation of the shape represented by a mesh. Assume that $\mathbb{X}$ is a triangulated surface which is a good approximation to $X \in \mathfrak{R}$, in some suitable sense. One family of approaches [23] for estimating the heat kernel on X rely on the spectral expansion (4.17): one computes approximately the eigenvalues and eigenfunctions of a discrete Laplace–Beltrami operator defined on $\mathbb{X}$ and then plugs these into (4.17), see [23] for example. Many schemes have been proposed to estimate the Laplace–Beltrami operator from discrete meshes [18,80–82]. Perhaps, the most commonly used method in the computer graphics community is the so-called cotangent scheme. It is not clear whether one can prove convergence of this type of approximation to the true underlying heat kernel.

A completely different and very promising idea is mentioned in [41, §3.4 and §5] which is related to the Smolyanov–Weizsaecker approach, see [83] and references therein.

The finite approximation of heat kernels has a bearing on the discrete side of the construction we have proposed in this article, since *sampling bounds* for the spectral GW distance are missing from the current formulation.

11. Proofs

Lemma 11.1. *For all* $\alpha > 0$, $\sup_{t>0} c(t) \cdot t^{-\alpha} \leqslant (e/\alpha)^{-\alpha}$.

Proof of Proposition 4.1. Non-negativity, the triangle inequality and symmetry for $d_{X;t}^{\mathrm{spec}}$ are easy to establish from (4.25). That for fixed $t > 0$ $d_{X;t}^{\mathrm{spec}}(x,x') = 0$ implies that $x = x'$ follows form a simple argument involving the alternative expression (4.24) for the diffusion distance. Let $\{\zeta_i\}_{i\geqslant 0} \in \mathcal{B}(X)$ with associated eigenvalues $\{\lambda_i\}_{i\geqslant 0}$. Then, from (4.24) we see that $d_{X;t}^{\mathrm{spec}}(x,x') = 0$ implies that

$$\sum_{i\geqslant 1} e^{-\lambda_i t}\big(\zeta_i(x) - \zeta\big(x'\big)\big)^2 = 0.$$

Thus, $\zeta_i(x) = \zeta_i(x')$ for all $i \geqslant 1$, and since the collection $\{\zeta_i\}_{i\geqslant 1}$ separates points in X, we conclude that $x = x'$.

The bound for $\max_{x,x'} d_{X;t}^{\mathrm{spec}}(x,x')$ follows from recalling (4.25), the spectral expansion of the heat kernel (4.17), and a direct application of Proposition 5.2. □

Proof of Proposition 5.1. By (5.27), since $e^{-d_X^2(x,x')/8t} \leqslant 1$ and $e^{-t^{-1}} \leqslant 1$ for all $x, x' \in X$ and $t > 0$,

$$c(t) \cdot k_X\big(t,x,x'\big) \leqslant C \cdot \begin{cases} 1 & \text{for } t \geqslant D^2, \\ V_{n,\kappa}(D) \cdot \frac{e^{-t^{-1}}}{V_{n,\kappa}(t^{1/2})} & \text{for } t \in (0, D^2]. \end{cases} \tag{11.44}$$

Now, for all $0 \leqslant u \leqslant D$, $\sinh(u) = \frac{e^u - e^{-u}}{2} = \frac{e^{2u}-1}{2e^u} \geqslant \frac{u}{e^D}$. Hence, for any $0 \leqslant r \leqslant D$,

$$\int_0^r \big(\sinh(\kappa s)\big)^{n-1}\,ds \geqslant \kappa^{n-1} e^{-D\kappa(n-1)}\frac{r^n}{n}.$$

It then follows from the definition of $V_{n,\kappa}$ (see Section 5) that for all $0 \leqslant t \leqslant D^2$

$$V_{n,\kappa}(\sqrt{t}\,) \geqslant \gamma(D,\kappa,n) \cdot t^{n/2},$$

where $\gamma(D,\kappa,n) = \frac{V_{n,\kappa}(D)}{n} \cdot e^{-D\kappa(n-1)}$. Now, from (11.44), for all $x, x' \in X$ and $t > 0$

$$c(t) \cdot k_X\big(t,x,x'\big) \leqslant C \cdot \begin{cases} 1 & \text{for } t \geqslant D^2, \\ \gamma(D,\kappa,n) \cdot e^{-t^{-1}} t^{-n/2} & \text{for } t \in (0, D^2], \end{cases}$$

which implies that

$$\sup_{t>0} c(t) \cdot t^{-\beta} \cdot \big\|k_X(t,\cdot,\cdot)\big\|_{L^\infty(X\times X)} \leqslant C \cdot \max\Big(D^{-2\beta}, \gamma(D,\kappa,n) \cdot \sup_{t>0} t^{-(n/2+\beta)} e^{-t^{-1}}\Big).$$

Simple calculus now yields that $h(t) = t^{-(n/2+\beta)}e^{-t^{-1}} \leqslant (n/2+\beta)^{n/2+\beta} e^{-(n/2+\beta)}$ for all $t > 0$ thus completing the proof. □

Proof of Proposition 5.2. From [47, Theorem 3(iii)], for all $x \in X$ and $t > 0$,

$$\sum_{j\geqslant 1} e^{-\lambda_j t}\zeta_j^2(x) \leqslant \frac{C(n,\kappa,D)}{\mathbf{Vol}(X)} \cdot t^{-n/2}, \tag{11.45}$$

where $D = \mathbf{diam}(X)$, $\mathbf{Ric}_X \geqslant (n-1)\kappa$ and $C(\cdot,\cdot,\cdot)$ depends only on the listed geometric invariants of X. For $x, x' \in X$ and $t > 0$, by the Cauchy–Schwartz inequality,

$$\sum_{j\geqslant 1} e^{-\lambda_j t}\big|\zeta_j(x)\big| \cdot \big|\zeta_j\big(x'\big)\big| \leqslant \Big(\sum_{j\geqslant 1} e^{-\lambda_j t}\zeta_j^2(x)\Big)^{1/2} \Big(\sum_{j\geqslant 1} e^{-\lambda_j t}\zeta_j^2\big(x'\big)\Big)^{1/2}.$$

The claim follows from (11.45). □

Proof of Theorem 6.1. The proof of the triangle inequality is easy. Let $X, Y, Z \in \mathfrak{R}$ be s.t. $d_{\mathcal{GW},p}^{\mathrm{spec}}(X,Y) < \varepsilon_1$ and $d_{\mathcal{GW},p}^{\mathrm{spec}}(Y,Z) < \varepsilon_2$. Let $\mu_1 \in \mathcal{M}(\mu_X, \mu_Y)$ and $\mu_2 \in \mathcal{M}(\mu_X, \mu_Y)$ be s.t. $c^2(t) \cdot \|\Gamma_{X,Y,t}^{\mathrm{spec}}\|_{L^p(\mu_1\otimes\mu_1)} < \varepsilon_1$ and $c^2(t) \cdot \|\Gamma_{Y,Z,t}^{\mathrm{spec}}\|_{L^p(\mu_2\otimes\mu_2)} < \varepsilon_2$ for all $t \in \mathbb{R}^+$. For fixed $t \in \mathbb{R}^+$, by the triangle inequality for the absolute value:

$$\left|\mathbf{Vol}(X)\cdot k_X\big(t,x,x'\big)-\mathbf{Vol}(Y)\cdot k_Y\big(t,y,y'\big)\right| \leqslant \left|\mathbf{Vol}(X)\cdot k_X\big(t,x,x'\big)-\mathbf{Vol}(Z)\cdot k_Z\big(t,z,z'\big)\right| + \left|\mathbf{Vol}(Z)\cdot k_Z\big(t,z,z'\big)-\mathbf{Vol}(Y)\cdot k_Y\big(t,y,y'\big)\right|$$

for all $x,x'\in X$, $y,y'\in Y$ and $z,z'\in Z$, and hence

$$\Gamma^{\mathrm{spec}}_{X,Y,t}\big(x,y,x',y'\big)\leqslant \Gamma^{\mathrm{spec}}_{X,Z,t}\big(x,z,x',z'\big)+\Gamma^{\mathrm{spec}}_{Z,Y,t}\big(z,y,z',y'\big), \tag{11.46}$$

for all $x,x'\in X$, $y,y'\in Y$ and $z,z'\in Z$. Now, by the Glueing Lemma [59, Lemma 7.6], there exists 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 (11.46), we obtain

$$\begin{aligned}\|\Gamma_{X,Y,t}\|_{L^p(\mu_3\otimes\mu_3)} &= \|\Gamma_{X,Y,t}\|_{L^p(\mu\otimes\mu)}\\ &\leqslant \|\Gamma_{X,Z,t}+\Gamma_{Z,Y,t}\|_{L^p(\mu\otimes\mu)}\\ &\leqslant \|\Gamma_{X,Z,t}\|_{L^p(\mu\otimes\mu)}+\|\Gamma_{Z,Y,t}\|_{L^p(\mu\otimes\mu)}\\ &= \|\Gamma_{X,Z,t}\|_{L^p(\mu_1\otimes\mu_1)}+\|\Gamma_{Z,Y,t}\|_{L^p(\mu_2\otimes\mu_2)}\\ &\leqslant (\varepsilon_1+\varepsilon_2)/c(t).\end{aligned}$$

