A Geometric Condition for Uniqueness of Fréchet Means of Persistence Diagrams

Yueqi Cao[1] and Anthea Monod[1,†]

1 Department of Mathematics, Imperial College London, UK

† Corresponding e-mail: a.monod@imperial.ac.uk

Abstract

The Fréchet mean is an important statistical summary and measure of centrality of data; it has been defined and studied for persistent homology captured by persistence diagrams. However, the complicated geometry of the space of persistence diagrams implies that the Fréchet mean for a given set of persistence diagrams is generally not unique, which prohibits theoretical guarantees for empirical means with respect to population means. In this paper, we derive a variance expression for a set of persistence diagrams exhibiting a multi-matching between the persistence points known as a grouping. Moreover, we propose a condition for groupings, which we refer to as flatness: sets of persistence diagrams that exhibit flat groupings give rise to unique Fréchet means. Together with recent results from Alexandrov geometry, this allows for the first derivation of a finite sample convergence rate for sets of persistence diagrams that exhibit flat groupings.

Keywords: Alexandrov geometry; Fréchet means; persistent homology; nonnegative curvature.

1 Introduction

Persistent homology is an important methodology from topological data analysis which has gained rapid interest and application over the past recent decades and by now has been widely implemented in many applications across diverse scientific domains. Given that its primary purpose is to summarize topological and geometric aspects of data—specifically, it captures the "shape" and "size" of a given dataset—studying statistical aspects of persistent homology is a crucial task to make topological data analysis a valid approach for data analysis.

The space of persistence diagrams is a viable setting for statistics and probability, satisfying conditions for the existence of important statistical and probabilistic quantities, such as means, variances, and probability measures (Mileyko et al., 2011). This paper studies the mean, in particular, which is perhaps the most fundamental statistic that captures the central tendency of data and provides an understanding of what we expect to see on average for a data generating process. The Fréchet mean is a generalization of the usual algorithmic mean to general metric spaces and has been previously defined and studied for sets of persistence diagrams (Turner, 2013). Significant results on Fréchet means for sets of persistence diagrams were provided by Turner et al. (2014), which include an algorithm for its computation along with the only known convergence result for Fréchet means to date. However, this result is valid only in quite restrictive settings, and most importantly, under the assumption of uniqueness of the Fréchet mean. Due to the complicated geometry of the space of persistence diagrams—in particular, it is an nonnegatively curved Alexandrov space (Turner et al., 2014) with the implication that geodesics are not even locally unique—it is far from clear that the Fréchet mean should ever be unique. This lack of a condition for uniqueness prohibits a comprehensive convergence analysis for empirical Fréchet means of persistence diagrams computed from real datasets (Cao and Monod, 2022). The practical implication of a lack of convergence guarantee is that it is difficult, if not impossible, to draw conclusions about the behavior of the general distribution and population from observed samples. We are thus restricted to only making descriptive and exploratory observations with Fréchet means computed from sampled data and cannot inference on the general behavior of the data generating process and the general unseen population with any theoretical guarantees.

In this paper, we propose a geometric condition on sets of persistence diagrams that guarantees uniqueness of Fréchet means. In particular, we consider a multi-matching representation between persistence points

known as a *grouping* (Munch et al., 2015) and derive a variance expression for groupings. Further, we propose a geometric condition on groupings, which we refer to as *flatness*, and show that flat groupings give rise to unique Fréchet means. Using recent computational and statistical results on Alexandrov spaces by Le Gouic et al. (2019), we then derive the first finite sample convergence rate for empirical Fréchet means to population means for sets of persistence diagrams that exhibit flat groupings.

The remainder of this paper is organized as follows. In Section 2, we provide background and details on persistent homology and metric geometry, and in particular, the metric geometry of persistence diagrams and the space of persistence diagrams. In Section 3, we recall definitions of a grouping and a Fréchet mean, which are our specific objects of interest in this paper. Here, we also present our contributions of an expression for the variance of general groupings, our proposed notion of flatness of groupings, and prove uniqueness of Fréchet means for sets of persistence diagrams for which there exist flat groupings. Section 4 presents the first finite sample convergence rate for empirical Fréchet means of persistence diagrams exhibiting flat groupings to population means. We close in Section 5 with a discussion of our findings and some ideas for future research based on our contributions in this paper.

2 Background: Persistent Homology, Metric Geometry, and Metric Geometry of Persistent Homology

In this section, we provide background and details on our setting and objects of study: persistent homology, which gives rise to persistence diagrams, and the space of all persistence diagrams. We also review some concepts from metric geometry that will be essential for our study and construction of our results.

2.1 Persistent Homology

The standard pipeline of persistent homology begins with a filtration, which is a nested sequence of topological spaces: $\mathcal{M}_0 \subseteq \mathcal{M}_1 \subseteq \cdots \subseteq \mathcal{M}_n = \mathcal{M}$. By applying the homology functor $H(\cdot)$ with coefficients in a field, we have the sequence of homology vector spaces $H(\mathcal{M}_0) \to H(\mathcal{M}_1) \to \cdots \to H(\mathcal{M}_n)$. The collection of vector spaces $H(\mathcal{M}_i)$, together with vector space homomorphisms $H(\mathcal{M}_i) \to H(\mathcal{M}_j), i < j$, is called a *persistence module*. When each $H(\mathcal{X}_i)$ is finite dimensional, a persistence module can be decomposed into a direct sum of irreducible summands called *interval modules*, which correspond to birth and death times of homology classes (Chazal et al., 2016). The collection of birth–death intervals $[\epsilon_i, \epsilon_j)$ are called *barcodes* and they represent the *persistent homology* of the filtration of $\mathcal{M}$. Each interval can also be identified as the coordinate of a point in the plane $\mathbb{R}^2$. In this way we have an alternate representation known as a *persistence diagram*. For a detailed introduction of persistence homology, see e.g., Edelsbrunner et al. (2008); Edelsbrunner and Harer (2010).

Definition 1. A *persistence diagram* D is a locally finite multiset of points in the half-plane $\Omega = \{(x, y) \in \mathbb{R}^2 \mid x < y\}$ together with points on the diagonal $\partial\Omega = \{(x, x) \in \mathbb{R}^2\}$ counted with infinite multiplicity. Points in Ω are called *off-diagonal points*. The persistence diagram with no off-diagonal points is called the *empty persistence diagram*, denoted by $D_\emptyset$.

The geometry and statistical properties of the space of persistence diagrams are the main focus of this paper.

2.2 Metric Geometry

We now outline essential concepts from metric geometry and refer to Burago et al. (2001) for a comprehensive and detailed discussion.

