Topological Optimal Transport for Geometric Cycle Matching

Stephen Y Zhang[1,2], Michael P H Stumpf[1,2,3], Tom Needham[4], and Agnese Barbensi[5]

[1]School of Mathematics and Statistics, the University of Melbourne, Parkville, Melbourne, 3010, VIC, Australia
[2]Melbourne Integrative Genomics, the University of Melbourne, Parkville, Melbourne, 3010, VIC, Australia
[3]School of Biosciences, the University of Melbourne, Parkville, Melbourne, 3010, VIC, Australia
[4]Mathematics Department, Florida State University, Academic Way, Tallahassee, 32306, FL, USA
[5]School of Maths and Physics, University of Queensland, St Lucia, Brisbane, 4067, QLD, Australia

March 29, 2024

Abstract

Topological data analysis is a powerful tool for describing topological signatures in real world data. An important challenge in topological data analysis is matching significant topological signals across distinct systems. In geometry and probability theory, optimal transport formalises notions of distance and matchings between distributions and structured objects. We propose to combine these approaches, constructing a mathematical framework for optimal transport-based matchings of topological features. Building upon recent advances in the domains of persistent homology and optimal transport for hypergraphs, we develop a transport-based methodology for topological data processing. We define measure topological networks, which integrate both geometric and topological information about a system, introduce a distance on the space of these objects, and study its metric properties, showing that it induces a geodesic metric space of non-negative curvature. The resulting Topological Optimal Transport (TpOT) framework provides a transport model on point clouds that minimises topological distortion while simultaneously yielding a geometrically informed matching between persistent homology cycles.

1 Introduction

Topological data analysis (TDA) is a quickly growing field in computational and applied topology. In recent years, TDA has established itself as an effective framework to analyse, cluster, and detect patterns in complex data [52]. One of the key algorithms in TDA is persistent homology (PH), a computational tool that describes the structure of data based on topological features persisting across different scales [35, 18, 22]. In a nutshell, the PH algorithm proceeds by building a nested sequence of discrete spaces describing the input data at increasingly coarse scales, known as a *filtration* of simplicial complexes. The topological features of this filtration are then quantified by computing the homology groups of each simplicial complex in the sequence. A structural theorem [56] guarantees that the birth, death, and evolution of homology classes in the filtration can be summarised as a multi-set, called a *persistence diagram* (PD). PDs have been shown to contain rich information about the initial data, and PH-based data analysis approaches appear in an ever-increasing number of applications across multiple fields in modern science [40, 42, 51, 44].

Optimal transport (OT) is a far-reaching mathematical theory that, in its most basic and classical formulation, formalises the problem of finding a matching of a given probability distribution to another as efficiently as possible, in terms of a cost function [50, 36]. When the cost function is induced by the distance function on a metric space (X, d), optimal transport provides a natural lifting of the ground metric to a metric on the space of probability distributions supported on X, commonly known as the *Wasserstein* distance. An extension of the optimal transportation problem which has recently received considerable attention is the *Gromov-Wasserstein* (GW) problem [31, 32, 8, 43], in which the probability distributions to be matched are defined over different metric spaces, rather than supported on a common space. The objective is then to find matchings that minimise *distortion* of pairwise distances. While the GW distance was originally introduced to compare metric measure spaces and can be viewed as a relaxation of the well known Gromov-Hausdorff distance [31], some of its most successful applications address the problem of matching more general structured objects such as (labelled) graphs [54, 53, 10, 48]. Among its numerous advantages, the GW distance has brought substantial improvements to graph processing tasks over classical methods, and can be efficiently approximated by a range of numerical algorithms in practice [53, 37, 20, 9]. On the theoretical front, the space of *gauged measure spaces* (i.e., a measure space (X,μ) together with a symmetric gauge function $k: X\times X \to \mathbb{R}$) endowed with the GW distance has the geometry of an Alexandrov space, and a Riemannian orbifold structure can be developed in this setting [43, 10]. More recently, inspired by the goal of encoding relations in complex systems as higher-order networks, a variant of the GW problem, referred to as *co-optimal transport* [46], was applied to model hypergraphs both from a theoretical and applied point of view [11].

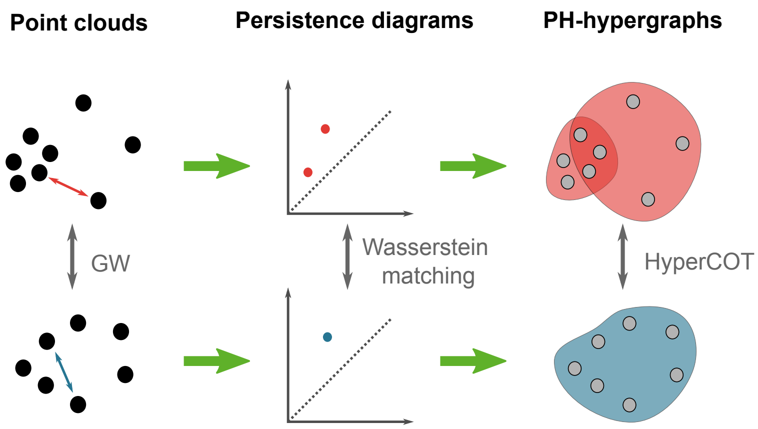

Figure 1: **The Topological Optimal Transport (TpOT) problem.** From left to right: two input point clouds, their persistence diagrams, and corresponding PH-hypergraphs. The objective of the TpOT problem combines a Gromov-Wasserstein distortion on the point clouds (geometric information), (partial) Wasserstein matching of points in the persistence diagrams (topological information) and HyperCOT on the PH-hypergraphs (coupling of geometric and topological information).

In this paper, we combine the TDA and OT approaches described above to develop a transport-based theory for topological data processing. Our approach is inspired by recent extensions built upon persistent homology computations: the hyperTDA framework [2] encodes each point in a persistence diagram, together with a geometric realisation of its homology class as a generating cycle, as a higher-order network called a *PH-hypergraph*. In a PH-hypergraph, vertices corresponds to data points used as input to the PH computation, and hyperedges are given by the generators of features in the persistence diagram. We define the *topological optimal transport (TpOT) problem*, by considering a trade-off between preservation of geometric relationships, preservation of topological features, and coupling the two via the PH-hypergraph structure (see Figure 1). The output is **(1)** a matching between points that is topologically driven and geometrically informed, coupled with **(2)** a matching between persistent homology classes that is geometrically driven and topologically informed.

Finding meaningful ways of matching topological features across distinct systems is an important challenge in TDA. The past few years have seen considerable effort in addressing this question, with a number of different solutions proposed [13, 3, 25, 38, 55, 21], many of which bridge between topological data analysis and optimal transport [23, 27, 30]. A key aspect of the TpOT method is leveraging existing transportation theory for persistence diagrams to drive a matching between topological features, while also taking into account geometric information and the spatial interconnectivity of homology generators. By endowing PH-hypergraphs with probability measures, we introduce the concept of *measure topological networks*, and we formalise a framework for studying the TpOT induced distance $d_{\mathrm{TpOT},p}$. Our main contributions are as follows:

- We develop a flexible measure-theoretic formalism for simultaneously encoding the geometry and topology of a finite point cloud, in the form of the aforementioned measure topological networks, as well as a family of distances $d_{\mathrm{TpOT},p}$, $p \geq 1$, between these objects; see Definitions 1 and 2, respectively.
- The distance $d_{\mathrm{TpOT},p}$ is shown to be a pseudometric on the space of measure topological networks $\mathscr{P}$, and the zero distance equivalence relation is completely characterised; see Theorem 1.

- We show that the metric induced by $d_{\mathrm{TpOT},p}$ is geodesic, characterise the exact form of the geodesics and use this characterisation to show that this metric space is non-negatively curved, in the sense of Alexandrov; see Theorems 2, 3 and 4, respectively.

- Our framework is centred on an optimisation problem for which we provide efficient numerical algorithms, which we demonstrate on a variety of examples; see Section 5.

The paper is organised as follows. Section 2 describes the necessary mathematical background on persistent homology and optimal transport. In Section 3, after motivating and constructing the TpOT problem, we define *measure topological networks* and the family of (pseudo-)metrics $d_{\mathrm{TpOT},p}$. We then discuss theoretical properties (Subsection 3.2) of the distance $d_{\mathrm{TpOT},p}$, and provide a characterisation of geodesics in Section 4. Finally, we provide details of the computational implementation (Subsection 3.4), as well as examples (Section 5).

2 Mathematical background

Our formalisation of the TpOT problem relies on constructions in topological data analysis and measure theory. The aim of this section is to outline the mathematical background needed to define measure topological networks and the family of distances $d_{\mathrm{TpOT},p}$ described in the introduction.

2.1 Persistent homology and persistent homology-hypergraphs

Let (X,d) be a finite metric space – for example, a point cloud $X = \{x_1,\cdots,x_N\} \subset \mathbb{R}^n$ as illustrated in Figure 2A. For any $\varepsilon > 0$, the *Vietoris-Rips complex* $K(X)_\varepsilon$ (see *e.g.* [18, Ch.III.2]) is the simplicial complex obtained from X by adding a k-simplex $[x_{i_0}, x_{i_1},\cdots,x_{i_k}]$ whenever the distance between all pairs of points in $\{x_{i_0}, x_{i_1},\cdots,x_{i_k}\}$ is less than ε. Note that $K(X)_\varepsilon$ is a sub-complex of $K(X)_{\varepsilon'}$ whenever $\varepsilon \le \varepsilon'$. Thus, as ε grows, this yields a nested sequence of simplicial complexes:

$$\mathcal{K}_{\mathrm{VR}}(X) := K(X)_{\varepsilon_0} \hookrightarrow K(X)_{\varepsilon_1} \hookrightarrow \cdots \hookrightarrow K(X)_{\varepsilon_M}.$$

Here, the numbers ε_i, $i = 1,\ldots,M$, are the parameters corresponding to the creation of new simplices; see Figure 2B.

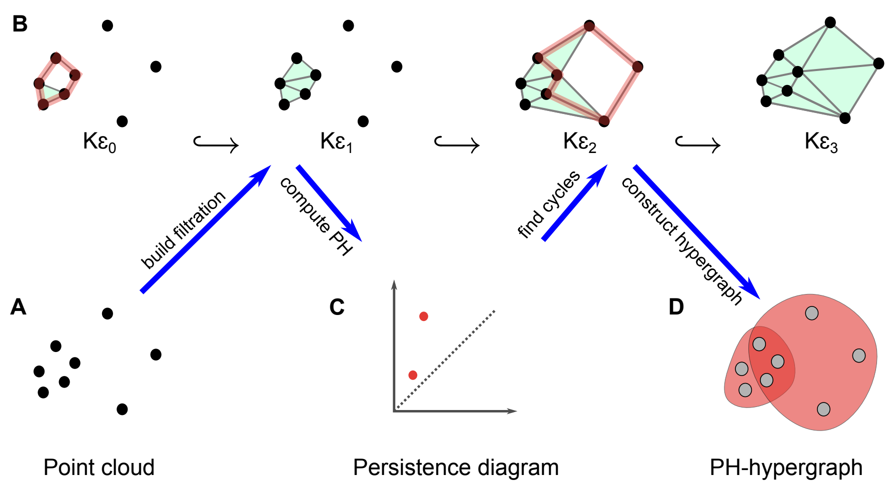

Figure 2: **Computing PH and the PH-hypergraph (A)** A point cloud and **(B)** a filtration of simplicial complexes built on it. **(C)** The 1-dimensional persistence diagram of the filtration in (B). **(D)** The PH-hypergraph constructed after computing a representative cycle (red cycles in (B)) for each class in (C).

More generally, for a finite metric space X, a *filtration* over X is a finite sequence $\mathcal{K} = (K_{\varepsilon_1},\ldots,K_{\varepsilon_M})$ of simplicial complexes, indexed by an increasing sequence of real numbers $\varepsilon_i \le \varepsilon_{i+1}$, together with simplicial maps $K_{\varepsilon_i} \to K_{\varepsilon_{i+1}}$ for all i, such that the vertex set of each K_{ε_i} is a subset of X. Common alternatives to the Vietoris-Rips complex in the literature with a similar interpretation include the Čech complex [18, Ch.III.2] and the Alpha complex [18, Ch.III.4]. Clearly, different constructions of the filtration $\mathcal{K}$ will influence the results of the persistent homology computation and their interpretation.

A filtration $\mathcal{K}$ over X can be analysed by computing the simplicial homology groups of each simplicial complex in the sequence and the corresponding induced maps. Collectively, this information is called the *persistent homology* (PH) of $\mathcal{K}$. The Structural Theorem [57], guarantees that, when using coefficients in $\mathbb{Z}/2\mathbb{Z}$, the birth, death, and evolution of homology classes in a filtration of simplicial complexes can be summarised in a multi-set, called a *persistence diagram*. We use Λ to denote the upper-diagonal cone in the upper-half plane, $\Lambda = \{(x, y) \in \mathbb{R}^2, y > x > 0\}$. The persistence diagram D associated to the filtration $\mathcal{K}$ is then a collection of points in Λ: $D = \{(b_i, d_i)\}_{i=1}^{|D|} \subset \Lambda$.

Each point $(b_i, d_i) \in D$ represents a homology class in the filtration $\mathcal{K}$, with coordinates corresponding to the birth b_i and death d_i parameters of the class (see Figure 2C). The *persistence* $d_i - b_i > 0$ of a homology class is a measure of significance of the underlying topological feature. The persistence diagram D can be interpreted as a topological descriptor of X. A local, geometric realisation of each homology class can be obtained by computing a representative cycle generating the class: see for instance the red cycles in Figure 2B. Computing cycles depends on a number of choices [5, 16, 34, 49, 29], and therefore different representatives may lead to differing downstream results.

The recently introduced PH-hypergraph [2] encodes the algebraic information of a persistence diagram, together with the local geometric interpretation of generating cycles, in a single higher-order network. Recall that a *hypergraph* is defined by a pair (V, E) consisting of a set of vertices V and a set of hyperedges E. Given a point cloud X, we may construct an associated filtration $\mathcal{K}$ together with a choice $g = \{c_1, \ldots, c_{|D|}\}$ of generating cycles for classes in the corresponding persistence diagram D. These data can be summarised in a PH-hypergraph $H = H(X, \mathcal{K}, g)$ in which we take X to be the vertex set, and one hyperedge added for each cycle in g (see Figure 2D). Moreover, for every cycle $c \in g$, a point $x_i \in X$ belongs to the corresponding hyperedge $e(c)$ if it is a vertex of any simplex in c. While the hypergraph structure H clearly depends on the choice of generators g and the choice of filtration $\mathcal{K}$, for a fixed filtration $\mathcal{K}$ the qualitative (hyper)network structure of the PH-hypergraph has been empirically shown to be stable to noise and different choices of generating cycles [2].