Hence, $d^{\mathrm{spec}}_{\mathcal{GW},p}(X,Y)\leqslant \sup_{t>0}c^2(t)\cdot\|\Gamma^{\mathrm{spec}}_{X,Y,t}\|_{L^p(\mu_3\otimes\mu_3)}<\varepsilon_1+\varepsilon_2$. The conclusion now follows by taking $\varepsilon_1\to d^{\mathrm{spec}}_{\mathcal{GW},p}(X,Y)$ and $\varepsilon_2\to d^{\mathrm{spec}}_{\mathcal{GW},p}(Y,Z)$.

We prove that $d^{\mathrm{spec}}_{\mathcal{GW},p}(X,Y)=0$ implies that X and Y are isometric. Assume first that $p\in[1,\infty)$. Let $(\varepsilon_n)\subset\mathbb{R}^+$ be s.t. $\varepsilon_n\to 0$ and $(\mu_n)\subset\mathcal{M}(\mu_X,\mu_Y)$ be s.t.

$$\big\|\Gamma^{\mathrm{spec}}_{X,Y,t}\big\|_{L^p(\mu_n\otimes\mu_n)}<\varepsilon_n/c^2(t)\quad\text{for all } n\in\mathbb{N}\text{ and } t\in\mathbb{R}^+. \tag{11.47}$$

Since $\mathcal{M}(\mu_X,\mu_Y)$ is compact for the weak topology on $\mathcal{P}(X\times Y)$ (see [59, pp. 49]), we can assume that up to extraction of a sub-sequence, μ_n converges to some $\mu_0\in\mathcal{M}(\mu_X,\mu_Y)$. We assume w.l.o.g. that $\mu_n\to\mu_0$ weakly. Then, $\mu_n\otimes\mu_n\to\mu_0\otimes\mu_0$ weakly as well. Since for fixed $t\in\mathbb{R}^+$, $\Gamma^{\mathrm{spec}}_{X,Y,t}$ is continuous on $X\times Y\times X\times Y$ and hence bounded (since we are considering only compact manifolds), one has that

$$\iint\limits_{X\times Y}\iint\limits_{X\times Y}\big(\Gamma^{\mathrm{spec}}_{X,Y,t}\big)^p\,d\mu_n\otimes\mu_n\to\iint\limits_{X\times Y}\iint\limits_{X\times Y}\big(\Gamma^{\mathrm{spec}}_{X,Y,t}\big)^p\,d\mu_0\otimes\mu_0$$

as $n\uparrow\infty$. By (11.47) we obtain $\|\Gamma^{\mathrm{spec}}_{X,Y,t}\|_{L^p(\mu_0\otimes\mu_0)}=0$ that for all $t>0$. It follows that $\Gamma^{\mathrm{spec}}_{X,Y,t}(x,y,x',y')=0$ for all $(x,y),(x',y')\in R(\mu_0)$ which is equivalent to

$$\mathbf{Vol}(X)\cdot k_X\big(t,x,x'\big)=\mathbf{Vol}(Y)\cdot k_Y\big(t,y,y'\big),$$

for all $(x,y),(x',y')\in R(\mu_0)$. By Lemma 2.1, $R(\mu_0)\in\mathcal{R}(X,Y)$. Consider a map $\phi:X\to Y$ s.t. $(x,\phi(x))\in R(\mu_0)$ for all $x\in X$. Then by the above we find that $\mathbf{Vol}(X)\cdot k_X(t,x,x')=\mathbf{Vol}(Y)\cdot k_Y(t,\phi(x),\phi(x'))$ for all $t>0$ and $x,x'\in X$. By Lemma 4.1, it follows then that $d_X(x,x')=d_Y(\phi(x),\phi(x'))$ for all $x,x'\in X$. Similarly, we can find a map $\psi:Y\to X$ s.t. $d_Y(y,y')=d_X(\psi(y),\psi(y'))$ for all $y,y'\in Y$. It follows that $\zeta:=\phi\circ\psi$ is a distance preserving map from Y into itself and since Y is compact, ζ has to be surjective, see [53, Theorem 1.6.14]. It follows that ϕ (and also ψ) is an isometry.

Now, for the case $p=\infty$, pick $(\varepsilon_n)\subset\mathbb{R}^+$ s.t. $\varepsilon_n\to 0$ and let $(\mu_n)\subset\mathcal{M}(\mu_X,\mu_Y)$ be s.t. $\|\Gamma^{\mathrm{spec}}_{X,Y,t}\|_{L^\infty(\mu_n\otimes\mu_n)}<\varepsilon_n/c^2(t)$ for all $n\in\mathbb{N}$ and $t\in\mathbb{R}^+$. Then, by Remark 2.1, (11.47) holds as well for finite p and the argument above applies.

We delay the proof of (6.29) to after the proof of Proposition 6.2. □

Proof of Proposition 6.1. The proof generalizes the argument given in Example 6.1. Let $\{\lambda_i\}_{i=0}^\infty$ denote the spectrum of Δ_Y and $\{\varphi_i\}_{i=0}^\infty$ an orthonormal basis of $L^2(Y)$ composed by eigenfunctions of Δ_Y, s.t. φ_i is the eigenfunctions corresponding to λ_i, each $i\geqslant 0$. Let $m=\dim(Y)$ and let $Z_\varepsilon=X\times(\varepsilon\cdot Y)$. Notice that (recall Remark 4.3, Example 4.3 and the fact that the heat kernel in product spaces factors) for all $t>0$ and $(x,y)\in X\times Y$ one has

$$k_{Z_\varepsilon}\big(t,(x,y),\big(x',y'\big)\big)=k_X\big(t,x,x'\big)\cdot\varepsilon^{-m}k_Y\big(t/\varepsilon^2,y,y'\big).$$

Since $\mathbf{Vol}(Z_\varepsilon)=\mathbf{Vol}(X)\,\mathbf{Vol}(Y)\varepsilon^m$, then, for all $t>0$, $x,x'\in X$ and $y,y'\in Y$:

$$\begin{aligned}c^2(t)\cdot\Gamma^{\mathrm{spec}}_{X,Z_\varepsilon,t}\big(x,(x,y),x',\big(x',y'\big)\big) &= c^2(t)\cdot\mathbf{Vol}(X)k_X\big(t,x,x'\big)\left|1-\mathbf{Vol}(Y)k_Y\big(t/\varepsilon^2,y,y'\big)\right|\\ &\leqslant c^2(t)\cdot\mathbf{Vol}(X)\,\mathbf{Vol}(Y)\cdot k_X\big(t,x,x'\big)\sum_{j\geqslant 1}e^{-\lambda_j t/\varepsilon^2}\left|\varphi_j(y)\right|\left|\varphi_j\big(y'\big)\right|\end{aligned}$$

$$\leqslant c^2(t)\cdot \mathbf{Vol}(X)k_X\big(t,x,x'\big)\cdot C(Y)\cdot t^{-m}\varepsilon^{2m} \quad \text{(by Proposition 5.2)}$$
$$\leqslant \mathbf{Vol}(X)C(Y)\cdot\varepsilon^{2m}\cdot\sup_{t>0}\big(c(t)t^{-m}\big\|k_X(t,\cdot,\cdot)\big\|_{L^\infty(X\times X)}\big)$$
$$\leqslant C(X,Y)\cdot\varepsilon^{2m} \quad \text{(by Proposition 5.1)},$$

where $C(X,Y)$ does not depend on ε. Denote by μ_{Z_ε} the normalized volume measure on Z_ε. Then (check p. 371),

$$\mu_{Z_\varepsilon}=\frac{\mathbf{vol}_{Z_\varepsilon}}{\mathbf{Vol}(Z_\varepsilon)}=\frac{\mathbf{vol}_X\otimes\varepsilon^m\,\mathbf{vol}_Y}{\varepsilon^m\,\mathbf{Vol}(X)\,\mathbf{Vol}(Y)}=\mu_X\otimes\mu_Y.$$