Let $(\mathcal{S}, d)$ be an arbitrary metric space. For any two points $x, y \in \mathcal{S}$, a *geodesic* connecting x and y is a continuous curve $\gamma : [a, b] \to \mathcal{S}$ such that for any $a \leq s \leq t \leq b$,

$$d(\gamma(s), \gamma(t)) = \frac{t - s}{b - a} d(x, y).$$

$(\mathcal{S}, d)$ is called a geodesic space if any two points can be joined by a geodesic. A geodesic space is an *Alexandrov space with nonnegative curvature* if for every triangle $\{x_0, x_1, y\} \subseteq \mathcal{S}$ and a geodesic $\gamma : [0,1] \to \mathcal{S}$ connecting x_0 and x_1 there exists an isometric triangle $\{\tilde{x}_0, \tilde{x}_1, \tilde{y}\}$ in $\mathbb{R}^2$ such that $d(y, \gamma(t)) \geq \|\tilde{y} - \tilde{\gamma}(t)\|$, where $\tilde{\gamma}(t) = t\tilde{x}_1 + (1-t)\tilde{x}_0$ is the line segment joining $\tilde{x}_0$ and $\tilde{x}_1$ in the Euclidean plane.

Given $z \in \mathcal{S}$, let Γ_z be the set of all geodesics emanating from z. For any two geodesics $\gamma_0, \gamma_1 \in \Gamma_z$, the *Alexandrov angle* $\angle_z(\gamma_0, \gamma_1)$ is defined by

$$\angle_z(\gamma_0, \gamma_1) = \lim_{s,t \to 0} \cos^{-1}\left(\frac{d^2(z, \gamma_0(t)) + d^2(z, \gamma_1(s)) - d^2(\gamma_0(t), \gamma_1(s))}{2d(z, \gamma_0(t))d(z, \gamma_1(s))}\right).$$

If $(\mathcal{S}, d)$ is an Alexandrov space with nonnegative curvature then $\angle_z : \Gamma_z \times \Gamma_z \to [0, \pi]$ is well-defined and a pseudo-metric on Γ_z. Therefore $\angle_z$ defined a metric on the quotient space $\Gamma_z / \sim$ where $\gamma_0 \sim \gamma_1$ if and only if $\angle_z(\gamma_0, \gamma_1) = 0$.

The completion $(\widehat{\Gamma}_z, \angle_z)$ of $(\Gamma_z / \sim, \angle_z)$ is called the space of directions. Let v_γ denote the direction of γ at z, i.e., the equivalence class of γ in $\widehat{\Gamma}_z$. The tangent cone $T_z\mathcal{S}$ is defined as $\widehat{\Gamma}_z \times \mathbb{R}_+ / \sim$ where $(v_\gamma, t) \sim (v_\eta, s)$ if and only if $t = s = 0$ or $(v_\gamma, t) = (v_\eta, s)$. Let $[v_\gamma, t]$, $[v_\eta, s] \in T_z\mathcal{S}$ be two tangent vectors, then define

$$\mathrm{C}_z([v_\gamma, t], [v_\eta, s]) = \sqrt{s^2 + t^2 - 2st\cos\angle_z(v_\gamma, v_\eta)}. \tag{1}$$

C_z is a metric on $T_z\mathcal{S}$ called *the cone metric*.

For a geodesic space, the *logrithmic map (log map) at* z $\log_z : \mathcal{S} \to T_z\mathcal{S}$ assigns x to $[v_\gamma, d(z,x)]$ where γ is a geodesic from z to x. The log map is a multimap since there can be different geodesics from z to x. By selecting an arbitrary direction for every x the log map is a well-defined map, and moreover $\log_z$ can be chosen to be measurable with respect to the Borel algebra of $T_z\mathcal{S}$ (Le Gouic et al., 2019).

2.3 Metric Geometry of Persistence Diagram Space

The collection of all persistence diagrams may be viewed as a space; in particular, it is a metric space and hence its metric geometry may be studied. Although there exist various possible metrics on the space of persistence diagrams, we focus on the following.

Definition 2. For any two persistence diagrams D_1 and D_2, define the *2-Wasserstein distance* by

$$\mathrm{W}_2(D_1, D_2) = \inf_{\phi}\left(\sum_{x \in D_1} \|x - \phi(x)\|^2\right)^{\frac{1}{2}}$$

where ϕ ranges over all bijections between D_1 and D_2, and $\|\cdot\|$ denotes the 2-norm on $\mathbb{R}^2$. The *total persistence* of a persistence diagram D is defined as $\mathrm{W}_2(D, D_\emptyset)$. Let $\mathcal{D}_2$ be the set of all persistence diagrams with finite total persistence.

Under the 2-Wasserstein distance, we now discuss several metric geometric characteristics of the space of persistence diagrams. We have that $(\mathcal{D}_2, \mathrm{W}_2)$ is an Alexandrov space with nonnegative curvature (Turner et al., 2014); the curvature behavior is largely determined by the boundary, see Figure 1. Moreover, we have the following characterization of geodesics between persistence diagrams: Let D_1 and D_2 be two persistence diagrams with finite total persistence and $\phi : D_1 \to D_2$ be an optimal matching, then the geodesic $\gamma : [0,1] \to \mathcal{D}_2$ joining D_1 to D_2 is such that $\gamma(t)$ is in fact a persistence diagram with points of the form $(1-t)x + t\phi(x)$ where x ranges all points from D_1.

We also have the following characterization concerning tangent vectors. Let $D \in \mathcal{D}_2$ be a persistence diagram. A tangent vector in the tangent cone $T_D\mathcal{D}_2$ can be represented as a set of vectors $\{v_i \in \mathbb{R}^2, i \in I\} \cup \{v_j \in \mathbb{R}^2, j \in J\}$ where I is the index set of off-diagonal points in D and J is the index set of vectors perpendicular to the diagonal such that $\sum_{i \in I} \|v_i\|^2 + \sum_{j \in J} \|v_j\|^2 < \infty$. Note that there may exist tangent vectors with no corresponding geodesics.

Example 3. Consider the persistence diagram $D = \{x_n = \left(0, \frac{1}{n^2}\right), n \in \mathbb{N}\}$. At each x_n, assign the vector $v_n = \left(\frac{1}{n}, -\frac{1}{n}\right)$. We claim that the collection $V = \{v_n, n \in \mathbb{N}\}$ is an element in $T_D\mathcal{D}_2$. In fact, let $V_N = \{\tilde{v}_n =$

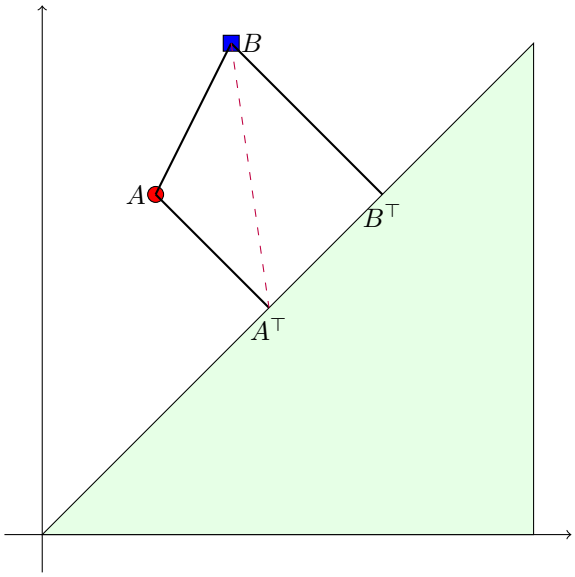

Figure 1: Curvature is determined by the boundary. Consider three persistence diagrams: diagram D_A with a single off-diagonal point A, diagram D_B with a single off-diagonal point B, and the empty diagram $D_\emptyset$ with no off-diagonal point. The three edges of triangle $\triangle D_\emptyset D_A D_B$ are plotted with solid lines. For the comparison triangle, given $\mathrm{W}_2(D_A, D_\emptyset) = \|AA^\top\|$, $\mathrm{W}_2(D_A, D_B) = \|AB\|$, and $\angle D_\emptyset D_A D_B = \angle A^\top AB$, the length of the third edge is $\|A^\top B\|$. We see that $\|A^\top B\| > \|BB^\top\| = \mathrm{W}_2(D_B, D_\emptyset)$, indicating nonnegative Alexsandrov curvature.