2.2 Optimal transport on persistence diagrams

Consider now two Borel probability measures μ and μ' supported on a Polish metric space (X, d). For $p \in [1, \infty)$, the *p-Wasserstein distance* [50, Definition 6.1] between μ and μ' is then defined as

$$d_{\mathrm{W},p}(\mu, \mu') := \left(\inf_{\pi \in \Pi(\mu, \mu')} \int_{X \times X} d(x, y)^p \, \mathrm{d}\pi(x, y) \right)^{1/p}. \tag{1}$$

This metric is the central object of study in the optimal transport (OT) literature. It is a standard result that the infimum in (1) is realised (i.e. it is in fact a *minimum*) and a minimiser π of (1) is referred to as an *optimal coupling*. Intuitively, this is a probabilistic matching between points in the supports of μ and μ'.

The optimal transport approach of finding matchings has inspired different notions of distance on the space of persistence diagrams: the most popular construction is known as the bottleneck distance [12], which is related to the ∞-Wasserstein distance. In this article, we are more interested in similar adaptations of the p-Wasserstein distances for $1 \le p < \infty$. The usual modification of the Wasserstein matching problem to compare persistence diagrams is to allow points to be matched to a virtual point ∂_Λ, representing the diagonal line $\{(a, a) \mid a \in \mathbb{R}\} \subset \mathbb{R}^2$ (i.e. homology classes with vanishing persistence). Formally, for $p \in [1, \infty)$, the *p-Wasserstein distance* between two persistence diagrams D and D' (see [28, Equation 6]) is defined as

$$d_{\mathrm{W},p}^{\mathrm{PD}}(D, D')^p = \min_{\pi \in \Pi(D, D')} \left(\sum_{(a,b) \in \pi} \|a - b\|_p^p + \sum_{s \in U_\pi} \|s - \mathrm{Proj}_{\partial_\Lambda}(s)\|_p^p \right), \tag{2}$$

where the various notations are described as follows:

- Slightly abusing notation, to illustrate the analogy to the OT Wasserstein distance, we denote by $\Pi(D, D')$ the set of all *partial matchings* between points representing homology classes in D and D'. For any $\pi \in \Pi(D, D')$, $\pi \subset D \times D'$ has the property that each $a \in D$ and $b \in D'$ appear in at most one ordered pair in π.

- The set U_π is the set of *unmatched points* for π: $s \in U_\pi$ if and only if $s \in D$ and $(s, b) \notin \pi$ for any $b \in D'$ or $s \in D'$ and $(a, s) \notin \pi$ for any $a \in D$.

- The map $\mathrm{Proj}_{\partial_\Lambda}$ is the metric projection (w.r.t. any $\|\cdot\|_p$) of Λ onto the diagonal, as represented by the point ∂_Λ. Explicitly, for $s=(x,y)$,

$$\mathrm{Proj}_{\partial_\Lambda}(s) = ((x+y)/2, (x+y)/2).$$

The output from solving (2) is a partial matching π between homology classes in D and D' which is optimal with respect to the cost induced by the $\|\cdot\|_p$-norm in $\mathbb{R}^2$, restricted to Λ. By construction, this matching only depends on the coordinates of points in the persistence diagram that correspond to homology classes. Notably, the matching is agnostic to geometric and topological aspects of the data that are not captured in the persistence diagram representations. While the generality of the persistence diagram representation may contribute to its wide applicability, in many contexts this may be a limitation.

Observe that the notation and concepts used to define $d_{\mathrm{W},p}^{\mathrm{PD}}$ are similar to those that appear in the optimal transport framework described above; indeed, the family of persistence diagram distances (2) can be generalised via the language of *partial* OT [17]. We define a *measure persistence diagram* to be a Radon measure μ supported on Λ. Then, any finite persistence diagram $D = \{(b_i, d_i)\}_{i=1}^{|D|}$ determines a measure persistence diagram:

$$\nu_D = \sum_{i=1}^{|D|} \delta_{x_i}, \tag{3}$$

where each δ_{x_i} denotes a Dirac supported at $x_i = (b_i, d_i)$.

Let $\overline{\Lambda} := \Lambda \cup \{\partial_\Lambda\}$ be obtained by adjoining to Λ the virtual point ∂_Λ, with the resulting space endowed with the disjoint union topology. Given two measure persistence diagrams μ and μ', we say a Radon measure π on $\overline{\Lambda}\times\overline{\Lambda}$ is *admissible* if it has marginals μ and μ' (we consider μ and μ' to be measures on $\overline{\Lambda}$ which are supported on Λ) and it satisfies the additional constraint $\pi((\partial_\Lambda, \partial_\Lambda)) = 0$, and write $\pi \in \Pi_{\mathrm{adm}}(\mu,\mu')$. Then for $p \in [1,\infty)$, we define the *p-Wasserstein distance between measure persistence diagrams* [17] as:

$$d_{\mathrm{W},p}^{\mathrm{MPD}}(\mu,\mu') = \inf_{\pi\in\Pi_{\mathrm{adm}}(\mu,\mu')} \left(\int_{\overline{\Lambda}\times\overline{\Lambda}} \|x-x'\|_p^p \,\mathrm{d}\pi(x,x')\right)^{1/p}. \tag{4}$$

If ν_D and $\nu_{D'}$ arise from finite persistence diagrams D and D', respectively, as in (3), then it is not hard to show that $d_{\mathrm{W},p}^{PD}(D,D') = d_{\mathrm{W},p}^{\mathrm{MPD}}(\nu_D,\nu_{D'})$ [17, Prop. 3.2]. In general, $d_{\mathrm{W},p}^{\mathrm{MPD}}$ is finite, provided that we work in the space of measure persistence diagrams μ with *finite p-persistence*, in the sense that

$$\int_\Lambda \|x - \mathrm{Proj}_{\partial_\Lambda}(x)\|_p^p \, d\mu(x) < \infty.$$

In what follows, for any given p, we will restrict to the space of measure persistence diagrams μ with finite p-persistence. With an abuse of notation, we will still indicate their space as MPD.

2.3 Gromov-Wasserstein and co-optimal transport distances

A *metric measure space (mm-space)* [43, 32] is a triple $M = (X, d, \mu)$, where X is a space endowed with a complete separable metric d and with a fully supported Borel measure μ. Given two mm-spaces $M = (X, d, \mu)$ and $M' = (X', d', \mu')$ and a coupling $\pi \in \Pi(\mu,\mu')$ and $p \in [1,\infty)$, we introduce the p-distortion $\mathrm{dis}_p^{\mathrm{GW}}(\pi)$:

$$\mathrm{dis}_p^{\mathrm{GW}}(\pi) = \left(\int_{(X\times X')^2} |d(x,y) - d'(x',y')|^p \,\mathrm{d}\pi(x,x')\,\mathrm{d}\pi(y,y')\right)^{1/p} = \|d - d'\|_{L^p(\pi\otimes\pi)}.$$

By taking the infimum over all feasible couplings, we obtain the *Gromov-Wasserstein (GW) p-distance* between M and M':

$$d_{\mathrm{GW},p}(M,M') := \inf_{\pi\in\Pi(\mu,\mu')} \mathrm{dis}_p^{\mathrm{GW}}(\pi).$$

This quantity is finite if we work in the space of mm-spaces whose distance functions have finite p-th moment; let us make the simplifying convention that we always work with bounded mm-spaces, in order to avoid this technical issue. The GW-distance induces a pseudo-metric on the space of mm-spaces [31, 32], such that $d_{\mathrm{GW},p}(M,M') = 0$ if and only if there is a measure-preserving isometry from M to M'.

For our purposes, we relax the requirement on the function $d : X\times X \to \mathbb{R}$ to be a metric and instead allow any symmetric, measurable and bounded function $k \in L^p_{\mathrm{sym}}(X\times X)$. Then, the structure $M = (X, k, \mu)$ is instead referred to as a gauged measure space [43] or a measure network [10] (in fact, the definition of a measure network in [10] even drops the symmetry condition). This allows us to consider pairwise relations on X to be modelled by a wider class of functions, such as affinity kernels. In this case, the space $(\mathscr{M}, d_{\mathrm{GW},p})$ of equivalence classes of measure networks (under the equivalence relation $M \sim M' \Leftrightarrow d_{\mathrm{GW},p}(M, M') = 0$) is a complete, geodesic, metric space [43, 10].

Recently, a GW-like distance has been developed for comparing hypergraph structures [11], based on ideas from the *co-optimal transport problem* introduced in [45]. Recall that a hypergraph is defined by a pair $H = (V, E)$, where V is a finite set of vertices and E is a set of hyperedges; each $e \in E$ is a subset of V. A general structure which encompasses the notion of a hypergraph is a *measure hypernetwork*, which is a quintuple $H = (X, \mu, Y, \nu, \omega)$, where (X, μ) and (Y, ν) are respectively Polish spaces with fully supported Borel probability measures, and ω is a non-negative, measurable and bounded function $\omega : X \times Y \to \mathbb{R}$. Indeed, from a hypergraph (V, E), one obtains a measure network (V, μ, E, ν, ω), where μ and ν are uniform probability measures and $\omega : V \times E \to \mathbb{R}$ is a binary incidence function $(\omega(v, e) = 1 \Leftrightarrow v \in e)$.

Given two measure hypernetworks $H = (X, \mu, Y, \nu, \omega)$ and $H' = (X', \mu', Y', \nu', \omega')$, the *p-co-optimal distortion* for a pair of couplings $(\pi^v, \pi^e) \in \Pi(\mu, \mu') \times \Pi(\nu, \nu')$ (the notation here is intended to evoke the idea that π^v is a coupling of vertices and π^e is a coupling of hyperedges) is defined to be

$$\mathrm{dis}_p^{\mathrm{COOT}}(\pi^v, \pi^e) := \left(\int_{X \times X' \times Y \times Y'} |\omega(x, y) - \omega'(x', y')|^p \, \mathrm{d}\pi^v(x, x') \, \mathrm{d}\pi^e(y, y') \right)^{1/p} = \|\omega - \omega'\|_{L^p(\pi^v \otimes \pi^e)}.$$

Similarly to the GW distance, the *hypernetwork p-co-optimal transport distance* is given by

$$d_{\mathrm{COOT},p}(H, H') := \inf_{\pi^v \in \Pi(\mu,\mu'),\, \pi^e \in \Pi(\nu,\nu')} \mathrm{dis}_p^{\mathrm{COOT}}(\pi^v, \pi^e). \tag{5}$$

The expression $d_{\mathrm{COOT},p}(H, H')$ defines a pseudo-metric on the space of measure hypernetworks, and induces a metric on the space of equivalence classes of measure hypernetworks up to isomorphism, which is also shown to be geodesic and complete in [11, Theorem 1].

3 Topological Optimal Transport

Having discussed various matching-based comparison approaches for persistence diagrams, point clouds, and hypergraphs, we turn to the aim of this article, which is to develop an optimal transportation framework for matching point clouds that couples geometric and topological information. Our starting point is to represent a point cloud X together with its topological features as a PH-hypergraph $H = H(X, \mathscr{K}, g)$. Given another point cloud and PH-hypergraph, our objective is to find a coupling between them that minimises topological distortion and optimally preserves topological features. A natural attempt to a find a topology-informed matching between point clouds X and X' is to simply apply the HyperCOT framework [11] on their PH-hypergraphs $H = H(X, \mathscr{K}, g)$ and $H' = H'(X', \mathscr{K}', g')$. However, this simplistic approach does not work due to two problems we list below.

- **Problem 1.** PH-hypergraphs do not contain any information on the significance (measured by persistence) of homology classes. One could naturally weight hyperedges by the persistence value of their corresponding homology classes [2]. However, in an optimal transport setting, the HyperCOT would promote "splitting" of mass from hyperedges with higher persistence to hyperedges with lower persistence, rather than matching significant hyperedges.

- **Problem 2.** PH-hypergraphs can be disconnected, and there might be points which do not belong to any hyperedge [2]. In this context, the desired property is to have geometric (spatial) information co-driving the transport plan.

The goal of this section is to develop a framework for geometric cycle matching which addresses these problems.

3.1 Measure topological networks

To solve the problems described above, we first introduce an appropriate general model for the structures that we wish to compare. Here we introduce the following notation. For a locally compact Polish metric space Y, we denote by $\overline{Y}$ the augmented space obtained as the disjoint union $\overline{Y} = Y \cup \{\partial_Y\}$ (with the disjoint union topology), where ∂_Y is an abstract point.

Definition 1 (Measure Topological Network)**.** *A* measure topological network*, or simply* topological network*, is a triple $P = \big((X,k,\mu),(Y,\iota,\nu),\omega\big)$, where*

- *(X,k,μ) is a gauged measure space (see Section 2.3),*
- *Y is a locally compact Polish space,*
- *ν is a Radon measure supported on Y,*
- *$\iota : Y \longrightarrow \Lambda$ is a continuous function,*
- *$\iota_\# \nu$ is a measure persistence diagram (see Section 2.2), and*
- *ω is a measurable and bounded function $\omega : X \times Y \to \mathbb{R}$, so that the quintuple $\big(X,\mu,Y,\nu,\omega\big)$ is a measure hypernetwork (see Section 2.3).*

Let us denote by $\mathscr{P}$ the class of all measure topological networks, with the simplifying conventions that gauged measure spaces have bounded kernels and that measure persistence diagrams have finite ∞-persistence.

Example 1 (Topological Network from a Metric Measure Space)**.** *A measure topological network arises from the data of a PH-hypergraph defined over a metric measure space, and this is inspiration for the definition and the source of all computational examples. Indeed, given a finite metric measure space (X,d,μ), an associated persistence diagram D (obtained via, say, Vietoris-Rips persistent homology) and a choice of a set of generating cycles g for D, we construct an associated topological network as follows. First, we take $Y = D$ as a multiset of points in Λ (one can consider this as a proper set by indexing its points, so that Y can be considered a finite topological space). Taking ν to be the uniform (i.e., counting) measure and $\iota : Y \to \Lambda$ to be inclusion, we have that $\iota_\# \nu = \nu_D$ as in* (3)*. Finally, ω is a kernel representing the hypergraph structure (X,g). In our computational experiments, we use the binary incidence function: for $x \in X$ and a point $(a,b) \in D$, represented by a cycle $c \in g$, $\omega(x,c) = 1$ if and only if x is a vertex of any simplex in c. The flexibility in the definition allows for other hypergraph kernels, such as those studied in [11]. We can also replace the metric d with a symmetric gauge function k that may better encode local geometry without issue—see Section 5 for examples.*

Remark 1. *In full generality, topological measure networks are not restricted to describing topological features of the underlying space. The construction described in Example 1 motivates the use of "topology" in the name.*

Although computational examples of topological networks in this paper will always be constructed as in Example 1, we provide the following additional construction to motivate the level of generality of our definition.

Example 2 (Topological Network from a Curvature Set)**.** *Let (X,d,μ) be a finite metric space. For fixed $k < |X|$, consider the map $d_k : X^k \to \mathbb{R}^{k\times k}$ defined by $d_k(x_1,\ldots,x_k) = (d(x_i,x_j))_{i,j=1}^k$; that is, this map takes a k-tuple of points to its distance matrix. The image of this map is an invariant of X introduced by Gromov in [26], called the kth* curvature set *of X. Given a $k \times k$ distance matrix, one can apply degree-ℓ Vietoris-Rips persistent homology, so that the composition yields a map $D_{k,\ell}$ from X^k into the space of persistence diagrams—this invariant and other related invariants were thoroughly studied in the recent paper [24]. This structure gives rise to a topological network as follows. Let $\widetilde{Y}$ be the multiset consisting of all persistence points arising in diagrams in the image of $D_{k,\ell}$, let Y be its underlying set, let ν be the pushforward of uniform measure on $\widetilde{Y}$ to Y (so that ν counts persistence points with multiplicity), and let $\iota : Y \to \Lambda$ be the inclusion map. We then define $\omega : X \times Y \to \mathbb{R}$ as*

$$\omega(x,y) = \mathbb{P}_{\mu^{\otimes(k-1)}}\big(y \in D_{k,\ell}(x,x_1,\ldots,x_{k-1})\big).$$

We now define a notion of distance between measure topological networks. This distance will be our main object of study throughout the rest of the paper.