Let $\Phi: X\times Y\to X\times X\times Y$ be given by $(x,y)\mapsto(x,x,y)$ and let $\mu=\Phi_\#\mu_{Z_\varepsilon}$. Clearly, $\mu\in\mathcal{M}(\mu_X,\mu_{Z_\varepsilon})$ and $\mathrm{supp}[\mu]\subset\{(x,x,y),x\in X,y\in Y\}$. Hence, for any $p\in[1,\infty]$:

$$\sup_{t>0}c^2(t)\cdot\big\|\Gamma^{\mathrm{spec}}_{X,Z_\varepsilon,t}\big\|_{L^p(\mu\otimes\mu)}\leqslant\varepsilon^{2m}\cdot C(X,Y).$$

The conclusion follows by (1) noting that the LHS is an upper bound for $d^{\mathrm{spec}}_{\mathcal{GW},p}(X,Z_\varepsilon)$, and (2) letting $\varepsilon\downarrow 0$. □

Proof of Proposition 6.2. Let $\{\lambda_i\}_{i\in\mathbb{N}}$ and $\{\phi_i\}_{i\in\mathbb{N}}$ denote the eigenvalues and corresponding eigenfunctions of Δ_Z chosen so as to form an orthonormal basis of $L^2(X)$. Now, by the triangle inequality for the absolute value we see that

$$\begin{aligned}\Big|\Gamma^{\mathrm{spec}}_{X,\varepsilon\cdot Z;t}\big(x,z,x',z'\big)-\big|\mathbf{Vol}(X)\cdot k_X\big(t,x,x'\big)-1\big|\Big| &\leqslant\big|\mathbf{Vol}(\varepsilon\cdot Z)\cdot k_{\varepsilon\cdot Z}\big(t,x,x'\big)-1\big|\\ &=\left|\sum_{i=0}^{\infty}e^{\lambda_i t/\varepsilon^2}\phi_i(z)\phi_i\big(z'\big)-1\right| \quad \text{(by Remark 4.3)}\\ &=\left|\sum_{i=1}^{\infty}e^{-\lambda_i t/\varepsilon^2}\phi_i(z)\phi_i\big(z'\big)\right|\\ &\leqslant\sum_{i=1}^{\infty}e^{-\lambda_i t/\varepsilon^2}\big|\phi_i(z)\big|\big|\phi_i\big(z'\big)\big|\\ &\leqslant C\cdot t^{-n}\cdot\varepsilon^{2n} \quad \text{(by Proposition 5.2)},\end{aligned}$$

where C depends only on Z. Now fix any $\mu\in\mathcal{M}(\mu_X,\mu_Z)$ and notice that

$$\big\|\mathbf{Vol}(X)\cdot k_X(t,\cdot,\cdot)-1\big\|_{L^p(\mu\otimes\mu)}=\big\|\mathbf{Vol}(X)\cdot k_X(t,\cdot,\cdot)-1\big\|_{L^p(\mu_X\otimes\mu_X)},$$

and hence

$$\sup_{t>0}c^2(t)\cdot\big\|\mathbf{Vol}(X)\cdot k_X(t,\cdot,\cdot)-1\big\|_{L^p(\mu\otimes\mu)}=\mathbf{var}^{\mathrm{spec}}_p(X).$$

Using all the above, plus the fact that $|\sup f-\sup g|\leqslant\sup|f-g|$ for real valued functions f and g, and the triangle inequality for the norm $L^p(\mu\otimes\mu)$ we have that

$$\begin{aligned}\Big|\sup_{t>0}c^2(t)\cdot\big\|\Gamma^{\mathrm{spec}}_{X,\varepsilon\cdot Z;t}\big\|_{L^p(\mu\otimes\mu)}-\mathbf{var}^{\mathrm{spec}}_p(X)\Big| &\leqslant\sup_{t>0}c^2(t)\cdot\Big|\big\|\Gamma^{\mathrm{spec}}_{X,\varepsilon\cdot Z;t}\big\|_{L^p(\mu\otimes\mu)}-\big\|\mathbf{Vol}(X)\cdot k_X(t,\cdot,\cdot)-1\big\|_{L^p(\mu\otimes\mu)}\Big|\\ &\leqslant\sup_{t>0}c^2(t)\cdot\Big\|\Gamma^{\mathrm{spec}}_{X,\varepsilon\cdot Z;t}-\big|\mathbf{Vol}(X)\cdot k_X(t,\cdot,\cdot)-1\big|\Big\|_{L^p(\mu\otimes\mu)}\\ &\leqslant C\cdot\Big(\sup_{t>0}c(t)\cdot t^{-n}\Big)\cdot\varepsilon^{2n}\\ &=C'(n)\cdot\varepsilon^{2n},\end{aligned}$$

for some constant $C'(n)>0$, by Lemma 11.1. From the arbitrariness of μ it then follows that

$$\big|d^{\mathrm{spec}}_{\mathcal{GW},p}(X,\varepsilon\cdot Z)-\mathbf{var}^{\mathrm{spec}}_p(X)\big|\leqslant C'(n)\cdot\varepsilon^{2n}.$$

Whence the claim. □

Continuation of the proof of Theorem 6.1. Choose any $Z\in\mathfrak{R}$ and $\varepsilon>0$. Then, by the triangle inequality for the spectral GW distance,

$$d^{\mathrm{spec}}_{\mathcal{GW},p}(X,\varepsilon\cdot Z)\leqslant d^{\mathrm{spec}}_{\mathcal{GW},p}(X,Y)+d^{\mathrm{spec}}_{\mathcal{GW},p}(Y,\varepsilon\cdot Z)$$

and

$$d_{\mathcal{GW},p}^{\mathrm{spec}}(X,Y) \leqslant d_{\mathcal{GW},p}^{\mathrm{spec}}(X,\varepsilon\cdot Z) + d_{\mathcal{GW},p}^{\mathrm{spec}}(Y,\varepsilon\cdot Z).$$

The conclusion follows from a direct application of Proposition 6.2. □

For the proofs of Theorems 8.1 and 8.2 we need a technical lemma:

Lemma 11.2. *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. Then*

$$\inf_{\mu}\int_{X\times Y}\big|f(x)-g(y)\big|\mu(dx,dy) \geqslant \int_{\mathbb{R}}\big|F(u)-G(u)\big|\,du,$$

where $\mu\in\mathcal{M}(\mu_X,\mu_Y)$, $F(s):=\mu_X\{x\in X|f(x)\leqslant s\}$ and $G(s):=\mu_Y\{y\in Y|g(y)\leqslant s\}$ are the distributions of f and g under μ_X and μ_Y, respectively.

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 [56] (applied to $T: X\times Y\to\mathbb{R}\times\mathbb{R}$, $(x,y)\mapsto(f(x),g(y))$) one has

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

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,

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

The conclusion follows from results on the transportation problem on the real line, see [59, Remark 2.19], and then from the fact that $\mu\in\mathcal{M}(\mu_X,\mu_Y)$ was arbitrary and the RHS does not depend on it. □

Proof of Theorem 8.1.