$v_n, 1 \le n \le N\} \cup \{\tilde{v}_n = 0, n > N\}$. The geodesic $\gamma_N(t) = \{x_n + t\tilde{v}_n, n \in \mathbb{N}\}$ is well-defined for $t \in [0, \frac{1}{2N}]$. For $1 \le N \le M$,

$$\begin{aligned}
\cos\angle_D(V_N, V_M) &= \lim_{s,t\to 0} \frac{\mathrm{W}_2^2(\gamma_N(t), D) + \mathrm{W}_2^2(\gamma_M(s), D) - \mathrm{W}_2^2(\gamma_N(t), \gamma_M(s))}{2\mathrm{W}_2(\gamma_N(t), D)\mathrm{W}_2(\gamma_M(s), D)} \\
&\ge \lim_{s,t\to 0} \frac{\sum_{n=1}^N \frac{2t^2}{n^2} + \sum_{n=1}^M \frac{2s^2}{n^2} - \sum_{n=1}^N \frac{2(t-s)^2}{n^2} - \sum_{n=N+1}^M \frac{2s^2}{n^2}}{2\sqrt{\sum_{n=1}^N \frac{2t^2}{n^2}}\sqrt{\sum_{n=1}^M \frac{2s^2}{n^2}}} = \frac{\sqrt{\sum_{n=1}^N \frac{1}{n^2}}}{\sqrt{\sum_{n=1}^M \frac{1}{n^2}}}
\end{aligned}$$

Hence under the cone metric,

$$\begin{aligned}
\mathrm{C}_D^2(V_N, V_M) &= \|V_N\|^2 + \|V_M\|^2 - 2\|V_N\|\|V_M\|\cos\angle_D(V_N, V_M) \\
&\le 2\sum_{n=1}^N \frac{1}{n^2} + 2\sum_{n=1}^M \frac{1}{n^2} - 2\sqrt{2\sum_{n=1}^N \frac{1}{n^2}} \cdot \sqrt{2\sum_{n=1}^M \frac{1}{n^2}} \cdot \frac{\sqrt{\sum_{n=1}^N \frac{1}{n^2}}}{\sqrt{\sum_{n=1}^M \frac{1}{n^2}}} \\
&= 2\sum_{n=N+1}^M \frac{1}{n^2} \to 0, \quad N, M \to \infty
\end{aligned}$$

Therefore, $\{V_N\}_{N\in\mathbb{N}}$ is a Cauchy sequence in $T_D\mathcal{D}_2$ and converges to V. However, V is not a tangent vector of any geodesic emanating from D as for any fixed t, $x_n + tv_n \notin \Omega$ when n is sufficiently large.

3 Groupings of Persistence Diagrams and their Fréchet Means

Let $(\mathcal{S}, d)$ be a metric space and μ be a (Borel) probability measure on $\mathcal{S}$. The Fréchet function is defined by

$$F(x) = \int_{\mathcal{S}} d^2(x, y)\, \mathrm{d}\mu(y).$$

If $F(x)$ is finite for some (hence, every) x, the probability measure μ is said to have finite second moment. The quantity $\mathbb{V} = \inf_{x \in \mathcal{S}} F(x)$ is the *variance* of μ. The set of points achieving the variance is the *Fréchet mean or expectation*.

Fréchet means for sets of persistence diagrams exist, given that a probability measure on $(\mathcal{D}_2, \mathrm{W}_2)$ has finite second moment and compact support (Mileyko et al., 2011; Turner et al., 2014). However, Fréchet means are not necessarily unique due to the nonnegative curvature of $(\mathcal{D}_2, \mathrm{W}_2)$. The lack of unique Fréchet means is problematic in many practical applications as well as theoretical settings—for example, averaging time-varying persistence diagrams (Munch et al., 2015) and establishing convergence of empirical Fréchet mean of persistence diagrams (Cao and Monod, 2022)—however, approximations for Fréchet means are computable.

Let $D_1, \ldots, D_L$ be a finite set of persistence diagrams with finite off-diagonal points, and $\mu = \frac{1}{L}\sum_{i=1}^{L} \delta_{D_i}$ be a discrete probability measure. The Fréchet function for this set of persistence diagrams is

$$F(D) = \frac{1}{L}\sum_{i=1}^{L} \mathrm{W}_2^2(D, D_i). \tag{2}$$

Turner et al. (2014) proposed a greedy algorithm to compute local minima of the Fréchet function (2). Other work by Lacombe et al. (2018) in more general contexts has also given rise to alternative algorithms to compute Fréchet means for persistence diagrams. Munch et al. (2015) introduced probabilistic Fréchet means to average time-varying persistence diagrams. We now recall and rephrase some definitions and results from some of this prior work which will be useful for our study.

Definition 4. For a finite set of persistence diagrams $D_1, \ldots, D_L$, each with $k_1, \ldots, k_L$ off-diagonal points, a *grouping* G is a $K \times L$ matrix where $K = k_1 + \cdots + k_L$. The jth column G^j consists of k_j off-diagonal points of D_j and $K - k_j$ copies of the diagonal $\partial\Omega$. Each row is called a *selection*. A trivial selection is a row with all $\partial\Omega$ entries.

Intuitively, a grouping is a matching of points between persistence diagrams. For the special case with $L = 2$, a grouping is equivalent to a bijective matching between two persistence diagrams, with each selection representing the one-to-one correspondence between points. In general cases, a grouping is thus a representation of multi-matching, i.e., any two columns of the grouping induce a bijective matching between corresponding persistence diagrams.

For any $x \in \Omega$, let $x^\top$ be the projection to the diagonal, and $x^\perp = x - x^\top$. Set $\|x - \partial\Omega\| = \|x^\perp\|$ and $\|\partial\Omega - \partial\Omega\| = 0$. Let $Q = \{x_1, \ldots, x_L\}$ be a multiset of off-diagonal points and copies of the diagonal. If $Q \subseteq \Omega$ consists of off-diagonal points only, the mean point $\bar{Q}$ is the usual algorithmic mean $\bar{Q} = \frac{1}{L}\sum_{i=1}^{L} x_i$. If $Q = \{x_1, \ldots, x_s\} \cup \{\partial\Omega, \ldots, \partial\Omega\}$ consists of $(L - s)$ copies of the diagonal, set $Q_o = \{x_1, \ldots, x_s\}$, and the mean point is then given by

$$\bar{Q} = \frac{s\bar{Q}_o + (L - s)(\bar{Q}_o)^\top}{L}. \tag{3}$$

If $Q = \{\partial\Omega, \ldots, \partial\Omega\}$, then $\bar{Q} = \partial\Omega$.