Definition 2 (Topological Optimal Transport (TpOT))**.** *Given two topological networks* $P = \big((X,k,\mu),(Y,\iota,\nu),\omega\big)$ *and* $P' = \big((X',k',\mu'),(Y',\iota',\nu'),\omega'\big)$*, we formulate the* Topological Optimal Transport (TpOT) problem *as:*

$$d_{\mathrm{TpOT},p}(P,P') := \inf_{\substack{\pi^\nu\in\Pi(\mu,\mu')\\ \pi^e\in\Pi_{\mathrm{adm}}(\nu,\nu')}} \bigg(\int_{\overline{Y}\times\overline{Y'}} \|\iota(y)-\iota'(y')\|_p^p \,\mathrm{d}\pi^e(y,y') \tag{6}$$

$$+\int_{(X\times X')^2} |k(x,y)-k'(x',y')|^p \,\mathrm{d}\pi^\nu(x,x')\,\mathrm{d}\pi^\nu(y,y') \tag{7}$$

$$+\int_{X\times X'\times\overline{Y}\times\overline{Y'}} |\omega(x,y)-\omega'(x',y')|^p \,\mathrm{d}\pi^\nu(x,x')\,\mathrm{d}\pi^e(y,y')\bigg)^{\frac{1}{p}}. \tag{8}$$

Letting d_p *denote the* ℓ^p*-distance on* $\mathbb{R}^2$*, this can be expressed more concisely as*

$$d_{\mathrm{TpOT},p}(P,P') = \inf_{\substack{\pi^\nu\in\Pi(\mu,\mu')\\ \pi^e\in\Pi_{\mathrm{adm}}(\nu,\nu')}} \Big(\|d_p\circ(\iota,\iota')\|^p_{L^p(\pi^e)} + \|k-k'\|^p_{L^p(\pi^\nu\otimes\pi^\nu)} + \|\omega-\omega'\|^p_{L^p(\pi^\nu\otimes\pi^e)}\Big)^{\frac{1}{p}}. \tag{9}$$

Similar to Section 2.2, we say a Radon measure π on $\overline{Y}\times\overline{Y'}$ is *admissible*, and we write $\pi\in\Pi_{\mathrm{adm}}(\nu,\nu')$, if it has marginals ν and ν', and it satisfies $\pi((\partial_Y,\partial_{Y'}))=0$. Here we are implicitly extending ω and ι to functions $\omega: X\times\overline{Y}\to\mathbb{R}$, $\iota:\overline{Y}\to\overline{\Lambda}$ by setting $\omega_{|X\times\partial_Y}=0$ and $\iota(\partial_Y)=\partial_\Lambda$. The restriction $\omega_{|X\times\partial_Y}$ can be interpreted as (trivial) a membership function (via the map ι) for the diagonal ∂_Λ.

Let us interpret the various components of this definition, when the measure topological spaces arise from PH-hypergraphs, as described in Example 1.

- The coupling π^ν defines a probabilistic matching between the points of the gauged measure spaces (see Figure 1, left column) and the admissible coupling π^e defines a probabilistic matching between generating cycles of the persistence diagrams (see Figure 1, middle column).
- The integral in (6) is a Wasserstein term which measures the quality of the matching of generating cycles. This term is intended to address Problem 1.
- The integral in (7) is a Gromov-Wasserstein term which matches the quality of the matching of points in the gauged measure spaces. This term is intended to address Problem 2.
- The integral in (8) is a co-optimal transport term which addresses the basic problem of matching PH-hypergraphs.

Remark 2. *In applications, it is frequently useful to include tunable weights on the terms of* (9)*, which can be used as hyperparameters to accentuate topology or geometry as is appropriate for a given task. To keep the exposition clean, we suppress these weights from theoretical considerations. They are explicitly included in the exposition of the computational pipeline—see Section 3.3.*

3.2 Metric properties of TpOT

Next we will show that $d_{\mathrm{TpOT,p}}$ defines a metric on the space of (certain equivalence classes of) topological networks. As a first step, we have the following proposition.

Proposition 1. *The infimum of Equation* (9) *is realised.*

Proof of Proposition 1. It is a classic result that the coupling space $\Pi(\mu,\mu')$ is (sequentially) compact in $P(X\times X')$, where $P(X\times X')$ is the set of Borel probability measures on $X\times X'$, see, *e.g.*, [50, Lemma 4.4] and [11, proof of Lemma 24], where we topologise spaces of measures with the weak topology. Similarly, $\Pi_{\mathrm{adm}}(\nu,\nu')$ is sequentially compact in the space $M(\overline{Y}\times\overline{Y'})$ of Radon measures on $\overline{Y}\times\overline{Y'}$ (see [17, proof of Prop 3.1], which extends to Polish spaces as in Definition 1), with the vague topology. These two together imply that $\Pi(\mu,\mu')\times\Pi_{\mathrm{adm}}(\nu,\nu')$ is sequentially compact in $P(X\times X')\times M(\overline{Y}\times\overline{Y'})$ with the product topology.

To complete the proof, we just need to show lower-semicontinuity of the function $C_p : \Pi(\mu,\mu') \times \Pi_{\mathrm{adm}}(\nu,\nu') \to \mathbb{R}$ defined by

$$C_p(\pi,\xi) = \int_{X\times X'\times\overline{Y}\times\overline{Y'}} |\omega-\omega'|^p \,\mathrm{d}(\pi\otimes\xi) + \int_{(X\times X')^2} |k-k'|^p \,\mathrm{d}(\pi\otimes\pi) + \int_{\overline{Y}\times\overline{Y'}} |d_p\circ(\iota,\iota')|^p \,\mathrm{d}\xi. \tag{10}$$

Continuity of each of the second two terms follows exactly as in the proofs of [11, Lemma 24] and [17, Proposition 3.1], respectively. For the first term, we consider a weakly converging $\pi_n^\nu \xrightarrow{\mathrm{w}} \pi^\nu$, and a vaguely converging $\pi_n^e \xrightarrow{\mathrm{v}} \pi^e$. Then, by the same argument as in [17, Proposition 3.1], the sequence $(\pi^\nu\otimes\pi^e)'_n : U \mapsto \int_U |\omega-\omega'|^p \,\mathrm{d}(\pi_n^\nu\otimes\pi_n^e)$ converges to $(\pi^\nu\otimes\pi^e)' : U \mapsto \int_U |\omega-\omega'|^p \,\mathrm{d}(\pi^\nu\otimes\pi^e)$. By using the Pormanteau Theorem ([17, Proposition A.4], which also holds for weak convergence) as in [17, Proposition 3.1], continuity follows. □

Proposition 2. *The function $d_{\mathrm{TpOT},p}$ defines a pseudo-metric on the space of topological networks $\mathscr{P}$.*

To prove Prop. 2 we only need to show that the triangle inequality holds, as symmetry is obvious from the definition of $d_{\mathrm{TpOT},p}$. We follow the usual approach, which relies on the following standard result (see *e.g.* [43, Lemma 1.4]).

Lemma 1 (Gluing Lemma)**.** *Let (X_i,μ_i), $i = 1,\cdots,n$, be Polish probability spaces. Consider couplings $\xi_i \in \Pi(\mu_{i-1}\mu_i)$, for $i = 2,\cdots,n$. There exists a unique probability measure ξ on $X_1\times\cdots\times X_n$ such that $(p_{i-1}\times p_i)_\#\xi = \xi_i$ for every $i = 2,\cdots,n$, where $p_i : X_1\times\cdots\times X_n \longrightarrow X_i$ is the projection on the ith factor.*

The unique measure guaranteed by the lemma will be referred to as the *gluing* of $\xi_2,\ldots,\xi_n$ and will sometimes be denoted as $\xi_2 \boxtimes \xi_3 \boxtimes \cdots \boxtimes \xi_n$. Lemma 1 can be adapted to the case of Radon measures and admissible couplings [17, Prop. 3.2].

Proof of Proposition 2. The proof follows adapts techniques used in the proof of [17, Prop. 3.2] and [11, Thm. 1]. To prove that the triangle inequality holds for $d_{\mathrm{TpOT},p}$, consider three topological networks P_1, P_2 and P_3, and let $(\pi_{12}^\nu,\pi_{12}^e)$ and $(\pi_{23}^\nu,\pi_{23}^e)$ be the couplings realising $d_{\mathrm{TpOT},p}(P_1,P_2)$ and $d_{\mathrm{TpOT},p}(P_2,P_3)$, which exist by Prop. 1. Then, using the gluing lemma, we construct a probability measure π^ν on $X_1\times X_2\times X_3$ with marginals π_{12}^ν and π_{23}^ν on $X_1\times X_2$ and $X_2\times X_3$, respectively. We denote by π_{13}^ν the marginal on $X_1\times X_3$. Similarly, the gluing lemma adapted to Radon measures and admissible couplings yields a measure π^e on $\overline{Y_1}\times\overline{Y_2}\times\overline{Y_3}$ with marginals that agree with π_{ij}^e when restricted to $\overline{Y_i}\times\overline{Y_j}\setminus\{(\partial_{Y_i},\partial_{Y_j})\}$ and that induces a zero cost on $(\partial_{Y_i},\partial_{Y_j})$. Then, we have

$$\begin{aligned}
d_{\mathrm{TpOT},p}(P_1,P_3) &\le \left(\|d_p\circ(\iota_1,\iota_3)\|^p_{L^p(\pi_{13}^e)} + \|k_1-k_3\|^p_{L^p(\pi_{13}^\nu\otimes\pi_{13}^\nu)} + \|\omega_1-\omega_3\|^p_{L^p(\pi_{13}^\nu\otimes\pi_{13}^e)}\right)^{1/p} && (11)\\
&= \left(\|d_p\circ(\iota_1,\iota_3)\|^p_{L^p(\pi^e)} + \|k_1-k_3\|^p_{L^p(\pi^\nu\otimes\pi^\nu)} + \|\omega_1-\omega_3\|^p_{L^p(\pi^\nu\otimes\pi^e)}\right)^{1/p} && (12)\\
&= \left\|\begin{pmatrix} \|d_p\circ(\iota_1,\iota_3)\|_{L^p(\pi^e)} \\ \|k_1-k_3\|_{L^p(\pi^\nu\otimes\pi^\nu)} \\ \|\omega_1-\omega_3\|_{L^p(\pi^\nu\otimes\pi^e)} \end{pmatrix}\right\|_p \\
&\le \left\|\begin{pmatrix} \|d_p\circ(\iota_1,\iota_2)\|_{L^p(\pi^e)} + \|d_p\circ(\iota_2,\iota_3)\|_{L^p(\pi^e)} \\ \|k_1-k_2\|_{L^p(\pi^\nu\otimes\pi^\nu)} + \|k_2-k_3\|_{L^p(\pi^\nu\otimes\pi^\nu)} \\ \|\omega_1-\omega_2\|_{L^p(\pi^\nu\otimes\pi^e)} + \|\omega_2-\omega_3\|_{L^p(\pi^\nu\otimes\pi^e)} \end{pmatrix}\right\|_p && (13)\\
&\le \left\|\begin{pmatrix} \|d_p\circ(\iota_1,\iota_2)\|_{L^p(\pi^e)} \\ \|k_1-k_2\|_{L^p(\pi^\nu\otimes\pi^\nu)} \\ \|\omega_1-\omega_2\|_{L^p(\pi^\nu\otimes\pi^e)} \end{pmatrix}\right\|_p + \left\|\begin{pmatrix} \|d_p\circ(\iota_2,\iota_3)\|_{L^p(\pi^e)} \\ \|k_2-k_3\|_{L^p(\pi^\nu\otimes\pi^\nu)} \\ \|\omega_2-\omega_3\|_{L^p(\pi^\nu\otimes\pi^e)} \end{pmatrix}\right\|_p && (14)\\
&= \left\|\begin{pmatrix} \|d_p\circ(\iota_1,\iota_2)\|_{L^p(\pi_{12}^e)} \\ \|k_1-k_2\|_{L^p(\pi_{12}^\nu\otimes\pi_{12}^\nu)} \\ \|\omega_1-\omega_2\|_{L^p(\pi_{12}^\nu\otimes\pi_{12}^e)} \end{pmatrix}\right\|_p + \left\|\begin{pmatrix} \|d_p\circ(\iota_2,\iota_3)\|_{L^p(\pi_{23}^e)} \\ \|k_2-k_3\|_{L^p(\pi_{23}^\nu\otimes\pi_{23}^\nu)} \\ \|\omega_2-\omega_3\|_{L^p(\pi_{23}^\nu\otimes\pi_{23}^e)} \end{pmatrix}\right\|_p && (15)\\
&= d_{\mathrm{TpOT},p}(P_1,P_2) + d_{\mathrm{TpOT},p}(P_2,P_3).
\end{aligned}$$

Here, we use $\|\cdot\|_p$ to indicate the standard ℓ^p norm on $\mathbb{R}^3$. The inequality (11) follows by suboptimality, while (12) and (15) by marginal conditions. The triangle inequalities in L^p and d_p give (13), and the triangle inequality in ℓ^p gives (14). □

In analogy with measure (hyper)networks [8, 11], we define an equivalence class on $\mathscr{P}$ as follows.

Definition 3 (Weak Isomorphism)**.** *A* weak isomorphism *of topological networks P, P' is a pair of measure-preserving maps $\phi : X \longrightarrow X'$ and $\psi : Y \longrightarrow Y'$ such that $\iota(y) = \iota'(\psi(y))$ for ν-almost every $y \in Y$, $\omega(x, y) = \omega'(\phi(x), \psi(y))$ for $\mu \otimes \nu$-almost every $(x, y) \in X \times Y$, and $k(x_1, x_2) = k'(\phi(x_1), \phi(x_2))$ for $\mu \otimes \mu$-almost every $(x_1, x_2) \in X \times X$. We say that P, P' are* weakly isomorphic *if there exists a topological network $\overline{P}$ and weak isomorphisms from $\overline{P}$ to P and to P'. We use $P \sim_w P'$ to denote that P and P' are weakly isomorphic; this is an equivalence relation on $\mathscr{P}$.*

Proposition 3. *Two topological measure networks $P = \big((X, k, \mu), (Y, \iota, \nu), \omega\big)$ and $P' = \big((X', k', \mu'), (Y', \iota', \nu'), \omega'\big)$ have distance $d_{\mathrm{TpOT},p}(P, P') = 0$ if and only if $P \sim_w P'$.*

Proof. The "if" direction is clear, by the triangle inequality. Conversely, assume $d_{\mathrm{TpOT},p}(P, P') = 0$ and let π^v and π^e be couplings realising this distance (Prop. 1).