(A) Assume $d_{\mathcal{GW},\infty}^{\mathrm{spec}}(X,Y)<\eta$ and let $\mu\in\mathcal{M}(\mu_X,\mu_Y)$ be s.t. $\|\Gamma_{X,Y,t}\|_{L^\infty(R(\mu)\times R(\mu))}\leqslant\eta/c^2(t)$ for all $t>0$. This means that $|\mathbf{Vol}(X)\cdot k_X(t,x,x')-\mathbf{Vol}(Y)\cdot k_Y(t,y,y')|\leqslant\eta/c^2(t)$ for all $(x,y),(x',y')\in R(\mu)$ and $t>0$. Then, for any $t>0$,

$$c^2(t)\cdot\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\|\Gamma_{X,Y,t}\|_{L^\infty(\mu\otimes\mu)}\leqslant\eta.$$

Now, since the RHS does not depend on t, we obtain

$$\sup_{t>0}c^2(t)\cdot\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\|\Gamma_{X,Y,t}\|_{L^\infty(\mu\otimes\mu)}\leqslant\eta.$$

The claim follows since $\eta>d_{\mathcal{GW},\infty}^{\mathrm{spec}}(X,Y)$ was arbitrary.

(B) Assume that

$$\sup_{t>0}c^2(t)\cdot\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\|\Gamma_{X,Y,t}\|_{L^\infty(\mu\otimes\mu)}<\eta$$

for some $\eta>0$. Then, for all $t>0$ there exists $\mu_t\in\mathcal{M}(\mu_X,\mu_Y)$ s.t.

$$\|\Gamma_{X,Y,t}\|_{L^\infty(\mu_t\otimes\mu_t)}\leqslant\eta/c^2(t).$$

Fix $t>0$, then in particular,

$$\big|\mathbf{Vol}(X)\cdot k_X(t,x,x)-\mathbf{Vol}(Y)\cdot k_Y(t,y,y)\big|\leqslant\eta/c^2(t)$$

for all $(x,y)\in R(\mu_t)$ and hence, recalling the definition of the HKS (4.18),

$$\begin{aligned}\eta &\geqslant c^2(t)\cdot\Big(\sup_{(x,y)\in R(\mu_t)}\big|\mathrm{hks}_X(x,t)-\mathrm{hks}_Y(y,t)\big|\Big)\\ &\geqslant c^2(t)\cdot\Big(\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\sup_{(x,y)\in R(\mu)}\big|\mathrm{hks}_X(x,t)-\mathrm{hks}_Y(y,t)\big|\Big)\end{aligned}$$

for all $t>0$. But then, $\eta\geqslant\sup_{t>0}c^2(t)\cdot\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\|\mathrm{hks}_X(\cdot,t)-\mathrm{hks}_Y(\cdot,t)\|_{L^\infty(\mu)}$.

Now, the claim follows since the $L^\infty(\mu)$ norm dominates all $L^p(\mu)$ norms for $p \geqslant 1$, any $\mu \in \mathcal{M}(\mu_X, \mu_Y)$.

(C) Follows directly from (B) for $p = 1$ and then invoking Lemma 11.2.

(D) Assume $\eta > 0$ is s.t. for all $t > 0$

$$\int_0^\infty \left|\mathcal{H}_X(t,s) - \mathcal{H}_Y(t,s)\right| ds < \eta / c^2(t). \tag{11.48}$$

Fix $t > 0$. Then, by standard results on transportation of probability measures on the real line [59, Section 2.2], the LHS is equal to $d^{\mathbb{R}}_{\mathcal{W},1}(d\mathcal{H}_X(t), d\mathcal{H}_Y(t))$, the 1-Wasserstein distance between $d\mathcal{H}_X(t)$ and $d\mathcal{H}_Y(t)$. Then, by the Kantorovich–Rubinstein duality formula [59, Theorem 1.3]:

$$d^{\mathbb{R}}_{\mathcal{W},1}\big(d\mathcal{H}_X(t), d\mathcal{H}_Y(t)\big) = \sup_{(\varphi,\psi)\in\Phi_c} \left(\int_0^\infty \varphi(a)\, d\mathcal{H}_X(t)(da) + \int_0^\infty \psi(b)\, d\mathcal{H}_Y(t)(db) \right), \tag{11.49}$$

where Φ_c denotes the set of all measurable functions $(\varphi, \psi) \in L^1(d\mathcal{H}_X(t)) \times L^1(d\mathcal{H}_Y(t))$ s.t. $\varphi(a) + \psi(b) \leqslant |a - b|$ for $d\mathcal{H}_X(t) \otimes d\mathcal{H}_Y(t)$-almost all $(a, b) \in \mathbb{R}^2$.

In particular we can choose for all $s \in \mathbb{R}$, $\varphi(s) = -\psi(s) = s$. That this choice leads to a pair (φ, ψ) in Φ_c follows from Remark 4.7 and the finiteness of the heat trace for $t > 0$. Now from (11.49) above we obtain

$$d^{\mathbb{R}}_{\mathcal{W},1}\big(d\mathcal{H}_X(t), d\mathcal{H}_Y(t)\big) \geqslant \int_0^\infty s\big(d\mathcal{H}_X(t) - d\mathcal{H}_Y(t)\big)(ds)$$

and by symmetry

$$d^{\mathbb{R}}_{\mathcal{W},1}\big(d\mathcal{H}_X(t), d\mathcal{H}_Y(t)\big) \geqslant \left| \int_0^\infty s\big(d\mathcal{H}_X(t) - d\mathcal{H}_Y(t)\big)(ds) \right|.$$

Now, by Remark 4.7, the RHS equals $|K_X(t) - K_Y(t)|$. So, putting all together, from (11.48) we obtain,

$$\left|K_X(t) - K_Y(t)\right| \leqslant \eta / c^2(t), \quad \text{for all } t > 0.$$

Since $t > 0$ is arbitrary we find $\sup_{t>0} c^2(t) \cdot |K_X(t) - K_Y(t)| \leqslant \eta$. This concludes the proof. □

Proof of Theorem 8.2. Fix any $R \in \mathcal{R}(X, Y)$ and $t > 0$.

Claim 11.1. *For all $(x, y), (x', y') \in R$ one has*

$$\left|d^{\text{spec}}_{X;t}(x, x') - d^{\text{spec}}_{Y;t}(y, y')\right|^2 \leqslant \left|\big(d^{\text{spec}}_{X;t}(x, x')\big)^2 - \big(d^{\text{spec}}_{Y;t}(y, y')\big)^2\right|.$$

Proof. The claim follows from the inequality $|a - b|^2 \leqslant |a^2 - b^2|$ valid for all $a, b \geqslant 0$. □

Claim 11.2. *One has*

$$\big(\|\Gamma_{X_t,Y_t}\|_{L^\infty(R\times R)}\big)^2 \leqslant 4 \cdot \left\|\Gamma^{\text{spec}}_{X,Y,t}\right\|_{L^\infty(R\times R)}.$$

Proof. Recall (4.25) and notice that from the previous claim and the triangle inequality for the absolute value one obtains that for all $(x, y), (x', y') \in R$,

$$\begin{aligned}
\left|d^{\text{spec}}_{X;t}(x, x') - d^{\text{spec}}_{Y;t}(y, y')\right|^2 &\leqslant \left|\mathbf{Vol}(X) \cdot k_X(t, x, x) - \mathbf{Vol}(Y) \cdot k_Y(t, y, y)\right| \\
&\quad + \left|\mathbf{Vol}(X) \cdot k_X\big(t, x', x'\big) - \mathbf{Vol}(Y) \cdot k_Y\big(t, y', y'\big)\right| \\
&\quad + 2 \cdot \left|\mathbf{Vol}(X) \cdot k_X\big(t, x, x'\big) - \mathbf{Vol}(Y) \cdot k_Y\big(t, y, y'\big)\right| \\
&\leqslant 4 \cdot \left\|\Gamma^{\text{spec}}_{X,Y,t}\right\|_{L^\infty(R\times R)}.
\end{aligned}$$

But then, since the RHS does not depend on x, x', y and y' we find the claim. □

Now, pick any $\mu \in \mathcal{M}(\mu_X, \mu_Y)$. By Lemma 2.1, $R(\mu) = \text{supp}[\mu] \in \mathcal{R}(X, Y)$. From the last claim we obtain first

$$\left(\frac{1}{2} \inf_\mu \|\Gamma_{X_t,Y_t}\|_{L^\infty(\mu\otimes\mu)}\right)^2 \leqslant \left\|\Gamma^{\text{spec}}_{X,Y,t}\right\|_{L^\infty(R(\mu)\times R(\mu))}.$$

According to (3.12), the LHS equals $(d_{\mathcal{GW},\infty}(X_t,Y_t))^2$ which does not depend on μ. Thus, from the above we obtain

$$\big(d_{\mathcal{GW},\infty}(X_t,Y_t)\big)^2 \leqslant \inf_{\mu}\big\|\Gamma^{\text{spec}}_{X,Y,t}\big\|_{L^\infty(R(\mu)\times R(\mu))}.$$

Multiplying both sides by $c^2(t)$ and taking sup over all $t>0$ we arrive at (A').