Definition 5. Let $\{D_1, \ldots, D_L\}$ be a set of persistence diagrams and G be a grouping of size $K \times L$. The *mean persistence diagram* $\mathrm{mean}(G)$ is the diagram where each off-diagonal point is given by $\bar{G}_i$ for each nontrivial selection G_i. The *variance* of G is defined as

$$\mathbb{V}(G) = \frac{1}{L}\sum_{j=1}^{L}\sum_{i=1}^{K} \|G_i^j - \bar{G}_i\|^2. \tag{4}$$

The following theorem establishes the relation between Fréchet means and groupings of persistence diagrams.

Theorem 6. *(Turner et al., 2014, Theorem 3.3) Given a finite set of persistence diagrams $D_1, \ldots, D_L$, if $D_\star$ is a Fréchet mean then $D_\star = \text{mean}(G_\star)$ for some grouping $G_\star$, and the optimal matching between $D_\star$ and each $D_i, i = 1, \ldots, L$ is induced by $G_\star$.*

This result allows us to consider the Fréchet variance as the minimal variance of groupings,

$$\sigma^2 = \min_D \frac{1}{L} \sum_{i=1}^{L} \mathrm{W}_2(D, D_i)^2 = \min_G \mathbb{V}(G).$$

Optimal groupings are groupings that achieve the Fréchet variance. For a general grouping we derive the following variance expression.

Theorem 7. *Let $\{D_1, \ldots, D_L\}$ be a set of persistence diagrams, and G be a grouping of size $K \times L$. Let s_i be the number of off-diagonal points in the ith row of G. The variance of G is*

$$\mathbb{V}(G) = \frac{1}{L^2} \sum_{i=1}^{K} \sum_{1 \le w < \ell \le L} \|G_i^w - G_i^\ell\|^2 + \sum_{i=1}^{K} \frac{L - s_i}{L^2 s_i} \left(\sum_{1 \le w < \ell \le s_i} \|(G_i^{j_w})^\top - (G_i^{j_\ell})^\top\|^2 \right), \tag{5}$$

where $G_i^{j_\ell}, \ell = 1, \ldots, s_i$ ranges over all off-diagonal points in the ith row of G. If $s_i = 0$, the summand is taken to be 0.

Proof. Let G_i be the ith row with $s_i > 0$ off-diagonal points. Then the variance for G_i is

$$\begin{aligned}
\mathbb{V}(G_i) =& \frac{1}{L} \left(\left(\sum_{\ell=1}^{s_i} \|G_i^{j_\ell} - \bar{G}_i\|^2 \right) + (L - s_i) \|\bar{G}_i - \partial\Omega\|^2 \right) \\
=& \frac{1}{L} \left(\left(\sum_{\ell=1}^{s_i} \|G_i^{j_\ell}\|^2 \right) - 2 \left\langle \sum_{\ell=1}^{s_i} G_i^{j_\ell}, \bar{G}_i \right\rangle + s_i \|\bar{G}_i^\top\|^2 + L \|\bar{G}_i^\perp\|^2 \right).
\end{aligned} \tag{6}$$

Note that

$$\bar{G}_i^\top = \frac{1}{s_i} \sum_{\ell=1}^{s_i} (G_i^{j_\ell})^\top, \quad \bar{G}_i^\perp = \frac{1}{L} \sum_{\ell=1}^{s_i} (G_i^{j_\ell})^\perp$$

and

$$\left\langle \sum_{\ell=1}^{s_i} G_i^{j_\ell}, \bar{G}_i \right\rangle = \left\langle \sum_{\ell=1}^{s_i} (G_i^{j_\ell})^\top, \bar{G}_i^\top \right\rangle + \left\langle \sum_{\ell=1}^{s_i} (G_i^{j_\ell})^\perp, \bar{G}_i^\perp \right\rangle.$$

Substituting these expressions into (6), we obtain

$$\mathbb{V}(G_i) = \frac{1}{L} \sum_{\ell=1}^{s_i} \|G_i^{j_\ell}\|^2 - \frac{s_i}{L} \|\bar{G}_i^\top\|^2 - \|\bar{G}_i^\perp\|^2. \tag{7}$$

For the last two terms, we have

$$\begin{aligned}
\frac{s_i}{L} \|\bar{G}_i^\top\|^2 + \|\bar{G}_i^\perp\|^2 =& \frac{1}{L s_i} \sum_{w=1}^{s_i} \sum_{\ell=1}^{s_i} \langle (G_i^{j_w})^\top, (G_i^{j_\ell})^\top \rangle + \frac{1}{L^2} \sum_{w=1}^{s_i} \sum_{\ell=1}^{s_i} \langle (G_i^{j_w})^\perp, (G_i^{j_\ell})^\perp \rangle \\
=& \frac{1}{L s_i} \sum_{w=1}^{s_i} \sum_{\ell=1}^{s_i} \langle G_i^{j_w}, G_i^{j_\ell} \rangle - \frac{L - s_i}{L^2 s_i} \sum_{w=1}^{s_i} \sum_{\ell=1}^{s_i} \|(G_i^{j_w})^\perp\| \|(G_i^{j_\ell})^\perp\| \\
=& \frac{1}{2 L s_i} \sum_{w=1}^{s_i} \sum_{\ell=1}^{s_i} \left(\|G_i^{j_w}\|^2 + \|G_i^{j_\ell}\|^2 - \|G_i^{j_w} - G_i^{j_\ell}\|^2 \right) - \frac{L - s_i}{L^2 s_i} \left(\sum_{\ell=1}^{s_i} \|(G_i^{j_\ell})^\perp\| \right)^2 \\
=& \frac{1}{L} \sum_{\ell=1}^{s_i} \|G_i^{j_\ell}\|^2 - \frac{1}{L s_i} \sum_{1 \le w < \ell \le s_i} \|G_i^{j_w} - G_i^{j_\ell}\|^2 - \frac{L - s_i}{L^2 s_i} \left(\sum_{\ell=1}^{s_i} \|(G_i^{j_\ell})^\perp\| \right)^2,
\end{aligned} \tag{8}$$

where we used the fact that $(G_i^{j_w})^\perp$ and $(G_i^{j_\ell})^\perp$ are parallel and in the same direction. Combining (7) and (8), we have

$$\mathbb{V}(G_i) = \frac{1}{Ls_i} \sum_{1\le w<\ell\le s_i} \|G_i^{j_w} - G_i^{j_\ell}\|^2 + \frac{L-s_i}{L^2 s_i}\Big(\sum_{\ell=1}^{s_i} \|(G_i^{j_\ell})^\perp\|\Big)^2. \tag{9}$$