We construct $\widetilde{P}$ by setting $\widetilde{X} = X \times X'$, $\widetilde{\mu} = \pi^v$. Similarly, we take $\widetilde{Y} = \big((Y \times Y') \cup (Y \times \partial_{Y'}) \cup (\partial_Y \times Y')\big)$ (where we write $Y \times \partial_{Y'}$ rather than $Y \times \{\partial_{Y'}\}$ for the sake of cleaner notation, and take similar conventions elsewhere), augmented to $\overline{Y} \times \overline{Y'} = \widetilde{Y} \cup \partial_Y \times \partial_{Y'}$, and with measure $\widetilde{\nu} = \pi^e$. Next, we define ϕ, ϕ' and ψ, ψ' as coordinate projection maps.

Given $\widetilde{x}_1 = (x_1, x_1'), \widetilde{x}_2 = (x_2, x_2') \in X \times X'$, we set $\widetilde{k}(\widetilde{x}_1, \widetilde{x}_2) = k(\phi(\widetilde{x}_1), \phi(\widetilde{x}_2)) = k(x_1, x_2)$. By optimality of π^v and since $d_{\mathrm{TpOT},p}(P, P') = 0$, we have that $k(\phi(\widetilde{x}_1), \phi(\widetilde{x}_2)) = k'(\phi'(\widetilde{x}_1), \phi'(\widetilde{x}_2))$ for almost every $\widetilde{x}_1, \widetilde{x}_1$.

Similarly, for $\widetilde{y}_1 = (y_1, y_1'), \widetilde{y}_2 = (y_2, y_2') \in Y \times Y'$ if we define $\widetilde{\omega}(\widetilde{x}, \widetilde{y}) = \omega(\phi(x), \psi(y))$, by optimality of π^v we have $\widetilde{\omega}(\widetilde{x}, \widetilde{y}) = \omega(\phi(x), \psi(y)) = \omega'(\phi'(x), \psi'(y))$ for $\pi^v \otimes \pi^e$-almost every $(\widetilde{x}, \widetilde{y}) \in \widetilde{X} \times \widetilde{Y}$. Finally, we set $\widetilde{\iota} : \widetilde{Y} \longrightarrow \Delta$ as $\widetilde{\iota}(\widetilde{y}) = \iota(\psi(y)) = \iota'(\psi'(y'))$ for almost every $\widetilde{y}$, again by optimality of π^e. This implies that P and P' are weakly isomorphic. □

The work above immediately implies the following.

Theorem 1. *The pseudometric $d_{\mathrm{TpOT},p}$ induces a metric on the quotient space $\mathscr{P}/\sim_w$.*

We abuse notation and also denote by $d_{\mathrm{TpOT},p}$ the induced metric on the quotient space.

3.3 A topology-driven metric on point clouds

We now summarise the metric described above in the discrete setting of point clouds with finite number of points. Consider a point cloud $X = \{x_1, \cdots, x_N\}$ in $\mathbb{R}^n$. Any choice of filtration $\mathscr{K}$ over X yields a corresponding persistence diagram $D = D(X, \mathscr{K})$. Choosing a representative cycle for each homology class in D yields the set of generators g, and the corresponding PH-hypergraph $H = H(X, \mathscr{K}, g)$. We can now endow X with a topological measure hypergraph structure by considering

$$P_X = \big((X, k, \mu_X), (Y, \iota, \nu), \omega_H\big),$$

where k is a gauge function of choice defined on X (for instance, a kernel function, or pairwise distances), μ_X is a chosen measure on X (for instance, uniform), Y is a set with $|D|$ elements (counted with multiplicity), ν and ι are such that $\iota_\# \nu = \nu_D$, where ν_D is as in (3), and ω_H is the binary incidence function for H (see Example 1).

Generating cycles are far from being unique, and there are currently several different algorithms and software available to compute them [2, 5, 14, 15, 19, 16, 29, 34, 49]. As explained in Section 5, here persistent homology computations are performed using the Julia software `Ripserer.jl` [14], which implements the `involutive` algorithm [14, 15] to compute homology and representatives.

Given two point clouds $X = \{x_1, \cdots, x_N\}$ and $X' = \{x_1', \cdots, x_{N'}'\}$ in $\mathbb{R}^n$ endowed with filtrations, let P_X and $P_{X'}$ be their associated topological networks. In this concrete setting, we focus on the $p = 2$ version of the metric. As was described in Remark 2, it is useful in applications to include tunable weights on the terms of TpOT. With this notation, the distance $d_{\mathrm{TpOT},2} = d_{\mathrm{TpOT},\alpha,\beta}$ can be reformulated, using notation similar to that of [10], as

$$d_{\mathrm{TpOT},\alpha,\beta}(P_X, P_{X'})^2 = \min_{\pi^v \in \Pi(\mu,\mu'), \pi^e \in \Pi(\widetilde{\nu},\widetilde{\nu}')} \alpha \left\langle L(C, C'), \pi^v \otimes \pi^v \right\rangle + (1-\alpha) \left\langle \widetilde{C}(D, D'), \pi^e \right\rangle + \beta \left\langle L(\omega_H, \omega_{H'}), \pi^v \otimes \pi^e \right\rangle. \quad (16)$$

The notations used here are as follows:

- All inner products in this expression are Frobenius inner products of matrices.

- The terms C, C' are the pairwise (Euclidean) distance matrices, and D, D' are the persistence diagrams associated to the given filtrations.

- The term $\widetilde{C}(D, D')$ denotes the augmented L^2 cost matrix of dimension $(|D|+1) \times (|D'|+1)$ for persistence diagrams, whose last column and row correspond to transport to and from ∂_Λ respectively, following the diagonal projection π_{∂_Λ} (see [28, Equation 8]).

- The expression $L(C, C')$ denotes the fourth-order distortion tensor, whose term with index (i, j, k, ℓ) is given by $L(C, C')_{ijk\ell} = \frac{1}{2}|C_{ik} - C'_{j\ell}|^2$. The term $\langle L(C, C'), \pi^v \otimes \pi^v \rangle$ can be understood by considering the measure π^v as a matrix, so that the product $\pi^v \otimes \pi^v$ is also naturally identified with a fourth order tensor (with the same dimensions as $L(C, C')$). Then the inner product is computed by isomorphically identifying these fourth order tensors with matrices.

- We denote by ω_H and $\omega_{H'}$ the hypernetwork functions of the PH-hypergraphs H and H', respectively. Specifically, ω_H and $\omega_{H'}$ are represented as matrices of size $N \times |D| + 1$, $N' \times |D'| + 1$ respectively, that are obtained as the hypergraph incidence matrices with an additional 0-row each representing membership for the diagonal edge.

- The expression $L(\omega_H, \omega_{H'})$ is the fourth-order given by:

$$\big(L(\omega_H, \omega_{H'})\big)_{i,k,j,l} = \frac{1}{2}|\omega_{ik} - \omega_{jl}|^2 \qquad \text{for } 1 \le i \le |D|, 1 \le j \le |D'|$$

$$\big(L(\omega_H, \omega_{H'})\big)_{|D|+1,k,j,l} = \frac{1}{2}|\omega_{jl}|^2 \qquad \text{for } 1 \le j \le |D'| \tag{17}$$

$$\big(L(\omega_H, \omega_{H'})\big)_{i,k,|D'|+1,l} = \frac{1}{2}|\omega_{ik}|^2 \qquad \text{for } 1 \le i \le |D|, j = |D'| + 1 \tag{18}$$

$$\big(L(\omega_H, \omega_{H'})\big)_{|D|+1,k,|D'|+1,l} = 0 \tag{19}$$

  In other words, the cost of matching any real edge to the diagonal is given by the L^2 norm of the edge (17 and 17), while matching the diagonal with the diagonal has zero cost (19).

- The parameter $\alpha \in [0, 1]$ controls the tradeoff between the cost of matchings in the persistence diagram space (coefficient $1 - \alpha$) and matchings in terms of the Gromov-Wasserstein distortion (coefficient α).

- The parameter $\beta \in [0, \infty)$ controls the degree to which the geometric and topological matchings are coupled by the hypergraph structure.

The solution of the TpOT problem in 16 involves determining an optimal pair of couplings π^v, π^e. The coupling π^v induces a transport plan between points in the point clouds X, X'. When $\beta > 0, \alpha \in [0, 1]$, this transport balances between preserving topological features and pairwise distances. The coupling π^e induces a matching between homology classes in the persistence diagrams D, D'. For $0 < \alpha < 1$, this matching is driven by the Wasserstein matching, and it is informed by pairwise Euclidean proximity of points forming representative cycles of classes in D, D'. Implementation details are described in Section 5, where we also provide numerical examples.

3.4 Numerical algorithms

We now aim at numerical algorithms for approximating the solution to the TpOT problem in practice. Starting from (16), we consider a regularised variant of the TpOT problem by adding an entropic regularisation to the transport plans (π^e, π^v). Writing

$$\mathsf{L}(\pi^v, \pi^e) = \alpha \langle L(C, C'), \pi^v \otimes \pi^v \rangle + (1 - \alpha) \langle \widetilde{C}(D, D'), \pi^e \rangle + \beta \langle L(\omega_H, \omega_{H'}), \pi^v \otimes \pi^e \rangle,$$

we have

$$\min_{\pi^v \in \Pi(\mu, \mu'),\, \pi^e \in \Pi(\widetilde{\nu}, \widetilde{\nu}')} \mathsf{L}(\pi^v, \pi^e) + \varepsilon_v \,\mathrm{KL}(\pi^v | \mu \otimes \mu') + \varepsilon_e \,\mathrm{KL}(\pi^e | \widetilde{\nu} \otimes \widetilde{\nu}').$$

For $\varepsilon_v, \varepsilon_e > 0$, we can take advantage of the entropic regularisation to utilise fast, smooth optimisation techniques. In particular, we can find a local minimum by projected gradient descent [37, Section 2.3], where both the gradient and the projection are calculated with respect to the KL-divergence. This leads to the following iterative scheme:

$$\begin{aligned} \pi^v_{t+1} &\leftarrow \mathrm{Proj}^{\mathrm{KL}}_{\Pi(\mu,\mu')}\left[-\varepsilon_v^{-1}\nabla_v \mathsf{L}(\pi^v_t, \pi^e_t)(\mu\otimes\mu')\right] \\ \pi^e_{t+1} &\leftarrow \mathrm{Proj}^{\mathrm{KL}}_{\Pi(\widetilde{\nu},\widetilde{\nu}')}\left[-\varepsilon_e^{-1}\nabla_e \mathsf{L}(\pi^v_t, \pi^e_t)(\widetilde{\nu}\otimes\widetilde{\nu}')\right]. \end{aligned} \tag{20}$$

Each of these projections can be calculated by matrix scaling using the Sinkhorn algorithm [36, Chapter 4], and the gradients of L have closed-form expressions:

$$\begin{aligned} \nabla_v \mathsf{L}(\pi^v, \pi^e) &= 2\alpha L(C, C') \otimes \pi^v + \beta L(\omega_H, \omega'_H) \otimes \pi^e \\ \nabla_e \mathsf{L}(\pi^v, \pi^e) &= (1-\alpha)\widetilde{C}[D, D'] + \beta L(\omega_H^\top, {\omega'_H}^\top) \otimes \pi^v. \end{aligned} \tag{21}$$

In the unregularised case when $\varepsilon_v = \varepsilon_e = 0$, an alternating minimisation in (π^v, π^e) can be used to find a local minimum. Fixing π^e, the minimisation problem in $\pi^v \in \Pi(\mu, \mu')$ is

$$\min_{\pi^v \in \Pi(\mu,\mu')} \alpha\langle L(C, C'), \pi^v \otimes \pi^v\rangle + \beta\langle L(\omega_H, \omega_{H'}) \otimes \pi^e, \pi^v\rangle. \tag{22}$$

This falls within the fused Gromov-Wasserstein framework which was introduced and studied in detail by [45, 48]. A local minimum can be found using a conditional gradient method [45, Algorithm 1].

Fixing π^v and minimising in $\pi^e \in \Pi(\widetilde{\nu}, \widetilde{\nu}')$, we have

$$\min_{\pi^e \in \Pi(\widetilde{\nu},\widetilde{\nu}')} \langle M, \pi^e\rangle, \quad M = \pi^v \otimes L(\omega_H, \omega_{H'}) + (1-\alpha)\widetilde{C}(D, D') \tag{23}$$

and we have the identity

$$\langle \pi \otimes L(X, X'), \xi\rangle = \langle L(X^\top, (X')^\top) \otimes \pi, \xi\rangle\rangle = \langle -X^\top \pi (X'), \xi\rangle.$$

This amounts to a standard optimal transport problem.

4 Characterisation of geodesics

Geodesic properties have been extensively studied for the space of measure networks under Gromov-Wasserstein distance [43, 10, 33] and the space of persistence diagrams under Wasserstein distance [47, 6]. In this section, we derive similar results for the TpOT metric when $p = 2$. These results are of theoretical interest, but we plan to explore practical implications in future work; for example, the geodesic structure and curvature properties can be used to study Fréchet means of ensembles of topological networks, extending ideas in [47, 10]. We begin by recalling some definitions from metric geometry—see [4] as a general reference.

4.1 Metric Geometry Concepts

Consider a metric space (X, d). A *geodesic* between points $x, y \in X$ is defined as a path $\gamma : [0,1] \longrightarrow X$ with $\gamma(0) = x$, $\gamma(1) = y$ and such that, for every $0 \le s \le t \le 1$, we have $d(\gamma(s), \gamma(t)) = (t-s)d(x, y)$. In fact, it suffices to show that $d(\gamma(s), \gamma(t)) \le (t-s)d(x, y)$ always holds, as the reverse inequality then follows for free—see, e.g., [7, Lemma 1.3].

We say that (X, d) is a *geodesic* metric space if, for any pair of points $x, y \in X$, there exists a geodesic γ with $\gamma(0) = x$ and $\gamma(1) = y$. Further, we say that (X, d) is *uniquely geodesic* if for any pair of points $x, y \in X$ the geodesic connecting them is unique.

For (X, d) a geodesic space, we say that it has *curvature bounded below by zero* [43, Section 4.2] if, for every geodesic $\gamma : [0,1] \longrightarrow X$ and every point $x \in X$, the following holds:

$$d(\gamma(t), x)^2 \ge (1-t)d(\gamma(0), x)^2 + t d(\gamma(1), x)^2 - t(1-t)d(\gamma(0), \gamma(1))^2, \quad 0 \le t \le 1.$$

Intuitively, this says that geodesic triangles in X are always “thicker" than the corresponding triangles in (flat) Euclidean space.

4.2 Geodesics between measure persistence diagrams

The space of persistence diagrams with the p-Wasserstein distance is well studied, and it is known that it admits geodesics which are essentially linear interpolations of diagrams [47, 6]. What follows is a discussion of existence and uniqueness of geodesics in the case of measure persistence diagrams; these results are a straightforward generalisation of known results for persistence diagrams.