Now, for (B') pick $\eta>0$ s.t.

$$d_{\mathcal{GW},\infty}(X_t,Y_t) < \eta/c(t), \quad \text{for all } t>0. \tag{$*$}$$

Notice that by Theorem 3.1(h),

$$d_{\mathcal{GW},1}(X_t,Y_t) < \eta/c(t), \quad \text{for all } t>0$$

as well. Fix $t>0$ and let $\mu_t\in\mathcal{M}(\mu_X,\mu_Y)$ be s.t. $\|\Gamma_{X_t,Y_t}\|_{L^1(\mu_t\otimes\mu_t))}<2\eta/c(t)$. This means that

$$\iint\limits_{X\times Y}\iint\limits_{X\times Y}\big|d^{\text{spec}}_{X;t}(x,x')-d^{\text{spec}}_{Y;t}(y,y')\big|\mu_t(dx\times dy)\mu_t(dx'\times dy') < 2\eta/c(t). \tag{11.50}$$

Now, obviously, $\mu_t\otimes\mu_t$ gives rise to a measure coupling in $\mathcal{M}(\mu_X\otimes\mu_X,\mu_Y\otimes\mu_Y)$. We then apply Lemma 11.2 with $X\leftrightarrow X\times X$, $Y\leftrightarrow Y\times Y$, $f(\cdot)=d^{\text{spec}}_{X;t}(\cdot)$ and $g(\cdot)=d^{\text{spec}}_{Y;t}(\cdot)$. It then turns out that for all $s\geqslant 0$, the distributions F and G of f and g, respectively, are

$$F(s)=(\mu_X\otimes\mu_X)\big\{(x,x')\in X\times X,\ \text{s.t.}\ d^{\text{spec}}_{X;t}(x,x')\leqslant s\big\}$$

and

$$G(s)=(\mu_Y\otimes\mu_Y)\big\{(y,y')\in Y\times Y,\ \text{s.t.}\ d^{\text{spec}}_{Y;t}(y,y')\leqslant s\big\}$$

which, according to (4.26), equal $\mathcal{G}_X(t,s)$ and $\mathcal{G}_Y(t,s)$, respectively. Now, from (11.50) and Lemma 11.2,

$$\int_0^\infty\big|\mathcal{G}_X(t,s)-\mathcal{G}_Y(t,s)\big|\,ds \leqslant 2\eta/c(t).$$

The conclusion now follows since $t>0$ and η satisfying $(*)$ were arbitrary.

For (C') pick $\mu\in\mathcal{M}(\mu_X,\mu_Y)$, $t>0$, $p\in[1,\infty)$ and write

$$\begin{aligned}\big\|d^{\text{spec}}_{X;t}-d^{\text{spec}}_{Y;t}\big\|^p_{L^p(\mu\otimes\mu)} &= \iint\limits_{X\times Y}\bigg[\bigg(\iint\limits_{X\times Y}\big|d^{\text{spec}}_{X;t}(x,x')-d^{\text{spec}}_{Y;t}(y,y')\big|^p\mu(dx'\times dy')\bigg)^{1/p}\bigg]^p\mu(dx\times dy)\\ &= \iint\limits_{X\times Y}\big(\big\|d^{\text{spec}}_{X;t}(x,\cdot)-d^{\text{spec}}_{Y;t}(y,\cdot)\big\|_{L^p(\mu)}\big)^p\mu(dx\times dy)\\ &\geqslant \iint\limits_{X\times Y}\big|\big\|d^{\text{spec}}_{X;t}(x,\cdot)\big\|_{L^p(\mu)}-\big\|d^{\text{spec}}_{Y;t}(y,\cdot)\big\|_{L^p(\mu)}\big|^p\mu(dx\times dy).\end{aligned} \tag{11.51}$$

But then since μ is a coupling between μ_X and μ_Y, $\|d^{\text{spec}}_{X;t}\|_{L^p(\mu)}=\|d^{\text{spec}}_{X;t}\|_{L^p(\mu_X)}=\text{ecc}^{\text{spec}}_{X;p}(t,x)$ and $\|d^{\text{spec}}_{Y;t}\|_{L^p(\mu)}=\|d^{\text{spec}}_{Y;t}\|_{L^p(\mu_Y)}=\text{ecc}^{\text{spec}}_{Y;p}(t,y)$ for all $x\in X$ and $y\in Y$.

From (11.51) we then have that

$$\big\|d^{\text{spec}}_{X;t}-d^{\text{spec}}_{Y;t}\big\|_{L^p(\mu\otimes\mu)} \geqslant \inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\big\|\text{ecc}^{\text{spec}}_{X;p}(t,\cdot)-\text{ecc}^{\text{spec}}_{Y;p}(t,\cdot)\big\|_{L^p(\mu)}.$$

From this and the definition of $d_{\mathcal{GW},p}(X,Y)$ (Definition 3.3) we see that

$$d_{\mathcal{GW},p}(X,Y) \geqslant \frac{1}{2}\inf_{\mu\in\mathcal{M}(\mu_X,\mu_Y)}\big\|\text{ecc}^{\text{spec}}_{X;p}(t,\cdot)-\text{ecc}^{\text{spec}}_{Y;p}(t,\cdot)\big\|_{L^p(\mu)}.$$

Invoking Theorem 3.1(h) and (A') we conclude the proof. □

Proof of Theorem 9.3. Symmetry are non-negativity are clear. The triangle inequality is easy and shall concern us first. Let $X,Y,Z\in\mathfrak{R}$ and $a\in\mathcal{B}(X)$, $b\in\mathcal{B}(Y)$ and $c\in\mathcal{B}(Z)$ be such that

$$d^{\mathcal{F}_{\text{T}}}_{\mathcal{H}}\big(\mathcal{X}(a),\mathcal{Y}(b)\big) < \delta_1$$

and

$$d_{\mathcal{H}}^{\mathcal{F}_{\mathrm{T}}}\big(\mathcal{Y}(b),\mathcal{Z}(c)\big)<\delta_2$$

for some $\delta_1,\delta_2>0$.

By the triangle inequality for $d_{\mathcal{H}}^{\mathcal{F}_{\mathrm{T}}}(,)$ it then follows that

$$d_{\mathcal{H}}^{\mathcal{F}_{\mathrm{T}}}\big(\mathcal{X}(a),\mathcal{Z}(c)\big)<\delta_1+\delta_2.$$

Since the LHS is an upper bound for $\mathcal{H}\hat{d}_{\mathrm{T}}(X,Z)$, we conclude the proof of the triangle inequality by invoking the arbitrariness of δ_1 and δ_2.