We expand the first term as follows:

$$\begin{aligned}
&\frac{1}{Ls_i} \sum_{1\le w<\ell\le s_i} \|G_i^{j_w} - G_i^{j_\ell}\|^2 = \big(\frac{1}{2L^2} + \frac{L-s_i}{2L^2 s_i}\big) \sum_{w=1}^{s_i}\sum_{\ell=1}^{s_i} \|G_i^{j_w} - G_i^{j_\ell}\|^2 \\
=&\frac{1}{2L^2}\Big(\sum_{w=1}^{s_i}\sum_{\ell=1}^{s_i} \|G_i^{j_w} - G_i^{j_\ell}\|^2 + \sum_{w=1}^{s_i}\sum_{\ell=s_i+1}^{L} \|G_i^{j_w} - \partial\Omega\|^2 + \sum_{w=s_i+1}^{L}\sum_{\ell=1}^{s_i} \|G_i^{j_\ell} - \partial\Omega\|^2\Big) \\
&- \frac{L-s_i}{L^2}\sum_{\ell=1}^{s_i} \|G_i^{j_\ell} - \partial\Omega\|^2 + \frac{L-s_i}{2L^2 s_i}\sum_{w=1}^{s_i}\sum_{\ell=1}^{s_i} \|G_i^{j_w} - G_i^{j_\ell}\|^2 \\
=&\frac{1}{L^2} \sum_{1\le w<\ell\le L} \|G_i^{j_w} - G_i^{j_\ell}\|^2 - \frac{L-s_i}{L^2}\sum_{\ell=1}^{s_i} \|G_i^{j_\ell} - \partial\Omega\|^2 + \frac{L-s_i}{2L^2 s_i}\sum_{w=1}^{s_i}\sum_{\ell=1}^{s_i} \|G_i^{j_w} - G_i^{j_\ell}\|^2.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\mathbb{V}(G_i) - \frac{1}{L^2} \sum_{1\le w<\ell\le L} \|G_i^{j_w} - G_i^{j_\ell}\|^2 = \\
=&-\frac{L-s_i}{L^2}\sum_{\ell=1}^{s_i} \|G_i^{j_\ell} - \partial\Omega\|^2 + \frac{L-s_i}{2L^2 s_i}\sum_{w=1}^{s_i}\sum_{\ell=1}^{s_i} \|G_i^{j_w} - G_i^{j_\ell}\|^2 + \frac{L-s_i}{L^2 s_i}\Big(\sum_{\ell=1}^{s_i} \|(G_i^{j_\ell})^\perp\|\Big)^2 \\
=&\frac{L-s_i}{L^2 s_i}\Big(\frac{1}{2}\sum_{w=1}^{s_i}\sum_{\ell=1}^{s_i} \|G_i^{j_w} - G_i^{j_\ell}\|^2 + \sum_{w=1}^{s_i}\sum_{\ell=1}^{s_i} \|(G_i^{j_w})^\perp\|\|(G_i^{j_\ell})^\perp\| - \sum_{w=1}^{s_i}\sum_{\ell=1}^{s_i} \|(G_i^{j_w})^\perp\|^2\Big) \\
=&\frac{L-s_i}{L^2 s_i}\sum_{w=1}^{s_i}\sum_{\ell=1}^{s_i}\Big(\frac{1}{2}\|G_i^{j_w} - G_i^{j_\ell}\|^2 + \|(G_i^{j_w})^\perp\|\|(G_i^{j_\ell})^\perp\| - \frac{1}{2}\|(G_i^{j_w})^\perp\|^2 - \frac{1}{2}\|(G_i^{j_\ell})^\perp\|^2\Big) \\
=&\frac{L-s_i}{L^2 s_i} \sum_{1\le w<\ell\le s_i} \|(G_i^{j_w})^\top - (G_i^{j_\ell})^\top\|^2.
\end{aligned}$$

Finally, summing $\mathbb{V}(G_i)$ for all rows, we obtain (5). □

Notice that if we disregard the diagonal $\partial\Omega$ and suppose G is a grouping of points in the plane $\mathbb{R}^2$, then the variance of G only consists of the first term in (5). The diagonal contributes the second term in the variance expression.

The derived variance expression motivates the following definition.

Definition 8. A grouping G is called *flat* if there exists $\lambda > 0$ such that

1. For each nontrivial selection G_i, the diameter is bounded above by λ, i.e., $\|G_i^w - G_i^\ell\| < \lambda$ for all $w, \ell = 1, \dots, L$;
2. For two distinct selections G_i, G_j, the distance between G_i and G_j is bounded below by λ, i.e., $\|G_i^w - G_j^\ell\| > \lambda$ for all $w, \ell = 1, \dots, L$;
3. Off-diagonal points are bounded away from the diagonal by λ, i.e., $\|G_i^w - \partial\Omega\| > \lambda$ for $G_i^w \neq \partial\Omega$.

A visual example of flatness is illustrated in Figure 2.

Given this notion of flatness, we now have a condition that gives rise to unique Fréchet means of persistence diagrams.

Theorem 9. *Let $\{D_1, \dots, D_L\}$ be a set of persistence diagrams. If there exists a flat grouping $G_\star$ for $\{D_1, \dots, D_L\}$, then* $\mathrm{mean}(G_\star)$ *is the unique Fréchet mean of $\{D_1, \dots, D_L\}$.*

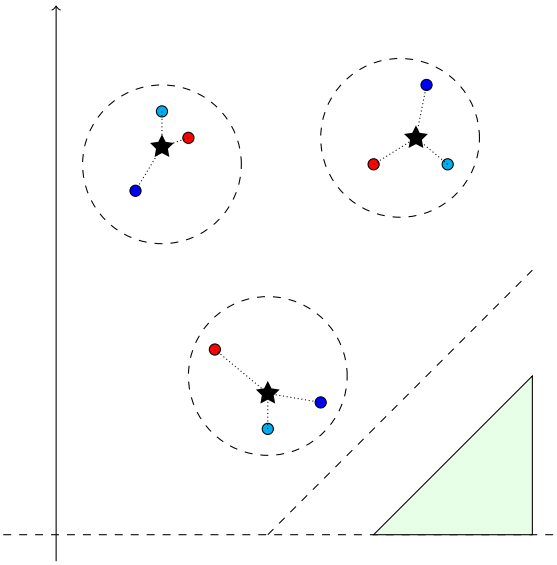

Figure 2: An example of flat groupings. The off-diagonal points of $D_{\text{red}}, D_{\text{blue}}, D_{\text{cyan}}$ are distributed as three clusters over the half-plane Ω. Every dashed circle indicates a selection of the grouping. The Fréchet mean is given by D_{black}.