Consider two measure persistence diagrams ν_0 and ν_1, and $\xi \in \Pi_{\mathrm{adm}}(\nu_0, \nu_1)$ a coupling between them. For $t \in [0,1]$ consider the map $\phi_t : \overline{\Lambda} \times \overline{\Lambda} \to \overline{\Lambda}$, $\phi_t(\lambda, \lambda') = (1-t)\lambda + t\lambda'$. Finally, we define

$$\nu_t^{\xi} := (\phi_t)_{\#}\xi. \tag{24}$$

Proposition 4. *Consider measure persistence diagrams ν_0, ν_1, and ξ an optimal coupling that realises $d_{\mathrm{W},p}^{\mathrm{MPD}}(\nu, \nu')$. The path $\gamma(t) = \nu_t^{\xi}$ defines a geodesic between ν_0, ν_1.*

Proof. Set ξ_{st} as the coupling in $\Pi_{\mathrm{adm}}(\nu_s^{\xi}, \nu_t^{\xi})$ given by $\xi_{st} = (\phi_s \times \phi_t)_{\#}\pi^e$, where we use the map

$$\begin{aligned} \phi_s \times \phi_t : (\overline{\Lambda} \times \overline{\Lambda}) &\to (\overline{\Lambda} \times \overline{\Lambda}) \\ (\lambda, \lambda') &\mapsto (\phi_s(\lambda, \lambda'), \phi_t(\lambda, \lambda')). \end{aligned}$$

By sub-optimality, we have that

$$d_{\mathrm{W},p}^{\mathrm{MPD}}(\nu_s^{\xi}, \nu_t^{\xi})^p \le \int_{\overline{\Lambda}^2} \|\phi_s(\overline{\lambda}) - \phi_t(\overline{\lambda})\|_p^p \xi_{st}(d\phi_s(\overline{\lambda}) \times d\phi_t(\overline{\lambda})) = \int_{\overline{\Lambda}^2} \|\phi_s - \phi_t\|_p^p \ \xi_{st},$$

where $\overline{\lambda} = (\lambda, \lambda')$. From the equality

$$\|\phi_s(\overline{\lambda}) - \phi_t(\overline{\lambda})\|_p^p = \|(t-s)\lambda - (t-s)\lambda'\|_p^p$$

we have

$$\int_{\overline{\Lambda}^2} \|\phi_s - \phi_t\|_p^p \ \xi_{st} = (t-s)^p \int_{\overline{\Lambda}^2} \|d_p\|_p^p \ \xi = (t-s)^p d_{\mathrm{W},p}^{\mathrm{MPD}}(\nu_0, \nu_1)^p.$$

Putting everything together, we have $d_{\mathrm{W},p}^{\mathrm{MPD}}(\nu_s^{\xi}, \nu_t^{\xi}) \le (t-s) d_{\mathrm{W},p}^{\mathrm{MPD}}(\nu_0, \nu_1)$, and this completes the proof. □

Remark 3. *Consider two measure persistence diagrams ν_0, ν_1. Take any path between them of the form $\nu_t = (\phi_t)_{\#}\xi$, where ξ is an optimal coupling between ν_0, ν_1, and for every $(\lambda, \lambda') \in \overline{\Lambda} \times \overline{\Lambda}$, the function $t \mapsto \phi_t(\lambda, \lambda')$ traces out a geodesic between λ and λ' in $(\overline{\Lambda}, \|\cdot\|_p)$. Then, the same argument as in Proposition 4 shows that ν_t is a geodesic between ν_0, ν_1. In particular, this shows that $d_{\mathrm{W},1}^{\mathrm{MPD}}$ admits geodesics which are not of the form ϕ_t.*

Remark 4. *This formula for geodesics in the space of measure persistence diagrams takes the same form as the well-known formula for geodesics in classical Wasserstein space—see, e.g., [1, Theorem 7.2.2].*

4.3 Geodesics in the space of measure topological networks

We now focus on geodesics in ${}^{\mathscr{P}}\!/\!{\sim_w}$, using techniques which follow the those employed by Sturm in the Gromov-Wasserstein setting [43]. Let $P = ((X, k, \mu), (Y, \iota, \nu), \omega)$ and $P' = ((X', k', \mu'), (Y', \iota', \nu'), \omega')$ be topological networks and let (π^v, π^e) be a pair of optimal couplings that realise the infimum in Definition 2. We construct the topological network

$$P_t = \big((\widetilde{X}, k_t, \pi^v), (\widetilde{Y}, \iota_t, \pi^e), \omega_t\big), \tag{25}$$

where $\widetilde{X} = X \times X'$, $\widetilde{Y} = \big((Y \times Y') \cup (Y \times \partial_{Y'}) \cup (\partial_Y \times Y')\big)$ are as in Proposition 3 (we once again simplify notation by using expressions such as $Y \times \partial_{Y'}$ rather than $Y \times \{\partial_{Y'}\}$),

$$\begin{aligned} \omega_t((x, x'), (y, y')) &= (1-t)\omega(x, y) + t\omega'(x', y'), \\ k_t((x_1, x_1'), (x_2, x_2')) &= (1-t)k(x_1, x_2) + tk'(x_1', x_2'), \\ \iota_t(y, y') &= (1-t)\iota(y) + t\iota'(y'). \end{aligned}$$

Note that $(\iota_t)_{\#}\pi^e = \nu_t^{(\iota \times \iota')_{\#}\pi^e}$ as in (24). In the following, we use $[P]$ to denote the equivalence class of $P \in \mathscr{P}$ with respect to $\sim_w$.

Theorem 2. *For $P, P' \in \mathscr{P}$, the path $\gamma : [0,1] \to \mathscr{P}/\sim_w$ defined by $\gamma_t = [P_t]$ defines a geodesic between $[P]$ and $[P']$, with respect to $d_{\mathrm{TpOT},p}$.*

Proof. Throughout the proof, let $\pi^v \in \Pi(\mu, \mu')$ and $\pi^e \in \Pi_{\mathrm{adm}}(\nu, \nu')$ be optimal couplings for P, P'. We fix $0 \le s \le t \le 1$, and we use the notation $\widetilde{y} = (y, y') \in \widetilde{Y}$, $\widetilde{x} = (x, x') \in \widetilde{X}$, and $\partial = (\partial_Y, \partial_{Y'})$. Consider now any pair of couplings π, ξ for P_s, P_t. By sub-optimality, we have

$$
\begin{aligned}
d_{\mathrm{TpOT},p}(P_s, P_t)^p \le \int_{\widetilde{X}^2 \times \widetilde{Y}^2} &|\omega_s(\widetilde{x}_1, \widetilde{y}_1) - \omega_t(\widetilde{x}_2, \widetilde{y}_2)|^p \,\mathrm{d}\pi(\widetilde{x}_1, \widetilde{x}_2)\,\mathrm{d}\xi(\widetilde{y}_1, \widetilde{y}_2)) \\
&+ \int_{\widetilde{X}^4} |k_s(\widetilde{x}_1, \widetilde{x}_2) - k_t(\widetilde{x}_3, \widetilde{x_4})|^p \,\mathrm{d}\pi(\widetilde{x}_1, \widetilde{x}_3)\,\mathrm{d}\pi(\widetilde{x}_2, \widetilde{x}_4) \\
&+ \int_{(\widetilde{Y} \cup \partial)^2} \|\iota_s(\widetilde{y}_1) - \iota_t(\widetilde{y}_2)\|_p^p \,\mathrm{d}\xi(\widetilde{y}_1, \widetilde{y}_2).
\end{aligned} \tag{26}
$$

Let $\mathbf{1}_{\pi^v}$ denote the identity coupling of π^v to itself, that is, $\mathbf{1}_{\pi^v} = (\mathrm{diag}_{\widetilde{X}})_\#(\pi^v)$, where $\mathrm{diag}_{\widetilde{X}}$ is the map

$$
\begin{aligned}
\mathrm{diag}_{\widetilde{X}} : \widetilde{X} &\to \widetilde{X} \times \widetilde{X} \\
\widetilde{x} &\mapsto (\widetilde{x}, \widetilde{x}).
\end{aligned}
$$

Similarly, set $\mathbf{1}_{\pi^e} = (\mathrm{diag}_{\widetilde{Y}})_\#(\pi^e)$, where $\mathrm{diag}_{\widetilde{Y}}$ is defined analogously. Note that we have that $\xi_{st} := (\iota_s \times \iota_t)_\# \mathbf{1}_{\pi^e} = (\phi_s \times \phi_t)_\# \pi^e$ as in the proof of Prop. 4.

By the same argument as in the proof of Prop. 4, we have that

$$
\int_{(\widetilde{Y} \cup \partial)^2} \|\iota_s(\widetilde{y}) - \iota_t(\widetilde{y})\|_p^p \,\mathrm{d}\mathbf{1}_{\pi^e}(\widetilde{y}, \widetilde{y}) = (t-s)^p \int_{\overline{Y} \times \overline{Y'}} \|\iota(y) - \iota'(y')\|_p^p \,\mathrm{d}\pi^e(y, y').
$$

Similarly, from the definition of k_s and k_t, we have

$$
|k_s(\widetilde{x}_1, \widetilde{x}_2) - k_t(\widetilde{x}_1, \widetilde{x}_2)|^p = |(t-s)k(x_1, x_2) - (t-s)k'(x_1', x_2')|^p,
$$

and it follows that

$$
\begin{aligned}
\int_{\widetilde{X}^4} &|k_s(\widetilde{x}_1, \widetilde{x}_2) - k_t(\widetilde{x}_1, \widetilde{x}_2)|^p \,\mathrm{d}\mathbf{1}_{\pi^v}(\widetilde{x}_1, \widetilde{x}_1)\,\mathrm{d}\mathbf{1}_{\pi^v}(\widetilde{x}_2, \widetilde{x}_2) \\
&= (t-s)^p \int_{\widetilde{X}^4} |k(x_1, x_2) - k'(x_1', x_2')|^p \,\mathrm{d}\mathbf{1}_{\pi^v}(\widetilde{x}_1, \widetilde{x}_1)\,\mathrm{d}\mathbf{1}_{\pi^v}(\widetilde{x}_2, \widetilde{x}_2) \\
&= (t-s)^p \int_{\widetilde{X}^2} |k(x_1, x_2) - k'(x_1', x_2')|^p \,\mathrm{d}\pi^v(x_1, x_1')\,\mathrm{d}\pi^v(x_2, x_2'),
\end{aligned}
$$

where we have applied the definition of $\mathbf{1}_{\pi^v}$ in the last line. Applying the same arguments to the remaining term in (26), we have

$$
|\omega_s(\widetilde{x}, \widetilde{y}) - \omega_t(\widetilde{x}, \widetilde{y})|^p = |(t-s)\omega(x, y) - (t-s)\omega'(x', y'))|^p
$$

and

$$
\begin{aligned}
\int_{\widetilde{X}^2 \times \widetilde{Y}^2} &|\omega_s(\widetilde{x}, \widetilde{y}) - \omega_t(\widetilde{x}, \widetilde{y})|^p \,\mathrm{d}\mathbf{1}_{\pi^v}(\widetilde{x}, \widetilde{x})\,\mathrm{d}\mathbf{1}_{\pi^e}(\widetilde{y}, \widetilde{y}) \\
&= (t-s)^p \int_{\widetilde{X}^2 \times \widetilde{Y}^2} |\omega(x, y) - \omega'(x', y')|^p \,\mathrm{d}\mathbf{1}_{\pi^v}(\widetilde{x}, \overline{x})\,\mathrm{d}\mathbf{1}_{\pi^e}(\widetilde{y}, \widetilde{y}) \\
&= (t-s)^p \int_{\widetilde{X} \times (\overline{Y} \times \overline{Y'})} |\omega(x, y) - \omega'(x', y')|^p \,\mathrm{d}\pi^v(x, x')\,\mathrm{d}\pi^e(y, y').
\end{aligned}
$$

Putting these together, we have that

$$
\begin{aligned}
d_{\mathrm{TpOT},p}(P_s, P_t)^p &\le (t-s)^p \bigg(\int_{\widetilde{X} \times (\overline{Y} \times \overline{Y'})} |\omega(x, y) - \omega'(x', y')|^p \,\mathrm{d}\pi^v(x, x')\,\mathrm{d}\pi^e(y, y') \\
&\qquad + \int_{\widetilde{X}^2} |k(x_1, x_2) - k'(x_1', x_2')|^p \,\mathrm{d}\pi^v(x_1, x_1')\,\mathrm{d}\pi^v(x_2, x_2') \\
&\qquad + \int_{\overline{Y} \times \overline{Y'}} \|\iota(y) - \iota'(y')\|_p^p \,\mathrm{d}\pi^e(y, y') \bigg) \\
&= \big((t-s)\, d_{\mathrm{TpOT},p}(P_s, P_t)\big)^p,
\end{aligned}
$$

and the result follows.

$\square$

A geodesic of the form described in Equation 25 will be called a *convex geodesic*. A natural question is whether all geodesics in $\mathscr{P}/\sim_w$ are convex. We will show below (Theorem 3) that this is the case for $p=2$. As a first observation, the proof of Theorem 2 easily generalises to give the following corollary. This corollary, in particular, shows that if $p=1$ then there are geodesics which are not convex (c.f. Remark 3).

Corollary 1. *Consider the path Q^t between topological networks Q^0 and Q^1 given by*

$$Q^t = \left((\widetilde{X}, k_t, \pi^v), (\widetilde{Y}, \iota_t, \pi^e), \omega_t\right),$$

where $k_t, \omega_t, \pi^v, \pi^e, \widetilde{X}, \widetilde{Y}$ are as in Prop. 2, and ι_t is such that the map

$$t \mapsto \iota_t((y^0, y^1))$$

defines a geodesic between $\iota^0(y)$ and $\iota^1(y^1)$ in $(\overline{\Lambda}, \|\cdot\|_p)$ as t varies in $[0,1]$ (as in Remark 3) for π^e-almost every $\widetilde{y} = (y^0, y^1) \in \widetilde{Y}$. Then Q^t defines a geodesic $[Q^t]$ in $\mathscr{P}/\sim_w$.

While some of the results below extend to $p>1$, from now on, we focus on the case where $p=2$, and write d_{TpOT} in place of $d_{\mathrm{TpOT},p}$. This is for the sake of simplifying notation, and because we will use $p=2$ in computational implementations and examples below.