The fact that $\mathcal{H}\hat{d}_{\mathrm{T}}(X,Y)=0$ implies that X and Y are isometric follows from the proof of Theorem 10 in [47]. Indeed, pick $(\varepsilon_n)\subset\mathbb{R}^+$ with zero limit. Then, there exist $(a_n)\subset\mathcal{B}(X)$ and $(b_n)\subset\mathcal{B}(Y)$ s.t.

$$d_{\mathcal{H}}^{\mathcal{F}_{\mathrm{T}}}\big(\mathcal{X}(a_n),\mathcal{Y}(b_n)\big)<\varepsilon_n\quad\text{for all } n\in\mathbb{N}.$$

Now, both $\mathcal{B}(X)$ and $\mathcal{B}(Y)$ are compact (in the topology generated by the metric $d_{\mathcal{B}}$ defined in [47, § IV]). One can therefore extract a subsequence $(n_k)\subset\mathbb{N}$ s.t. both (a_{n_k}) and (b_{n_k}) converge to some $a\in\mathcal{B}(X)$ and $b\in\mathcal{B}(Y)$, respectively, as $k\to\infty$. Furthermore, it follows from the above that

$$d_{\mathcal{H}}^{\mathcal{F}_{\mathrm{T}}}\big(\mathcal{X}(a),\mathcal{Y}(b)\big)=0.$$

In particular, one must have that $d_{\mathcal{H}}^{\ell^2}(I_X^a[X](t),I_Y^b[Y](t))=0$ for any $t\in\mathrm{T}$. From this point onwards, the same proof as [47, Theorem 10] applies. □

Proof of Corollary 9.1. See proof of Corollary 9.2 below. □

Proof of Lemma 9.1. Let $\{\lambda_i\}_{i\geqslant 0}$ denote the eigenvalues of the Laplace–Beltrami operator on X and let $\{\zeta_i\}_{i\geqslant 0}$ be a basis of $L^2(X)$ composed by associated eigenfunctions. From [48, Proposition 2.8], $\lambda_i\geqslant C\cdot i^{2/n}$ and $\|\zeta_i^2\|_{L^\infty(X)}\leqslant C'\cdot(\kappa^2\cdot D^2+i^2)^{n/2}$ for all $i\in\mathbb{N}$, where $n=\dim(X)$, κ is s.t. $\mathbf{Ric}_X\geqslant-(n-1)\kappa^2$ and $D=\mathbf{diam}(X)$ and C and C' are constants depending only on the geometry of X. Then, using the inequality $|a-b|^2\leqslant|a^2-b^2|$ valid of all $a,b\in\mathbb{R}^+$ and the definition of the diffusion distance (4.24), we have that for all $x,x'\in X$,

$$\big|d_{X;t}^{\mathrm{spec}}(x,x')-d_{X;t;m}^{\mathrm{spec}}(x,x')\big|^2\leqslant\mathbf{Vol}(X)\sum_{i>m}e^{-\lambda_i t}\big(\zeta_i(x)-\zeta_i(x')\big)^2.$$

With simple estimates and invoking the bounds on λ_i and $\|\zeta_i\|_{L^\infty(X)}$ provided above, one finds

$$\big|d_{X;t}^{\mathrm{spec}}(x,x')-d_{X;t;m}^{\mathrm{spec}}(x,x')\big|^2\leqslant 4\,\mathbf{Vol}(X)2^{n/2}C'\sum_{i>m}e^{-Ct\cdot i^{2/n}}i^n.$$

Now, since the series $\sum_{i\geqslant 1}e^{-Ct\cdot i^{2/n}}i^n$ is convergent for fixed $t>0$, we obtain that the RHS above tends to zero as $m\uparrow\infty$ and hence the claim. □

Proof of Theorem 9.4. See proof of Theorem 9.6 below. □

Proof of Theorem 9.5. The proof of this result is a direct application of [11, Theorem 2]. □

Proof of Theorems 9.4 and 9.6. Both claims follow easily from the following

Claim 11.3. *For all $R\in\mathrm{iso}(\ell^2)$, $a\in\mathcal{B}(X)$, $b\in\mathcal{B}(Y)$, $x,x'\in X$, $y,y'\in Y$ and $t>0$ it holds that*

$$c^2(t)\cdot\big|\mathbf{Vol}(X)\cdot k_X(t,x,x')-\mathbf{Vol}(Y)\cdot k_Y(t,y,y')\big|\leqslant A_X\cdot\big\|I_X^a[x'](t)-RI_Y^b[y'](t)\big\|_{\ell^2}+A_Y\cdot\big\|I_X^a[x](t)-RI_Y^b[y](t)\big\|_{\ell^2},$$

where $A_X:=\sup\{\|I_X^a[x](t)\|_{\ell^2},x\in X,t>0,a\in\mathcal{B}(X)\}$ and $A_Y:=\sup\{\|I_Y^b[y](t)\|_{\ell^2},y\in Y,t>0,b\in\mathcal{B}(Y)\}$.

The proof of the claim follows from the definition (4.22) of the spectral embeddings $I_X^a[\cdot](t)$, Eq. (4.23), and the Cauchy–Schwartz inequality.

We notice now that by (4.23) and the definition of spectral variances (5.1):

$$A_X=\big(\sup\big\{c(t)\cdot\big(\mathbf{Vol}(X)k_X(t,x,x)-1\big),x\in X,t>0\big\}\big)^{1/2}\leqslant\big(\mathbf{var}_\infty^{\mathrm{spec}}(X)\big)^{1/2}.$$

A similar claim holds for A_Y.

For the proof of Theorem 9.4, pick $C \in \mathcal{R}(X,Y)$ and $R \in \mathrm{iso}(\ell^2)$. From the claim above it follows that

$$\sup_{t>0} c^2(t) \cdot \big\| \Gamma^{\mathrm{spec}}_{X,Y,t} \big\|_{L^\infty(C\times C)} \leqslant 2\big(\mathbf{var}^{\mathrm{spec}}_\infty(X), \mathbf{var}^{\mathrm{spec}}_\infty(Y)\big)^{1/2} \sup_{t>0} \big\| \big\| I^a_X[\cdot](t) - R I^b_Y[\cdot](t) \big\|_{\ell^2} \big\|_{L^\infty(C)}.$$

Now, by definition of the spectral Gromov–Hausdorff distance (Section 9.1) the LHS is greater than or equal to $d^{\mathrm{spec}}_{\mathcal{GH}}(X,Y)$ so that one obtains

$$\begin{aligned} d^{\mathrm{spec}}_{\mathcal{GH}}(X,Y) &\leqslant 2\big(\mathbf{var}^{\mathrm{spec}}_\infty(X), \mathbf{var}^{\mathrm{spec}}_\infty(Y)\big)^{1/2} \sup_{t>0} \big\| \big\| I^a_X[\cdot](t) - R I^b_Y[\cdot](t) \big\|_{\ell^2} \big\|_{L^\infty(C)} \\ &\leqslant 2\big(\mathbf{var}^{\mathrm{spec}}_\infty(X), \mathbf{var}^{\mathrm{spec}}_\infty(Y)\big)^{1/2} \big\| \sup_{t>0} \big\| I^a_X[\cdot](t) - R I^b_Y[\cdot](t) \big\|_{\ell^2} \big\|_{L^\infty(C)} \end{aligned}$$

for all $R \in \mathrm{iso}(\ell^2)$ and $C \in \mathcal{R}(X,Y)$. From the arbitrariness of C it follows that

$$d^{\mathrm{spec}}_{\mathcal{GH}}(X,Y) \leqslant 2\big(\mathbf{var}^{\mathrm{spec}}_\infty(X), \mathbf{var}^{\mathrm{spec}}_\infty(Y)\big)^{1/2} \inf_{C\in\mathcal{R}(X,Y)} \big\| \big\| I^a_X[\cdot](\cdot) - R I^b_Y[\cdot](\cdot) \big\|_{\mathrm{T}} \big\|_{L^\infty(C)}$$

for all $R \in \mathrm{iso}(\ell^2)$. By Proposition 2.1 the RHS equals $d^{\mathcal{F}_{\mathrm{T}}}_{\mathcal{H}}(\mathcal{X}(a), R^*\mathcal{Y}(b))$.

The conclusion follows from the arbitrariness of R and (9.36), the definition of $\mathcal{H}D^{\ell^2}_{\mathbb{R}^+}$.

An almost identical argument applies to the proof of Theorem 9.6. We omit details. □

Proof of Corollaries 9.1 and 9.2. For all the metrics involved, symmetry and non-negativity is clear. By Theorems 9.4, 9.6, 9.1 and 6.1, and the inequalities (9.35), (9.37) and (9.43), if for any pair $X, Y \in \mathfrak{R}$ any of the metrics involved yields the value 0, then X and Y must be isometries. In order to establish that they are proper metrics, it therefore suffices to establish the triangle inequality for these metrics.