Proof. Suppose $G_\star$ is a flat grouping. By conditions 1 and 3, each nontrivial selection of G does not contain the diagonal. Thus, the variance of $G_\star$ is

$$\mathbb{V}(G_\star) = \frac{1}{L^2} \sum_{i=1} \sum_{1 \le w < \ell \le L} \|(G_\star)_i^w - (G_\star)_i^\ell\|^2.$$

Let G be any grouping. By (5), we have

$$\mathbb{V}(G) \ge \frac{1}{L^2} \sum_{1 \le w < \ell \le L} \sum_{i=1} \|G_i^w - G_i^\ell\|^2.$$

Now, fix any two columns w and ℓ. Without loss of generality, we may assume $(G_\star)_i^w = G_i^w$ (otherwise, we may apply row permutation to achieve this form). By conditions 2 and 3,

$$\min\{\|G_i^w - G_i^\ell\|, \|G_i^w - \partial\Omega\|\} > \lambda > \|G_i^w - (G_\star)_i^\ell\|$$

for any $G_i^\ell \neq (G_\star)_i^\ell$. Therefore, $\mathbb{V}(G) > \mathbb{V}(G_\star)$ if $G \neq G_\star$. Thus mean$(G_\star)$ is the unique Fréchet mean. □

Remark 10. If there exists a flat grouping for $D_1, \ldots, D_L$, then the off-diagonal points are distributed as several clusters over the half-plane Ω; see Figure 2. Though flat groupings are special, there are counterexamples if we drop any one of the three conditions; see Figure 3.

4 A Finite Sample Convergence Rate for Flat Groupings

With a guarantee of uniqueness of Fréchet means for sets of persistence diagrams given by Theorem 9 above, we are now in a position to derive a finite sample convergence rate for the empirical Fréchet mean for sets of persistence diagrams exhibiting flat groupings to the population Fréchet mean, which is the main topic in this section. Such a result paves the way to establishing the Fréchet mean as a viable tool with theoretical guarantees in important practical settings, such as those discussed by Cao and Monod (2022) on finding an appropriate representation to approximate the true persistent homology of a very large, yet finite, dataset.

For $\rho = \frac{1}{L}\sum_{i=1}^{L} \delta_{D_i}$ a discrete probability measure supported on $\{D_1, \ldots, D_L\}$ and $D'_1, \ldots, D'_B$ i.i.d. samples drawn from ρ, Turner et al. (2014) proved that if the Fréchet mean for ρ is unique, then with probability

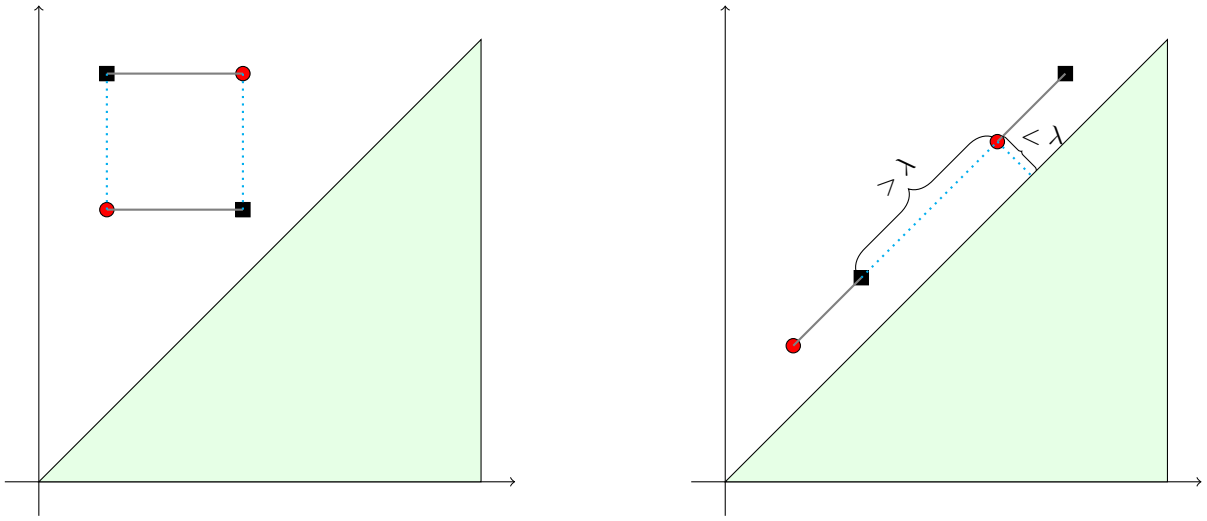

Figure 3: Counterexamples violating the conditions of flat groupings. On the left panel, a grouping for two persistence diagrams $D_{\text{red}}, D_{\text{black}}$ is depicted by solid lines. Four off-diagonal points form the corners of a square, hence the grouping violates conditions 1 and 2. The Fréchet mean is not unique as the grouping depicted by dotted lines gives another Fréchet mean. On the right panel, the grouping depicted by the solid lines satisfies conditions 1 and 2, but violates condition 3 as all off-diagonal points are near the diagonal. The mean of the grouping is not a Fréchet mean. The optimal grouping here will match all off-diagonal points with the diagonal.

one the empirical Fréchet mean converges to the population Fréchet mean under the Hausdorff distance. This is the only existing convergence result for Fréchet means of sets of persistence diagrams; there is no finite sample convergence rate for Fréchet means of persistence diagrams. Recently, Le Gouic et al. (2019) established a general theory on the convergence rate of empirical Fréchet means in Alexandrov spaces with curvature bounded from below. Although it is tempting to apply these results directly to the space $(\mathcal{D}_2, \mathrm{W}_2)$ which is also an Alexandrov space with curvature bounded from below, the general theory is unfortunately not applicable for technical reasons unique to the space of persistence diagrams that will be elaborated upon, following more in-depth discussions in this section. We borrow the main idea, but need to reconstruct new results for our specific setting. We begin by outlining some properties of Fréchet means presented by Le Gouic et al. (2019) and then present our convergence result.

4.1 Metric Properties of Fréchet Means of Persistence Diagrams

Let $(\mathcal{S}, d)$ be a geodesic space. Given two tangent vectors $[u, s], [v, t] \in T_z\mathcal{S}$, $[u, s]$ is said to be opposite to $[v, t]$ if $s = t = 0$ or $s = t \neq 0$ and $\angle_z(u, v) = \pi$. Define

$$H_z\mathcal{S} = \{[u, s] \in T_z\mathcal{S} \mid \exists\ [v, t] \in T_z\mathcal{S} \text{ opposite to } [u, s]\}.$$

Let $o_z = [v, 0]$ be the tip of the tangent cone. Note that $o_z \in H_z\mathcal{S}$, thus $H_z\mathcal{S}$ is nonempty. Alexander et al. (2022) show that $H_z\mathcal{S}$ with the inherited cone metric is in fact a Hilbert space when $\mathcal{S}$ is an Alexandrov space with nonnegative curvature. $H_z\mathcal{S}$ is referred to as the Hilbert subcone of the tangent cone at z.