Theorem 3. *All geodesics in $\left(\mathscr{P}/\sim_w, d_{\mathrm{TpOT}}\right)$ are convex.*

Proof. The proof follows by adapting techniques from [43, Thm. 3.1] and [6, Thm. 10]. Let $[Q^t]$ be an arbitrary geodesic, with

$$Q^t = \left((X^t, k^t, \mu^t), (Y^t, \iota^t, \nu^t), \omega^t\right)$$

Pick a dyadic decomposition of the unit interval $t_0 = 0, t_1 = \frac{1}{2^n}, \cdots, t_i = \frac{i}{2^n}, \cdots, t_{2^n} = 1$. For each $i \in \{0, \cdots, 2^n\}$, let π_i^v, π_i^e be optimal couplings for $Q^{t_i}, Q^{t_{i-1}}$. Consider the gluings

$$\overline{\pi^v} = \pi_0^v \boxtimes \pi_{\frac{1}{2^n}}^v \boxtimes \cdots \boxtimes \pi_1^v \quad \text{and} \quad \overline{\pi^e} = \pi_0^e \boxtimes \pi_{\frac{1}{2^n}}^e \boxtimes \cdots \boxtimes \pi_1^e$$

(see Lemma 1 and the ensuing discussion) and the couplings

$$\pi^v = (p_0 \times p_1)_\# \overline{\pi^v} \quad \text{and} \quad \pi^e = (p_0 \times p_1)_\# \overline{\pi^e},$$

with $p_i : X^0 \times X^{t_1} \times \cdots \times X^1 \to X^i$ projection on the ith factor. Then, by suboptimality, we have

$$\begin{aligned} d_{\mathrm{TpOT}}(Q^0, Q^1)^2 &\le \|\omega^0 - \omega^1\|^2_{L^2(\pi^v \otimes \pi^e)} + \|k^0 - k^1\|^2_{L^2(\pi^v \otimes \pi^v)} + \|d_2 \circ (\iota^0, \iota^1)\|^2_{L^2(\pi^e)} \\ &=: A + B + C \end{aligned} \tag{27}$$

For any choice of $t \in \{0, 1/2^n, 2/2^n, \cdots, 1\}$, let

$$\xi_t^v = (p_0 \times p_1 \times p_t)_\# \overline{\pi^v} \quad \text{and} \quad \xi_t^e = (p_0 \times p_1 \times p_t)_\# \overline{\pi^e}.$$

We now estimate (27) term-by-term. First observe that

$$\begin{aligned} A = \|\omega^0 - \omega^1\|^2_{L^2(\pi^v \otimes \pi^e)} &= \left\| t\left(\frac{1}{t}(\omega^0 - \omega^t)\right) + (1-t)\left(\frac{1}{1-t}(\omega^t - \omega^1)\right)\right\|^2_{L^2(\xi_t^v \otimes \xi_t^e)} \\ &= \frac{1}{t}\|\omega^0 - \omega^t\|^2_{L^2(\xi_t^v \otimes \xi_t^e)} + \frac{1}{1-t}\|\omega^t - \omega^1\|^2_{L^2(\xi_t^v \otimes \xi_t^e)} - \frac{1}{t(1-t)}\|(1-t)(\omega^0 - \omega^t) - t(\omega^t - \omega^1)\|^2_{L^2(\xi_t^v \otimes \xi_t^e)}, \end{aligned} \tag{28}$$

with the last line following by the general identity

$$|ta + (1-t)b|^2 = t|a|^2 + (1-t)|b|^2 - t(1-t)|a-b|^2,$$

applied to $a = \frac{1}{t}(\omega^0 - \omega^t)$ and $b = \frac{1}{1-t}(\omega^t - \omega^1)$, pointwise. Similarly,

$$B = \frac{1}{t}\|k^0 - k^t\|^2_{L^2(\xi_t^v \otimes \xi_t^v)} + \frac{1}{1-t}\|k^t - k^1\|^2_{L^2(\xi_t^v \otimes \xi_t^v)} - \frac{1}{t(1-t)}\|(1-t)(k^0 - k^t) - t(k^t - k^1)\|^2_{L^2(\xi_t^v \otimes \xi_t^v)} \tag{29}$$

and

$$C = \frac{1}{t}\|\iota^0 - \iota^t\|^2_{L^2(\xi_t^e)} + \frac{1}{1-t}\|\iota^t - \iota^1\|^2_{L^2(\xi_t^e)} - \frac{1}{t(1-t)}\|(1-t)(\iota^0 - \iota^t) - t(\iota^t - \iota^1)\|^2_{L^2(\xi_t^e)}. \tag{30}$$

Recalling that $t = k2^{-n}$ for some k, the first term in (28) satisfies

$$\begin{aligned}
\frac{1}{t}\left\|\omega^0 - \omega^t\right\|^2_{L^2(\xi_t^v \otimes \xi_t^e)} = 2^n \cdot \frac{1}{k}\left\|\omega^0 - \omega^{k2^{-n}}\right\|^2_{L^2(\xi_t^v \otimes \xi_t^e)} &= 2^n \cdot \frac{1}{k}\left\|\sum_{\ell=1}^{k}(\omega^{(\ell-1)2^{-n}} - \omega^{\ell 2^{-n}})\right\|^2_{L^2(\xi_t^v \otimes \xi_t^e)} \\
&\leq 2^n \cdot \frac{1}{k}\left(\sum_{\ell=1}^{k}\left\|\omega^{(\ell-1)2^{-n}} - \omega^{\ell 2^{-n}}\right\|_{L^2(\xi_t^v \otimes \xi_t^e)}\right)^2 \qquad (31) \\
&\leq 2^n \sum_{\ell=1}^{k}\left\|\omega^{(\ell-1)2^{-n}} - \omega^{\ell 2^{-n}}\right\|^2_{L^2(\xi_t^v \otimes \xi_t^e)}, \qquad (32)
\end{aligned}$$

where (31) follows by the triangle inequality for the L^2-norm and (32) is Jensen's inequality. Applying the same argument to the second term of (28), as well as to (29) and (30), it follows that

$$A \leq 2^n \sum_{j=1}^{2^n}\|\omega^{(j-1)2^{-n}} - \omega^{j2^{-n}}\|_{L^2(\pi_j^v \otimes \pi_j^e)} - \frac{1}{t(1-t)}\|(1-t)(\omega^0 - \omega^t) - t(\omega^t - \omega^1)\|^2_{L^2(\xi_t^v \otimes \xi_t^e)}, \tag{33}$$

$$B \leq 2^n \sum_{j=1}^{2^n}\|k^{(j-1)2^{-n}} - k^{j2^{-n}}\|_{L^2(\pi_j^v \otimes pi_j^v)} - \frac{1}{t(1-t)}\|(1-t)(k^0 - k^t) - t(k^t - k^1)\|^2_{L^2(\xi_t^v \otimes \xi_t^v)}, \tag{34}$$

$$C \leq 2^n \sum_{j=1}^{2^n}\|\iota^{(j-1)2^{-n}} - \iota^{j2^{-n}}\|_{L^2(\pi_j^e)} - \frac{1}{t(1-t)}\|(1-t)(\iota^0 - \iota^t) - t(\iota^t - \iota^1)\|^2_{L^2(\xi_t^e)}. \tag{35}$$

From (33),(34), (35), we deduce that

$$\begin{aligned}
d_{\mathrm{TpOT}}(Q^0, Q^1)^2 &\leq A + B + C \\
&\leq \left(\sum_{j=1}^{2^n}\left(\|\omega^{(j-1)2^{-n}} - \omega^{j2^{-n}}\|_{L^2(\pi_j^v \otimes pi_j^e)} + \|k^{(j-1)2^{-n}} - k^{j2^{-n}}\|_{L^2(\pi_j^v \otimes \pi_j^v)} + \|\iota^{(j-1)2^{-n}} - \iota^{j2^{-n}}\|_{L^2(\pi_j^e)}\right)\right) \\
&\quad - \frac{1}{t(1-t)}\Big(\|(1-t)(\omega^0 - \omega^t) - t(\omega^t - \omega^1)\|^2_{L^2(\xi_t^v \otimes \xi_t^e)} + \|(1-t)(\omega^0 - \omega^t) - t(\omega^t - \omega^1)\|^2_{L^2(\xi_t^v \otimes \xi_t^e)} \\
&\qquad + \|(1-t)(\iota^0 - \iota^t) - t(\iota^t - \iota^1)\|^2_{L^2(\xi_t^e)}\Big) \qquad (36) \\
&= d_{\mathrm{TpOT}}(Q^0, Q^1)^2 \\
&\quad - \frac{1}{t(1-t)}\Big(\|(1-t)(\omega^0 - \omega^t) - t(\omega^t - \omega^1)\|^2_{L^2(\xi_t^v \otimes \xi_t^e)} + \|(1-t)(\omega^0 - \omega^t) - t(\omega^t - \omega^1)\|^2_{L^2(\xi_t^v \otimes \xi_t^e)} \\
&\qquad + \|(1-t)(\iota^0 - \iota^t) - t(\iota^t - \iota^1)\|^2_{L^2(\xi_t^e)}\Big), \qquad (37)
\end{aligned}$$

where (37) follows by the assumption that $[Q^t]$ is a geodesic and by the optimality of π_t^e, π_t^v. This inequality implies that π_v, π_e are optimal for Q^0, Q^1, and that

$$\begin{aligned}
&\|(1-t)(\omega^0 - \omega^t) - t(\omega^t - \omega^1)\|^2_{L^2(\xi_t^v \otimes \xi_t^e)} + \|(1-t)(\omega^0 - \omega^t) - t(\omega^t - \omega^1)\|^2_{L^2(\xi_t^v \otimes \xi_t^e)} \\
&\qquad + \|(1-t)(\iota^0 - \iota^t) - t(\iota^t - \iota^1)\|^2_{L^2(\xi_t^e)} = 0
\end{aligned}$$

for every dyadic number t. This implies that $d_{\mathrm{TpOT}}(Q^t, P_t) = 0$ for every dyadic number, with P_t as in Theorem 2. Continuity of the function $t \mapsto Q^t$ and density of dyadic numbers in $[0, 1]$ complete the proof. □

From the characterisation of geodesics provided above, we easily deduce curvature bounds for $\mathscr{P}/\sim_w$.

Theorem 4. *The metric space* $\left(\mathscr{P}/\sim_w, d_{\mathrm{TpOT}}\right)$ *has curvature bounded below by zero.*

Proof. Let $[P^t]$ be a geodesic in $\mathscr{P}/\sim_w$ between P^0 and P^1. By Theorem 3 we can assume that

$$P_t = \left((\widetilde{X}, k_t, \pi^v), (\widetilde{Y}, \iota_t, \pi^e), \omega_t\right),$$

as in (25); in particular, π^v, π^e are optimal for P^0, P^1. Let $P' = \left((X', k', \pi'), (Y', \iota', \nu'), \omega'\right)$ be an arbitrary topological network, and let ξ^v, ξ^e be optimal for P^t and P'. Then we have

$$\begin{aligned}
&d_{\mathrm{TpOT},2}(P^t, P')^2 + t(1-t)d_{\mathrm{TpOT},2}(P^0, P^1)^2 \\
&\quad= \|\omega_t - \omega'\|^2_{L^2(\xi^v\otimes\xi^e)} + \|k_t - k'\|^2_{L^2(\xi^v\otimes\xi^v)} + \|d_2\circ(\iota_t, \iota')\|^2_{L^2(\xi^e)} \\
&\qquad\qquad + t(1-t)\left(\|\omega_0 - \omega_1\|^2_{L^2(\pi^v\otimes\pi^e)} + \|k_0 - k_1\|^2_{L^2(\pi^v\otimes\pi^v)} + \|d_2\circ(\iota_0, \iota_1)\|^2_{L^2(\pi^e)}\right) \\
&\quad= \|\omega_t - \omega'\|^2_{L^2(\xi^v\otimes\xi^e)} + \|k_t - k'\|^2_{L^2(\xi^v\otimes\xi^v)} + \|d_2\circ(\iota_t, \iota')\|^2_{L^2(\xi^e)} \\
&\qquad\qquad + t(1-t)\left(\|\omega_0 - \omega_1\|^2_{L^2(\xi^v\otimes\xi^e)} + \|k_0 - k_1\|^2_{L^2(\xi^v\otimes\xi^v)} + \|d_2\circ(\iota_0, \iota_1)\|^2_{L^2(\xi^e)}\right) \qquad (38)\\
&\quad= (1-t)\left(\|\omega_0 - \omega'\|^2_{L^2(\xi^v\otimes\xi^e)} + \|k_0 - k'\|^2_{L^2(\xi^v\otimes\xi^v)} + \|d_2\circ(\iota_0, \iota')\|^2_{L^2(\xi^e)}\right) \\
&\qquad\qquad + t\left(\|\omega_1 - \omega'\|^2_{L^2(\xi^v\otimes\xi^e)} + \|k_1 - k'\|^2_{L^2(\xi^v\otimes\xi^v)} + \|d_2\circ(\iota_1, \iota')\|^2_{L^2(\xi^e)}\right) \qquad (39)\\
&\quad\geq (1-t)d_{\mathrm{TpOT},2}(P^0, P')^2 + t d_{\mathrm{TpOT},2}(P^1, P')^2, \qquad (40)
\end{aligned}$$

where 38 follows by marginal properties of ξ^v and ξ^e, 39 by definition of w_t, k_t and ν_t, and 40 by suboptimality. □

5 Examples

In this section, we demonstrate our computational framework and the capabilities of TpOT on a range of numerical examples. In what follows, persistent homology computations are performed using the Julia software `Ripserer.jl` [14]. Specifically, we compute the Vietoris-Rips filtration and we use the `involutive` algorithm [14, 15] to compute homology and representatives. We emphasise that the choice of filtration and representatives influences the resulting transport plan and analysis.

5.1 Matching point clouds

In Figure 3, we consider a source point cloud X consisting of four circles of uniform size (Figure 3(A)) and a target point cloud X' resembling a "flower" with four ellipses as "petals" (Figure 3(B)). The point clouds X and X', together with the uniform distribution on points, are then measure spaces. We choose to endow these measure spaces with Gaussian kernels k, k' so that $k_{ij} = K(x_i, x_j)$ and $k'_{ij} = K(x'_i, x'_j)$ where

$$K(x, y) = \exp\left(-\frac{\|x-y\|_2^2}{h}\right),$$

where the bandwidth h is chosen for each set of points $\{x_1, \ldots, x_N\}$ such that $h^{-1}\left(N^{-2}\sum_{ij}\|x_i - x_j\|_2^2\right) = 1$.

In Figure 3(B), points of the target point cloud are coloured by their corresponding source points under the Gromov-Wasserstein matching. Clearly, topological features are not preserved by this matching. Instead, each of the circles in the source point cloud is split up among multiple ellipses in the target point cloud.