The triangle inequality for $\mathcal{H}d_{\mathbb{R}^+}$ follows from considerations in [47, Theorem 10]. For the triangle inequality for $\mathcal{H}\hat{d}_{\mathbb{R}^+}$ see the proof of Theorem 9.3 above. The triangle inequality for $\mathcal{H}D^{\ell^2}_{\mathrm{T}}$ is also easy. Finally, the triangle inequalities for the Wasserstein versions follow from arguments similar to those of their corresponding Hausdorff counterparts and we omit them.

Acknowledgments

The author gratefully acknowledges Gunnar Carlsson for his patient mentorship and support. Thanks to Maks Ovsjanikov, Martin Reuter, Jian Sun, Triet Le, and the anonymous reviewers for their input and feedback about this work. This work was supported by ONR grant N00014-09-1-0783.

References

[1] P. Frosini, A distance for similarity classes of submanifolds of Euclidean space, Bull. Austral. Math. Soc. 42 (3) (1990) 407–416.
[2] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 509–522, http://dx.doi.org/10.1109/34.993558.
[3] S. Manay, D. Cremers, B. Hong, A. Yezzi, S. Soatto, Integral invariants for shape matching, IEEE Trans. Pattern Anal. Mach. Intell. 28 (10) (2006) 1602–1618.
[4] A. Hamza, H. Krim, Geodesic object representation and recognition, in: Lect. Notes Control Comput. Sci., vol. 2886, Springer, 2003, pp. 378–387.
[5] R. Osada, T. Funkhouser, B. Chazelle, D. Dobkin, Shape distributions, ACM Trans. Graph. 21 (4) (2002) 807–832, http://dx.doi.org/10.1145/571647.571648.
[6] A. Elad, R. Kimmel, On bending invariant signatures for surfaces, IEEE Trans. Pattern Anal. Mach. Intell. 25 (10) (2003) 1285–1295.
[7] M. Reuter, F.-E. Wolter, N. Peinecke, Laplace–Beltrami spectra as "Shape-DNA" of surfaces and solids, Comput. Aided Design 38 (4) (2006) 342–366.
[8] F. Mémoli, G. Sapiro, Comparing point clouds, in: Proc. of SGP, ACM, 2004, pp. 32–40, http://dx.doi.org/10.1145/1057432.1057436.
[9] F. Mémoli, G. Sapiro, A theoretical and computational framework for isometry invariant recognition of point cloud data, Found. Comput. Math. 5 (3) (2005) 313–347.
[10] M. Gromov, Metric structures for Riemannian and non-Riemannian spaces, Progr. Math., vol. 152, Birkhäuser Boston Inc., 1999.
[11] F. Mémoli, Gromov–Hausdorff distances in Euclidean spaces, NORDIA, 2008, pp. 1–8, http://dx.doi.org/10.1109/CVPRW.2008.4563074.
[12] F. Chazal, D. Cohen-Steiner, L. Guibas, F. Mémoli, S. Oudot, Gromov–Hausdorff stable signatures for shapes using persistence, in: Proc. of SGP, 2009.
[13] F. Mémoli, On the use of Gromov–Hausdorff distances for shape comparison, in: Proceedings of Point Based Graphics 2007, Prague, Czech Republic, 2007.
[14] F. Mémoli, Gromov–Wasserstein distances and the metric approach to object matching, Tech. rep., preprint, 2010.
[15] B. Levy, Laplace–Beltrami eigenfunctions towards an algorithm that understands geometry, in: Proc. of SMI, 2006, pp. 13–13, doi:10.1109/SMI.2006.21.
[16] H. Qiu, Spectral methods for computer vision problems, Ph.D. thesis, University of York, November 2006.
[17] S. Lafon, Diffusion maps and geometric harmonics, Ph.D. thesis, Yale University, 2004.
[18] 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, ACM Press, New York, NY, USA, 2005, pp. 101–106, http://dx.doi.org/10.1145/1060244.1060256.
[19] M. Reuter, Hierarchical shape segmentation and registration via topological features of Laplace–Beltrami eigenfunctions, Int. J. Comput. Vision 89 (2) (2010) 287–308.
[20] R. Rustamov, Laplace–Beltrami eigenfunctions for deformation invariant shape representation, in: Proc. of SGP, 2007, pp. 225–233.
[21] B. Vallet, B. Lévy, Spectral geometry processing with manifold harmonics, Comput. Graph. Forum 27 (2) (2008) 251–260.
[22] M. Ovsjanikov, J. Sun, L. Guibas, Global intrinsic symmetries of shapes, Comput. Graph. Forum 27 (5) (2008) 1341–1348, http://dblp.uni-trier.de/db/journals/cgf/cgf27.html#OvsjanikovSG08.