Let $\log_z$ be the log map at z. Suppose $\log_z(x) = [v_z^x, d(z, x)]$ and $\log_z(y) = [v_z^y, d(z, y)]$. Denote $\langle \log_z(x), \log_z(y) \rangle_z := d(z, x)d(z, y) \cos \angle(v_z^x, v_z^y)$. Let μ be a probability measure on $(\mathcal{S}, d)$ with finite second moment and $z_\star$ be a Fréchet mean of μ. The tangent cone at $z_\star$ then exhibits the following properties.

Theorem 11. *(Le Gouic et al., 2019, Theorem 7) Let $(\mathcal{S}, d)$ be an Alexandrov space with nonnegative curvature. Then*

1. *At $z_\star$, the following equality holds*

$$\iint \langle \log_{z_\star}(x), \log_{z_\star}(y) \rangle_{z_\star} \, \mathrm{d}\mu(x)\mathrm{d}\mu(y) = 0; \tag{10}$$

2. *The Hilbert subcone at* $z_\star$ *satisfies* $\log_\star(\mathrm{supp}(\mu)) \subseteq H_\star\mathcal{S}$;

3. *For any probability measure* ν *with finite second moment and* $\log_{z_\star}(\mathrm{supp}(\nu)) \subseteq H_{z_\star}\mathcal{S}$, *and any* $y \in \mathcal{S}$,

$$\int_{\mathcal{S}} \langle \log_{z_\star}(x), \log_{z_\star}(y)\rangle_{z_\star} \,\mathrm{d}\nu(x) = \left\langle \int_{H_{z_\star}\mathcal{S}} u \,\mathrm{d}\nu_\#(u), \log_{z_\star}(y) \right\rangle_{z_\star}, \tag{11}$$

where $\nu_\# = (\log_{z_\star})_\#(\nu)$ *is the pushforward measure on* $H_{z_\star}\mathcal{S}$.

For the first property, at any point $z \in \mathcal{S}$ the following inequality holds

$$\iint \langle \log_z(x), \log_z(y)\rangle_z \,\mathrm{d}\mu(x)\mathrm{d}\mu(y) \geq 0$$

as a consequence of the Lang–Schroeder inequality (Le Gouic, 2020; Lang and Schroeder, 1997). Furthermore, if $z = z_\star$ is a Fréchet mean, then we have

$$\int \langle \log_{z_\star}(x), \log_{z_\star}(y)\rangle_{z_\star} \,\mathrm{d}\mu(x) \leq 0$$

for all $y \in \mathcal{S}$, which yields (10).

For the second property, note that although the Hilbert subcone is defined at any point in $\mathcal{S}$, it can be trivial as there may not exist a pair of tangent vectors with opposite directions, as in the space of persistence diagrams, which we formalize below.

Proposition 12. *The Hilbert subcone at the empty persistence diagram* $D_\emptyset$ *is trivial with a single point, i.e.,* $H_{D_\emptyset}(\mathcal{D}_2) = \{o_{D_\emptyset}\}$.

Proof. For any two nonempty persistence diagrams D_1, D_2, note that assigning all points to the diagonal gives a trivial bijection between D_1 and D_2. We have

$$\mathrm{W}_2^2(D_1, D_\emptyset) + \mathrm{W}_2^2(D_2, D_\emptyset) \geq \mathrm{W}_2^2(D_1, D_2),$$

meaning that the angle between any two directions at $D_\emptyset$ is bounded by $\frac{\pi}{2}$. Thus, the Hilbert subcone only consists of the tip $o_{D_\emptyset}$. □

Thus the second property of Theorem 11 is crucial to guarantee that the Hilbert subcone at the Fréchet mean is not trivial given that the probability measure is not a Dirac measure at a single point.

Definition 13. Fix $z, y \in \mathcal{S}$, the *hugging function* at z with respect to y is defined as

$$\kappa_z^y(x) = 1 - \frac{\mathrm{C}_z^2(\log_z(x), \log_z(y)) - d^2(x,y)}{d^2(y,z)}.$$

Intuitively, the hugging function at z measures the proximity of $\mathcal{S}$ to the tangent cone $T_z\mathcal{S}$. More importantly, at the Fréchet mean, we have the following equality.

Theorem 14. *(Le Gouic et al., 2019, Theorem 8) Let* $(\mathcal{S}, d)$ *be an Alexandrov space with nonnegative curvature and* $z_\star$ *be a Fréchet mean for the probability measure* μ. *Then*

$$d(y, z_\star) \int \kappa_{z_\star}^y(x) \,\mathrm{d}\mu(x) = \int (d^2(x,y) - d^2(x, z_\star)) \,\mathrm{d}\mu(x) \tag{12}$$

for all $y \in \mathcal{S}$.

Limitations. The work of Le Gouic et al. (2019) assumes that the hugging function at the barycenter has a positive lower bound for all points in the entirety of the space. This assumption is closely related to the bi-extendibility of geodesics, meaning that a geodesic can be extended for a positive amount of time at both the start and end points. However, in the space of persistence diagrams, no geodesic can extend beyond the diagonal, which prohibits the direct application of these existing results.

4.2 Convergence of Empirical Fréchet Means of Flat Groupings

Let $\rho = \frac{1}{L}\sum_{i=1}^{L} D_i$ be a discrete probability measure on $\mathcal{D}_2$, and $D'_1, \ldots, D'_B$ be i.i.d. samples from ρ. Assume $\mathbf{G}$ is an flat grouping for $D_1, \ldots, D_L$. By Theorem 9, $D_\star = \mathrm{mean}(\mathbf{G})$ is the unique population Fréchet mean for ρ. Let G' be the induced grouping for $D'_1, \ldots, D'_B$, i.e., each column $(G')^j$ is the corresponding column of D'_j in $\mathbf{G}$. Since the induced groupings are also flat groupings, $\bar{D} = \mathrm{mean}(G')$ is thus the unique empirical Fréchet mean for $D'_1, \ldots, D'_B$.

We begin with a computation of the hugging function analogous to Definition 13 for the space of persistence diagrams.

Lemma 15. *For any $D_j \in \{D_1, \ldots, D_L\}$, we have*

$$\kappa_{D_\star}^{\bar{D}}(D_j) = \kappa_{\bar{D}}^{D_\star}(D_j) = 1 \tag{13}$$