The 1-dimensional persistence diagram of the Vietoris-Rips filtration $\mathscr{K}_{\mathrm{VR}}(X)$ contains 4 significant, almost identical, homology classes (one per circle), and the same is true for $\mathscr{K}_{\mathrm{VR}}(X')$ (one class per ellipse). A choice of generating cycle for each class in the diagrams yields the PH-hypergraphs $H_X = H(X, \mathscr{K}_{\mathrm{VR}}(X), g_X)$ and $H_{X'} = H(X', \mathscr{K}_{\mathrm{VR}}(X'), g_{X'})$ and the corresponding topological networks (see Section 3.3)

$$\begin{aligned}
P_X &= \left((X, d_{\mathbb{R}^n}, \mu_X), (Y, \iota, \nu), \omega_H\right) \\
P_{X'} &= \left((X', d_{\mathbb{R}^n}, \mu_{X'}), (Y', \iota', \nu'), \omega_{H'}\right).
\end{aligned}$$

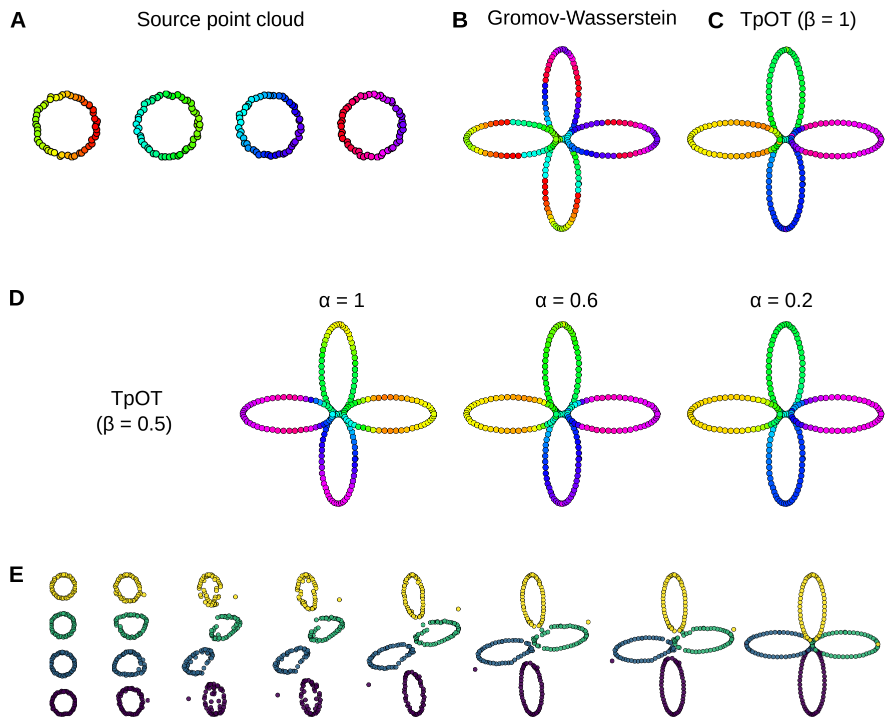

Figure 3: **(A)** Source point cloud: four circles in $\mathbb{R}^2$. **(B)** Target point cloud, with points coloured according to the Gromov-Wasserstein coupling. **(C)** Target point cloud, coloured according to the TpOT coupling with $\alpha = 0.5$ and $\beta = 1$. **(D)** As α varies, TpOT interpolates between a coupling driven by geometry (Gromov-Wasserstein, $\alpha \uparrow 1$) and topological features (persistence diagram matchings, $\alpha \downarrow 0$). **(E)** Geodesic between the point clouds for $\beta = 1$.

Using the algorithms we developed in Section 3.3, we are able to simultaneously find matchings between points and topological features in X and X' as the parameters (α, β) vary.

Intuitively, the parameter α controls the relative contributions of the geometric distortion (measured in terms of a Gromov-Wasserstein loss, with coefficient α) and topological distortion (measured in terms of transport cost on persistence features, with coefficient $1-\alpha$). When $\beta = 0$, point clouds and persistence diagrams are matched independently. For $\beta > 0$, the hypergraph distortion $(\pi^v, \pi^e) \mapsto \langle L(\omega_H, \omega_{H'}), \pi^v \otimes \pi^e \rangle$ couples the matching of the point clouds and persistence diagrams, allowing geometric information to inform the topological matching and vice versa. In the limit as $\beta \to \infty$, this term dominates the overall objective in (16) and in the limit we recover a partial matching variant of HyperCOT [11] in which hyperedges can be transported to a null edge (corresponding to the diagonal in the persistence diagram setting) for a cost equal to the squared L^2-norm of the corresponding incidence function.

Theorem 2 allows us to explicitly construct a geodesic $[P_t]$, $t \in [0,1]$ between P_X and $P_{X'}$.

At each time $t \in [0,1]$, the corresponding gauged measure space P_t can be represented as a point cloud with a gauge function ω_t given by interpolation of ω and ω'. In order to visualise the family of interpolating points, we consider the function $C_t((x,x'),(y,y')) \mapsto (1-t)C(x,y) + tC'(x',y')$ where $C(x,y)$ and $C'(x',y')$ are the squared Euclidean distances on the source and target point clouds respectively; this amounts to linear interpolation of C and C' in $L^2_{\pi^v \otimes \pi^v}((X \times X')^2)$. At each value of t, positions of points in the interpolating point cloud are obtained by applying multidimensional scaling (MDS) algorithm to C_t followed by an alignment step to remove issues due to the invariance of MDS under rigid transformation. Figure 3(E) shows a snapshot of this geodesic computed numerically for $\alpha = 0.5, \beta = 1$. We note that the geodesic almost perfectly recovers a homeomorphisms connecting each loop with their matched petal.

By construction of X, representative cycles for each of the four persistent homology classes span the circles almost entirely, and similarly for X'. This makes the example in Figure 3 a relatively simple one, as most of the points in X and X' are part of "topologically significant" cycles. We next investigate what happens for noisy point clouds which contain topological features with short lifetimes, or "topological noise". Figure 4 shows an example

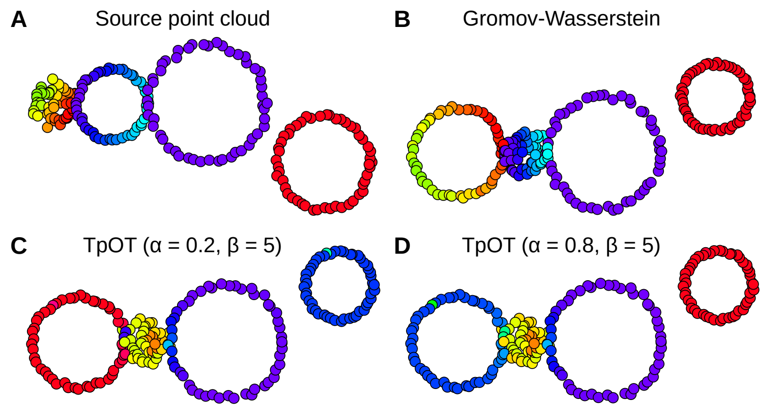

Figure 4: **(A)** Source point cloud and target point cloud have almost identical persistence diagrams. **(B)** Target point cloud, matching induced by GW distance. The colour code represents the matching. **(C)** Target point cloud, matching induced by TpOT with $\alpha = 0.2, \beta = 5$, and **(D)** matching induced by TpOT with $\alpha = 0.8, \beta = 5$.

of transport between point clouds containing noisy disks, and in this case the significant features in each of the persistent diagrams all have different lifespans. For small values of α (e.g. $\alpha = 0.2$, Figure 4(C)) the influence of the topological distortion in the PD-space is greater. As a result, TpOT matches points in cycles with similar lifespan. On the other hand, for higher values (e.g. $\alpha = 0.8$, Figure 4(D)) the geometric distortion has greater influence, and we observe that the matching tends to preserve local proximity between loops, rather than their relative sizes.

A more challenging example is shown in Figure 5. Here the objective is to transport a noisy point cloud M resembling a mug (see Figure 5(A)), into a noisy solid torus T (see Figure 5(B)) whilst preserving the topological features. In this context, matching of persistent homology features amounts to mapping the "handle" in M into an *essential* (i.e. not bounding a disk) closed curve in T. As shown in Figure 5(D), the persistence diagrams of T and M have only one significant class each. That implies that outside of the points involved in the two chosen representative cycles, points in M and T have almost trivial topological signal. For this example, we set $\beta = 5$. For low values of α, the π^v matches points successfully maps the handle in an essential closed curve spanning the doughnut hole (Figure 5(C), left), while for $\alpha = 1$ (Figure 5(C), right) this effect is lost. This can be explained by looking at the matching between cycles induced by π^e, Figure 5(D): for low values of α, the Wasserstein component promotes matching the highly persistent classes together, while for $\alpha = 1$ these are split into less significant classes.

5.2 Geometric cycle matching

The past few years have seen increasing efforts in addressing the problem of matching topological features across different systems, and many solutions have been proposed, including some inspired by techniques in optimal transport [30, 27, 23]. Perhaps, the simplest approach is to straightforwardly use the matching induced by Wasserstein-like distances on persistence diagrams, such as the bottleneck distance[12] and the $d^{\mathrm{PD}}_{\mathrm{W},p}$ Wasserstein distances [6]. More nuanced solutions rely on statistical considerations [13, 38, 21], on the existence of maps between the initial data [3, 25], or, when such a map is unknown, on leveraging the algebraic topology of the PH construction to define a notion of dissimilarity between the underlying complexes [55].

In our case, the topological feature coupling π^e yields a matching between homology classes which is informed by proximity of points creating the corresponding geometric cycles. Figure 6 shows comparisons between the Wasserstein matching on persistence diagrams (PD Wasserstein) and joint geometric-topological matching (TpOT). Similarly to Figure 4, the first examples concerns mapping classes between point clouds X (Figure 6(A)) and X' (Figure 6(B)) having almost identical persistence diagrams (Figure 6(C)). In this case, joint matching of geometric and topological features by TpOT allows us to preserve spatial proximity between representative cycles, see Figure 6(C). This can be particularly important when homology classes in a persistence diagram are indistinguishable, as shown in the example in Figure 6(D-G). Here the TpOT matching (Figure 6(G)) correctly maintains spatial proximity between representative cycles, in contrast with PD Wasserstein matching (Figure 6(F)) which is essentially random.

Finally, we consider the problem of tracking topological features as a system evolves over time. Figure 7(A) shows a piecewise linear spatial curve forming a trefoil knot as it undergoes thermal relaxation. The simulation consists of

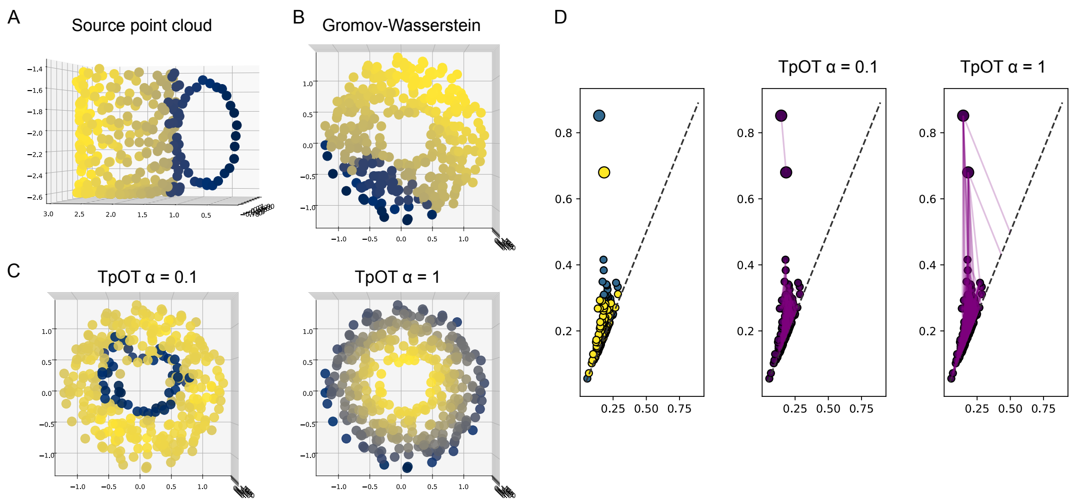

Figure 5: **(A)** Source point cloud: a noisy mug in $\mathbb{R}^3$. **(B)** Target point cloud, a noisy solid torus in $\mathbb{R}^3$. Matching induced by GW distance. The colour code represents the matching. **(C)** Target point cloud, matching induced by TpOT with $\alpha = 0.1$ and $\alpha = 1$. The colour code represents the matching. **(D)** Persistence diagrams for source (blue) and target (yellow) point clouds. Both diagrams contain a unique significant homology class.

20 steps and it was generated using KnotPlot [41]. At each time t_i, the knot takes a spatial configuration C_i, that can be represented as the point cloud X_i of its vertices. At each step, we interpret X_i as the topological network obtained by computing 1-dimensional PH and PH-hypergraph. We can then match homology classes between consecutive persistence diagrams D_i, D_{i+1} using PD Wasserstein distance and d_{TpOT}, respectively, see Figure 7(B). The TpOT matching significantly improves the accuracy of generator matchings compared to PD Wasserstein, as shown in Figure 7(B). Since the ground truth correspondence of points in the curve is known, in Figure 7(C) we show the correlation between generators and their images under the PD Wasserstein as well as TpOT matchings. We find that the TpOT matching provides a clear improvement over the PD Wasserstein matchings, which appears to be close to random.

Remark 5. *Consider two point clouds, and the convex GW geodesic connecting them (see [43]). Consider now the Vietoris-Rips filtration along the geodesic path. A natural question is whether the path induced by the persistent diagrams follows a convex geodesic. A counterexample to this question is given by Figure 3, as the GW-mapping between points suggests that the GW-geodesic "breaks" the loops, and thus, the four persistent homology classes. This can be verified by computing the path of Vietoris-Rips PDs obtained by inputting the various interpolation metrics. The same question applies for $TpOT$-geodesics. In this context, the answer depends on whether the GW-term contributes to cycle matching or not. When it doesn't (e.g. Figure 3(E)), persistent diagrams computed from Vietoris-Rips filtration along the TpOT-geodesic do indeed follow the convex-geodesic in the persistent diagram space. When GW contributes (e.g. Figure 4 and Figure 6), then the persistent diagram path is non geodesic.*

6 Conclusion

Inspired by recent advances in computational topology [2] and optimal transport for structured objects [48, 39, 11], we propose to encode geometric and topological information about a point cloud jointly in a hypergraph structure. Our approach lays the foundation for a topology-driven analysis of the geometry of gauged measure spaces. As in the co-optimal setting [39, 11], the TpOT problem outputs a pair of coupled transport plans. The first one provides a matching of the underlying metric spaces (point clouds) that preservers topological features. The second one matches the topological features (cycles representing persistent homology classes) in a way that maintains local metric structure. This geometric cycle matching represents a key novelty of our method, and, to the best of our knowledge, it is the first one taking into account the geometric nature of data, and the spatial interconnectivity of generating cycles.