[23] J. Sun, M. Ovsjanikov, L. Guibas, A concise and provably informative multi-scale signature based on heat diffusion, in: SGP, 2009.
[24] A.M. Bronstein, M.M. Bronstein, R. Kimmel, M. Mahmoudi, G. Sapiro, A Gromov–Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching, Int. J. Comput. Vision 89 (2–3) (2010) 266–286, http://dx.doi.org/10.1007/s11263-009-0301-6.
[25] F. de Goes, S. Goldenstein, L. Velho, A hierarchical segmentation of articulated bodies, Comput. Graph. Forum 27 (5) (2008) 1349–1356, http://dblp.uni-trier.de/db/journals/cgf/cgf27.html#GoesGV08.
[26] M. Gori, M. Maggini, L. Sarti, Exact and approximate graph matching using random walks, IEEE Trans. Pattern Anal. Mach. Intell. 27 (7) (2005) 1100–1111, http://dx.doi.org/10.1109/TPAMI.2005.138.
[27] L. Lieu, N. Saito, Pattern recognition by diffusion maps and earth mover's distance, Tech. rep., 2008, http://www.math.ucdavis.edu/~llieu/PRbyDMEMD.pdf.
[28] S. Lafon, Y. Keller, R.R. Coifman, Data fusion and multicue data matching by diffusion maps, IEEE Trans. Pattern Anal. Mach. Intell. 28 (11) (2006) 1784–1797, http://dx.doi.org/10.1109/TPAMI.2006.223.
[29] B. Nadler, S. Lafon, R. Coifman, I. Kevrekidis, Diffusion maps, spectral clustering and reaction coordinates of dynamical systems, Appl. Comput. Harmon. Anal. 21 (1) (2006) 113–127.
[30] R. Coifman, S. Lafon, Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions, Appl. Comput. Harmon. Anal. 21 (1) (2006) 31–52.
[31] S. Lafon, A.B. Lee, Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization, IEEE Trans. Pattern Anal. Mach. Intell. 28 (9) (2006) 1393–1403, http://dx.doi.org/10.1109/TPAMI.2006.184.
[32] X. Bai, Heat kernel analysis on graphs, Ph.D. thesis, The University of York, 2007.
[33] V. Jain, H. Zhang, O. van Kaick, Non-rigid spectral correspondence of triangle meshes, Int. J. Shape Model. 13 (1) (2007) 101–124.
[34] J. Milnor, Eigenvalues of the Laplace operator on certain manifolds, Proc. Natl. Acad. Sci. USA 51 (1964) 542.
[35] S. Bando, H. Urakawa, Generic properties of the eigenvalues of the Laplacian for compact Riemannian manifolds, Tohoku Math. J. 35 (2) (1983) 155–172.
[36] K. Gebal, A. Baerentzen, H. Aanæs, R. Larsen, Shape analysis using the auto diffusion function, in: SGP, 2009.
[37] M. Ovsjanikov, Q. Mérigot, F. Mémoli, L. Guibas, One point isometric matching with the heat kernel, Comput. Graph. Forum 29 (5) (2010) 1555–1564.
[38] P. Skraba, M. Ovsjanikov, F. Chazal, L. Guibas, Persistence-based segmentation of deformable shapes, in: CVPR Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment, 2010, in press, http://dx.doi.org/10.1109/CVPRW.2010.5543285.
[39] T.K. Dey, K. Li, C. Luo, P. Ranjan, I. Safa, Y. Wang, Persistent heat signature for pose-oblivious matching of incomplete models, in: Eurographics Symposium on Geometry Processing (SGP), 2010, in press, doi:10.1111/j.1467-8659.2010.01763.x.
[40] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396, http://dx.doi.org/10.1162/089976603321780317.
[41] R. Coifman, S. Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (1) (2006) 5–30.
[42] U. Vaidya, G. Hagen, A. Banaszuk, S. Lafon, I. Mezic, R. Coifman, Comparison of systems using diffusion maps, in: 44th IEEE Conference on Decision and Control, 2005 and 2005 European Control Conference, CDC-ECC '05, 2005, pp. 7931–7936.
[43] P.W. Jones, M. Maggioni, R. Schul, Universal local parametrizations via heat kernels and eigenfunctions of the Laplacian, 2007, http://www.citebase.org/abstract?id=oai:arXiv.org:0709.1975.
[44] R.R. Coifman, I.G. Kevrekidis, S. Lafon, M. Maggioni, B. Nadler, Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems, Multiscale Model. Simul. 7 (2) (2008) 842–864, http://dx.doi.org/10.1137/070696325.
[45] M. Hilaga, Y. Shinagawa, T. Kohmura, T. Kunii, Topology matching for fully automatic similarity estimation of 3d shapes, in: Proc. of SIGGRAPH, ACM, 2001, pp. 203–212, http://dx.doi.org/10.1145/383259.383282.
[46] M. Gromov, Groups of polynomial growth and expanding maps, Publ. Math. Inst. Hautes Études Sci. 53 (1981) 53–73.
[47] P. Berard, G. Besson, S. Gallot, Embedding Riemannian manifolds by their heat kernel, Geom. Funct. Anal. 4 (4) (1994) 373–398.
[48] A. Kasue, H. Kumura, Spectral convergence of Riemannian manifolds, Tohoku Math. J. 46 (1994) 147–179.
[49] J. Lott, C. Villani, Ricci curvature for metric-measure spaces via optimal transport, Ann. of Math. (2) 169 (3) (2009) 903–991.
[50] K.-T. Sturm, On the geometry of metric measure spaces. I, Acta Math. 196 (1) (2006) 65–131.
[51] Y. Ollivier, Ricci curvature of Markov chains on metric spaces, J. Funct. Anal. 256 (3) (2009) 810–864.
[52] R.R. Coifman, M. Maggioni, Diffusion wavelets, Appl. Comput. Harmon. Anal. 21 (1) (2006) 53–94, doi:10.1016/j.acha.2006.04.004.
[53] D. Burago, Y. Burago, S. Ivanov, A Course in Metric Geometry, AMS, 2001.
[54] F. Mémoli, Lower bounds for the Gromov–Hausdorff distance between spheres, Tech. rep., Stanford University, 2010.
[55] F. Mémoli, Spectral Gromov–Wasserstein distances for shape matching, in: Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (ICCV workshop, NORDIA'09), 2009, http://math.stanford.edu/~memoli.
[56] R.M. Dudley, Real Analysis and Probability, Cambridge Stud. Adv. Math., vol. 74, Cambridge Univ. Press, Cambridge, 2002, revised reprint of the 1989 original.
[57] I. Chavel, Riemannian Geometry: A Modern Introduction, Cambridge Univ. Press, 1997.
[58] T. Sakai, Riemannian Geometry, Transl. Math. Monogr., vol. 149, Amer. Math. Soc., Providence, RI, 1996.
[59] C. Villani, Topics in Optimal Transportation, AMS, 2003.
[60] C.R. Givens, R.M. Shortt, A class of Wasserstein metrics for probability distributions, Michigan Math. J. 31 (2) (1984) 231–240.
[61] A. Grigor'yan, Heat kernels on weighted manifolds and applications, in: The Ubiquitous Heat Kernel, in: Contemp. Math., vol. 398, Amer. Math. Soc., Providence, RI, 2006, pp. 93–191.
[62] A. Bronstein, M. Bronstein, R. Kimmel, Efficient computation of isometry-invariant distances between surfaces, SIAM J. Sci. Comput. 28 (5) (2006) 1812–1836, doi:10.1137/050639296, http://link.aip.org/link/?SCE/28/1812/1.
[63] I. Chavel, Eigenvalues in Riemannian Geometry, Academic Press, 1984.
[64] E.P. Hsu, Stochastic Analysis on Manifolds, AMS, 2002.
[65] A. Grigor'yan, Heat kernels on weighted manifolds and applications, Contemp. Math. 398 (2006) 93–191.
[66] S. Rosenberg, The Laplacian on a Riemannian Manifold, London Math. Soc. Stud. Texts, vol. 31, Cambridge Univ. Press, Cambridge, 1997, An introduction to analysis on manifolds, http://dx.doi.org/10.1017/CBO9780511623783.
[67] A. Grigor'yan, Estimates of heat kernels on Riemannian manifolds, in: Spectral Theory and Geometry, Edinburgh, 1998, in: London Math. Soc. Lecture Note Ser., vol. 273, Cambridge Univ. Press, Cambridge, 1999, pp. 140–225.
[68] P.B. Gilkey, Invariance Theory, the Heat Equation, and the Atiyah–Singer Index Theorem, 2nd edn., Stud. Adv. Math., CRC Press, Boca Raton, FL, 1995.
[69] C. Müller, Analysis of Spherical Symmetries in Euclidean Spaces, Appl. Math. Sci., vol. 129, Springer-Verlag, New York, 1998.
[70] J.R. Norris, Heat kernel asymptotics and the distance function in Lipschitz Riemannian manifolds, Acta Math. 179 (1) (1997) 79–103.
[71] S. Minakshisundaram, A. Pleijel, Some properties of eigenfunctions of the Laplace operator on Riemannian manifolds, Canad. J. Math. 11 (1949) 243–256.
[72] H. Qiu, E.R. Hancock, Clustering and embedding using commute times, IEEE Trans. Pattern Anal. Mach. Intell. 29 (11) (2007) 1873–1890, http://dx.doi.org/10.1109/TPAMI.2007.1103.
[73] E.B. Davies, Gaussian upper bounds for the heat kernels of some second-order operators on Riemannian manifolds, J. Funct. Anal. 80 (1) (1988) 16–32.

[74] A. Bensoussan, J.-L. Lions, G. Papanicolaou, Asymptotic Analysis for Periodic Structures, Stud. Math. Appl., vol. 5, North-Holland Publishing, 1978.
[75] T. Tsuchida, Long-time asymptotics of heat kernels for one-dimensional elliptic operators with periodic coefficients, Proc. Lond. Math. Soc. (3) 97 (2) (2008) 450–476.
[76] E.B. Davies, Heat kernels in one dimension, Quart. J. Math. Oxford Ser. (2) 44 (175) (1993) 283–299.
[77] Z. Chen, Q. Zhongmin, H. Yaozhong, Z. Weian, Stability and approximations of symmetric diffusion semigroups and kernels, J. Funct. Anal. 152 (1) (1998) 255–280.
[78] Z.I. Szabó, A cornucopia of isospectral pairs of metrics on spheres with different local geometries, Ann. of Math. (2) 161 (1) (2005) 343–395.
[79] P.W. Jones, T.M. Le, Local scales and multiscale image decompositions, Appl. Comput. Harmon. Anal. 26 (3) (2009) 371–394, doi:10.1016/j.acha.2008.08.001.
[80] U. Pinkall, K. Polthier, Computing discrete minimal surfaces and their conjugates, Experiment. Math. 2 (1) (1993) 15–36.
[81] M. Meyer, M. Desbrun, P. Schröder, A.H. Barr, Discrete differential geometry operators for triangulated 2-manifolds, in: Proc. VisMath'02, Berlin, Germany, 2002.
[82] M. Belkin, J. Sun, Y. Wang, Discrete Laplace operator on meshed surfaces, in: Proc. of SOCG, 2008, pp. 278–287.
[83] E. Shamarova, An improvement of a result on Smolyanov–Weizsaecker surface measures, arXiv preprint math/0701281.