Constructing PH-hypergraphs, and thus the TpOT pipeline, depends on computing explicit cycles representing

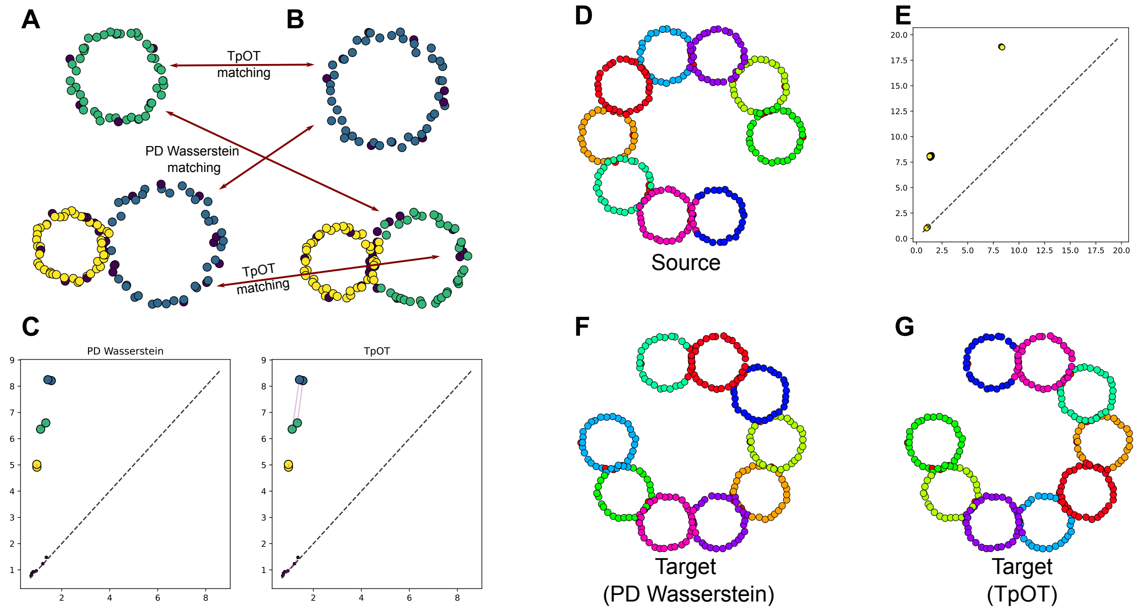

Figure 6: **(A)** A point cloud X consisting of 3 noisy loops of various sizes. **(B)** A point cloud X consisting of the same 3 noisy loops as X, placed in different ways. Points forming the three persistent cycles in $\mathcal{K}_{\mathrm{VR}}(X)$ and $\mathcal{K}_{\mathrm{VR}}(X)$ are coloured in blue, green, and yellow. **(C)** Persistence diagrams of $\mathcal{K}_{\mathrm{VR}}(X)$ and $\mathcal{K}_{\mathrm{VR}}(X)$ plotted super-imposed to each other and showing Wasserstein (left) and Geometric (right) matchings. The three highly persistent classes are coloured as the corresponding cycles in (A) and (B). Geometric and Wasserstein matchings are shown with arrows in (A) and (B) as well. **(D)** A point cloud X_1 consisting of a chain of 9 noisy circles. **(E)** Persistence diagrams of X_1 super-imposed to that of X_2 (point cloud in (F-G)). Colour indicates points forming the 9 second-most persistent cycles in each point cloud. **(F)** The result of cycle matching induced by Wasserstein distance. **(G)** The result of geometric cycle matching. Colour indicates matched cycles.

homology classes. It is well known that, even in the case of simplicial complexes, generating cycles are far from being unique, and that different choices can potentially bring important biases in the subsequent analysis [29]. Given their potential and demonstrate usefulness, the search for new methods of producing “good” representatives, and the research into homology generators in general [5, 16, 19, 29, 34, 49, 2], is a very active sub-field of topological data analysis. New, important advancements in this area currently provide a wide variety of algorithms to compute representatives, that can all be implemented in the TpOT pipeline, depending on the specific goal or application. In addition to that, the experimentally shown robustness of the hyperTDA method [2] suggests that, when dealing with systems sufficiently complex from a topological point of view, the qualitative behaviour of a TpOT induced matching should be robust by changes to the representative cycles.

Data availability

Data and software used to produce results in this paper are available at the GitHub repository `https://github.com/zsteve/TPOT`

Acknowledgments

SYZ gratefully acknowledge funding from the Australian Government Research Training Program and an Elizabeth and Vernon Puzey Scholarship. MPHS is funded through the University of Melbourne DRM initiative, through an ARC Laureate Fellowship (FL220100005) and acknowledges financial support from the Volkswagen Foundation through a “Life?” program grant. TN was partially supported by NSF grants DMS 2107808 and DMS 2324962.

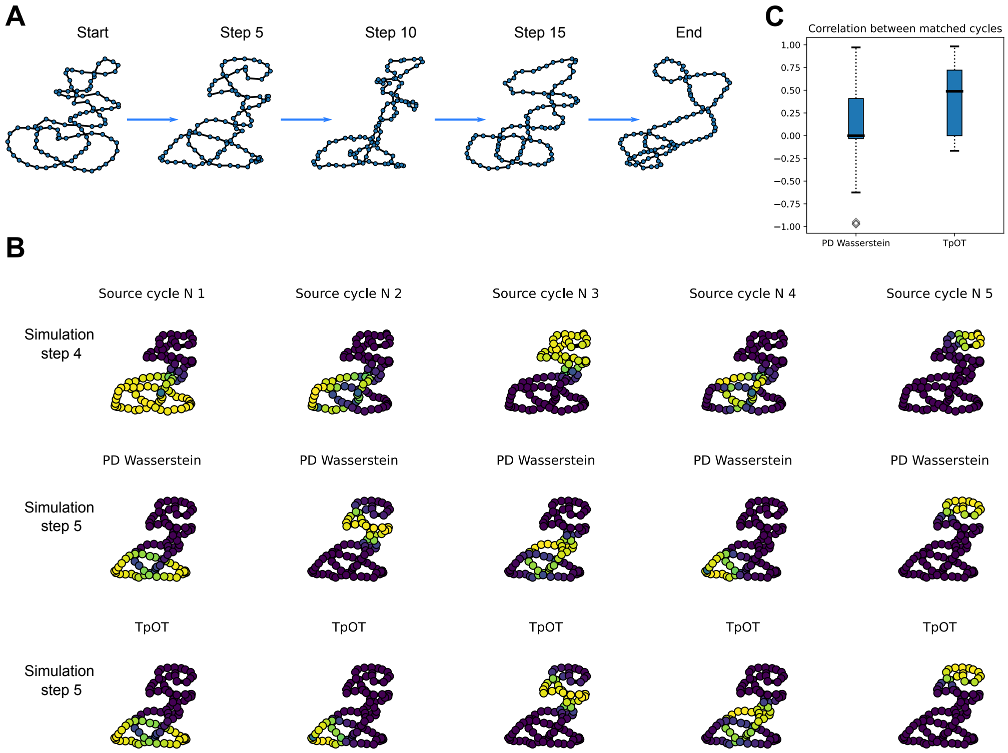

Figure 7: **(A)** Snapshots of a trefoil undergoing thermal relaxation. **(B)** Top rows shows the trefoil at step 4 of the simulation, while mid and bottom rows show the trefoil at step 5. Each column correspond to one of the most persistent homology classes of the trefoil at step 4. Mid and bottom rows shows the matched class at step 5, according to Wasserstein distance (mid) and TpOT (bottom). **(C)** Box plot of correlation values between cycles and their images.

Declarations

The authors declare no conflict of interest or competing interests.

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.

[2] Agnese Barbensi, Hee Rhang Yoon, Christian Degnbol Madsen, Deborah O Ajayi, Michael PH Stumpf, and Heather A Harrington. Hypergraphs for multiscale cycles in structured data. *arXiv preprint arXiv:2210.07545*, 2022.

[3] Ulrich Bauer and Michael Lesnick. Induced matchings and the algebraic stability of persistence barcodes. *arXiv preprint arXiv:1311.3681*, 2013.

[4] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33. American Mathematical Society, 2022.

[5] Chao Chen and Daniel Freedman. Measuring and computing natural generators for homology groups. *Computational Geometry*, 43:169–181, 02 2010.

[6] Samir Chowdhury. Geodesics in persistence diagram space. *arXiv preprint arXiv:1905.10820*, 2019.

[7] Samir Chowdhury and Facundo Mémoli. Explicit geodesics in gromov-hausdorff space. *arXiv preprint arXiv:1603.02385*, 2016.

[8] Samir Chowdhury and Facundo Mémoli. The Gromov–Wasserstein distance between networks and stable network invariants. *Information and Inference: A Journal of the IMA*, 8(4):757–787, 2019.

[9] Samir Chowdhury, David Miller, and Tom Needham. Quantized gromov-wasserstein. In *Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part III 21*, pages 811–827. Springer, 2021.

[10] Samir Chowdhury and Tom Needham. Gromov-Wasserstein averaging in a Riemannian framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 842–843, 2020.

[11] Samir Chowdhury, Tom Needham, Ethan Semrad, Bei Wang, and Youjia Zhou. Hypergraph co-optimal transport: Metric and categorical properties. *Journal of Applied and Computational Topology*, pages 1–60, 2023.

[12] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. In *Proceedings of the twenty-first annual symposium on Computational geometry*, pages 263–271, 2005.

[13] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Dmitriy Morozov. Persistent homology for kernels, images, and cokernels. In *Proceedings of the twentieth annual ACM-SIAM symposium on Discrete algorithms*, pages 1011–1020. SIAM, 2009.

[14] Matija Čufar. Ripserer.jl: flexible and efficient persistent homology computation in julia. *Journal of Open Source Software*, 5(54):2614, 2020.

[15] Matija Čufar and Žiga Virk. Fast computation of persistent homology representatives with involuted persistent homology. *arXiv preprint arXiv:2105.03629*, 2021.

[16] Tamal K Dey, Tao Hou, and Sayan Mandal. Persistent 1-cycles: Definition, computation, and its application. In *International Workshop on Computational Topology in Image Context*, pages 123–136. Springer, 2019.

[17] Vincent Divol and Théo Lacombe. Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport. *Journal of Applied and Computational Topology*, 5:1–53, 2021.

[18] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

[19] Kevin Emmett, Benjamin Schweinhart, and Raul Rabadan. Multiscale topology of chromatin folding. *arXiv preprint arXiv:1511.01426*, 2015.

[20] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, et al. Pot: Python optimal transport. *J. Mach. Learn. Res.*, 22(78):1–8, 2021.

[21] Inés García-Redondo, Anthea Monod, and Anna Song. Fast topological signal identification and persistent cohomological cycle matching. *arXiv preprint arXiv:2209.15446*, 2022.

[22] Robert Ghrist. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

[23] Chad Giusti and Darrick Lee. Signatures, lipschitz-free spaces, and paths of persistence diagrams. *SIAM Journal on Applied Algebra and Geometry*, 7(4):828–866, 2023.

[24] Mario Gómez and Facundo Mémoli. Curvature sets over persistence diagrams. *arXiv preprint arXiv:2103.04470*, 2021.

[25] Rocío Gonzalez-Diaz and Manuel Soriano-Trigueros. Basis-independent partial matchings induced by morphisms between persistence modules. *arXiv preprint arXiv:2006.11100*, 2020.

[26] Mikhael Gromov, Misha Katz, Pierre Pansu, and Stephen Semmes. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152. Springer, 1999.

[27] Yasuaki Hiraoka, Yusuke Imoto, Killian Meehan, Théo Lacombe, and Toshiaki Yachimura. Topological node2vec: Enhanced graph embedding via persistent homology. *arXiv preprint arXiv:2309.08241*, 2023.

[28] Théo Lacombe, Marco Cuturi, and Steve Oudot. Large scale computation of means and clusters for persistence diagrams using optimal transport. *Advances in Neural Information Processing Systems*, 31, 2018.

[29] Lu Li, Connor Thompson, Gregory Henselman-Petrusek, Chad Giusti, and Lori Ziegelmeier. Minimal cycle representatives in persistent homology using linear programming: An empirical study with user's guide. *Frontiers in Artificial Intelligence*, 4, 2021.

[30] Mingzhe Li, Xinyuan Yan, Lin Yan, Tom Needham, and Bei Wang. Flexible and probabilistic topology tracking with partial optimal transport. *arXiv preprint arXiv:2302.02895*, 2023.

[31] Facundo Mémoli. On the use of Gromov-Hausdorff Distances for Shape Comparison. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007.

[32] Facundo Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11:417–487, 2011.

[33] Facundo Mémoli and Zhengchao Wan. Characterization of gromov-type geodesics. *Differential Geometry and its Applications*, 88:102006, 2023.

[34] Ippei Obayashi. Volume optimal cycle: Tightest representative cycle of a generator on persistent homology. *SIAM J. Appl. Algebra Geom.*, 2:508–534, 2018.

[35] Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6:1–38, 2017.

[36] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[37] Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-wasserstein averaging of kernel and distance matrices. In *International conference on machine learning*, pages 2664–2672. PMLR, 2016.

[38] Yohai Reani and Omer Bobrowski. Cycle registration in persistent homology with applications in topological bootstrap. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5579–5593, 2022.

[39] Ievgen Redko, Titouan Vayer, Rémi Flamary, and Nicolas Courty. Co-optimal transport. *arXiv preprint arXiv:2002.03731*, 2020.

[40] Manish Saggar, Olaf Sporns, Javier Gonzalez-Castillo, Peter A Bandettini, Gunnar Carlsson, Gary Glover, and Allan L Reiss. Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature communications*, 9(1):1399, 2018.

[41] Robert G Scharein and Kellogg S Booth. Interactive knot theory with knotplot. In *Multimedia Tools for Communicating Mathematics*, pages 277–290. Springer, 2002.

[42] Søren S Sørensen, Christophe AN Biscio, Mathieu Bauchy, Lisbeth Fajstrup, and Morten M Smedskjaer. Revealing hidden medium-range order in amorphous materials using topological data analysis. *Science Advances*, 6(37):eabc2320, 2020.

[43] Karl-Theodor Sturm. *The Space of Spaces: Curvature Bounds and Gradient Flows on the Space of Metric Measure Spaces*, volume 290. American Mathematical Society, 2023.

[44] Thomas Thorne, Paul DW Kirk, and Heather A Harrington. Topological approximate bayesian computation for parameter inference of an angiogenesis model. *Bioinformatics*, 38(9):2529–2535, 2022.

[45] Vayer Titouan, Nicolas Courty, Romain Tavenard, and Rémi Flamary. Optimal transport for structured data with application on graphs. In *International Conference on Machine Learning*, pages 6275–6284. PMLR, 2019.

[46] Vayer Titouan, Ievgen Redko, Rémi Flamary, and Nicolas Courty. Co-optimal transport. *Advances in neural information processing systems*, 33:17559–17570, 2020.

[47] Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry*, 52:44–70, 2014.

[48] Titouan Vayer, Laetitia Chapel, Rémi Flamary, Romain Tavenard, and Nicolas Courty. Fused gromov-wasserstein distance for structured objects. *Algorithms*, 13(9):212, 2020.

[49] Matija Čufar and Žiga Virk. Fast computation of persistent homology representatives with involuted persistent homology, 2021.

[50] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

[51] Oliver Vipond, Joshua A Bull, Philip S Macklin, Ulrike Tillmann, Christopher W Pugh, Helen M Byrne, and Heather A Harrington. Multiparameter persistent homology landscapes identify immune cell spatial patterns in tumors. *Proceedings of the National Academy of Sciences*, 118(41):e2102166118, 2021.

[52] Larry Wasserman. Topological data analysis. *Annual Review of Statistics and Its Application*, 5:501–532, 2018.

[53] Hongteng Xu, Dixin Luo, and Lawrence Carin. Scalable gromov-wasserstein learning for graph partitioning and matching. *Advances in neural information processing systems*, 32, 2019.

[54] Hongteng Xu, Dixin Luo, Hongyuan Zha, and Lawrence Carin Duke. Gromov-wasserstein learning for graph matching and node embedding. In *International conference on machine learning*, pages 6932–6941. PMLR, 2019.

[55] Hee Rhang Yoon, Robert Ghrist, and Chad Giusti. Persistent extensions and analogous bars: data-induced relations between persistence barcodes. *Journal of Applied and Computational Topology*, pages 1–47, 2023.

[56] Afra Zomorodian. Topological data analysis. *Advances in applied and computational topology*, 70:1–39, 2012.

[57] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.