Proof. For any $\lambda_1, \ldots, \lambda_L \geq 0$ with $\sum_{j=1}^{L} \lambda_j = 1$, consider the persistence diagram D_Λ such that the off-diagonal points are given by $\mathbf{G}_i^\Lambda = \sum_{i=j}^{L} \lambda_j \mathbf{G}_i^j$ for every selection $\mathbf{G}_i$. Let $\gamma_\Lambda(t)$ be the geodesic from $D_\star$ to D_Λ. For any $0 \leq t, s \leq 1$, the optimal matching between $\gamma_\Lambda(t)$ and $\gamma_{\Lambda'}(s)$ is given by $t\mathbf{G}_i^\Lambda + (1-t)\bar{\mathbf{G}}_i \mapsto s\mathbf{G}_i^{\Lambda'} + (1-s)\bar{\mathbf{G}}_i$. Therefore,

$$\begin{aligned}
&\cos\angle_{D_\star}(\log_{D_\star}(D_\Lambda), \log_{D_\star}(D_{\Lambda'})) \\
=&\lim_{t,s\to 0} \frac{\mathrm{W}_2^2(D_\star, \gamma_\Lambda(t)) + \mathrm{W}_2^2(D_\star, \gamma_{\Lambda'}(s)) - \mathrm{W}_2^2(\gamma_\Lambda(t), \gamma_{\Lambda'}(s))}{2\mathrm{W}_2(D_\star, \gamma_\Lambda(t))\mathrm{W}_2(D_\star, \gamma_{\Lambda'}(s))} \\
=&\lim_{t,s\to 0} \frac{\sum_{i=1} t^2\|\mathbf{G}_i^\Lambda - \bar{\mathbf{G}}_i\|^2 + \sum_{i=1} s^2\|\mathbf{G}_i^{\Lambda'} - \bar{\mathbf{G}}_i\|^2 - \sum_{i=1}\|t\mathbf{G}_i^\Lambda - s\mathbf{G}_i^{\Lambda'} - (t-s)\bar{\mathbf{G}}_i\|^2}{2ts\sqrt{\sum_{i=1}\|\mathbf{G}_i^\Lambda - \bar{\mathbf{G}}_i\|^2}\sqrt{\sum_{i=1}\|\mathbf{G}_i^{\Lambda'} - \bar{\mathbf{G}}_i\|^2}} \\
=&\frac{\sum_{i=1}\langle \mathbf{G}_i^\Lambda - \bar{\mathbf{G}}_i, \mathbf{G}_i^{\Lambda'} - \bar{\mathbf{G}}_i\rangle}{\sqrt{\sum_{i=1}\|\mathbf{G}_i^\Lambda - \bar{\mathbf{G}}_i\|^2}\sqrt{\sum_{i=1}\|\mathbf{G}_i^{\Lambda'} - \bar{\mathbf{G}}_i\|^2}} \\
=&\frac{\mathrm{W}_2^2(D_\star, D_\Lambda) + \mathrm{W}_2^2(D_\star, D_{\Lambda'}) - \mathrm{W}_2^2(D_\Lambda, D_{\Lambda'})}{2\mathrm{W}_2(D_\star, D_\Lambda)\mathrm{W}_2(D_\star, D_{\Lambda'})}.
\end{aligned}$$

By definition of the cone metric (cf. (1)), we have

$$\begin{aligned}
&\mathrm{C}^2_{D_\star}(\log_{D_\star}(D_\Lambda), \log_{D_\star}(D_{\Lambda'})) \\
=&\mathrm{W}_2^2(D_\star, D_\Lambda) + \mathrm{W}_2^2(D_\star, D_{\Lambda'}) - 2\mathrm{W}_2(D_\star, D_\Lambda)\mathrm{W}_2^2(D_\star, D_{\Lambda'})\cos\angle_{D_\star}(\log_{D_\star}(D_\Lambda), \log_{D_\star}(D_{\Lambda'})) \\
=&\mathrm{W}_2^2(D_\Lambda, D_{\Lambda'}),
\end{aligned}$$

which implies that $\kappa_{D_\star}^{D_\Lambda}(D_{\Lambda'}) = 1$. Specifically, $\kappa_{D_\star}^{\bar{D}}(D_j) = 1$ for all $j = 1, \ldots, L$.

For the hugging function at $\bar{D}$, a similar computation gives

$$\cos\angle_{\bar{D}}(\log_{\bar{D}}(D_\Lambda), \log_{\bar{D}}(D_{\Lambda'})) = \frac{\mathrm{W}_2^2(\bar{D}, D_\Lambda) + \mathrm{W}_2^2(\bar{D}, D_{\Lambda'}) - \mathrm{W}_2^2(D_\Lambda, D_{\Lambda'})}{2\mathrm{W}_2(\bar{D}, D_\Lambda)\mathrm{W}_2(\bar{D}, D_{\Lambda'})}.$$

Thus, the cone metric satisfies $\mathrm{C}^2_{\bar{D}}(\log_{\bar{D}}(D_\Lambda), \log_{\bar{D}}(D_{\Lambda'})) = \mathrm{W}_2^2(D_\Lambda, D_{\Lambda'})$, which implies $\kappa_{\bar{D}}^{D_\star}(D_j) = 1$ for all $j = 1, \ldots, L$. ∎

We now prove the following finite sample convergence rate for flat groupings.

Theorem 16. *Let $\rho = \frac{1}{L}\sum_{i=1}^{L} D_i$ be a discrete probability measure on $\mathcal{D}_2$, and $D'_1, \ldots, D'_B$ be i.i.d. samples from ρ. Assume $\mathbf{G}$ is an flat grouping for $D_1, \ldots, D_L$. Let $D_\star$ be the population Fréchet mean and $\bar{D}$ be the empirical Fréchet mean. Then*

$$\mathbb{E}[\mathrm{W}_2^2(\widehat{D}, D_\star)] \leq \frac{\sigma^2}{B}, \tag{14}$$

where $\sigma^2 = \mathbb{V}(\mathbf{G})$ is the variance of the grouping.

Proof. By Theorem 14 and Lemma 15, we have

$$\mathrm{W}_2^2(D_\star, \bar{D}) = \int (\mathrm{W}_2^2(D_\star, D) - \mathrm{W}_2^2(\bar{D}, D))\,\mathrm{d}\mu(D).$$

By the equality of (11) in Theorem 11, we have

$$\begin{aligned}
2\mathrm{W}_2^2(D_\star, \bar{D}) &= \int (\mathrm{W}_2^2(D_\star, D) + \mathrm{W}_2^2(D_\star, \bar{D}) - \mathrm{W}_2^2(\bar{D}, D))\,\mathrm{d}\mu(D) \\
&= 2\int \mathrm{W}_2(D_\star, D)\mathrm{W}_2(D_\star, \bar{D})\cos\angle_\star(\log_\star D, \log_\star \bar{D})\,\mathrm{d}\mu(D) \\
&= 2\int \langle \log_\star D, \log_\star \bar{D}\rangle_\star\,\mathrm{d}\mu(D) \\
&= 2\langle \overline{\log_\star D}, \log_\star \bar{D}\rangle_\star,
\end{aligned}$$

where $\overline{\log_\star D}$ denotes the mean of the pushforward empirical measure $(\log_\star)_\# \mu$. Note that $\log_\star(\mathrm{supp}(\mu)) \subseteq \log_\star(\mathrm{supp}(\boldsymbol{\mu})) \subseteq H_\star \mathcal{D}_2$. In the Hilbert subcone, we have $\mathrm{W}_2^2(D_\star, \bar{D}) \leq \mathrm{C}_\star(\overline{\log_\star D}, o_\star)\mathrm{C}_\star(\log_\star \bar{D}, o_\star)$. Since $\mathrm{C}_\star(\log_\star \bar{D}, o_\star) = \mathrm{W}_2(D_\star, \bar{D})$, then $\mathbb{E}[\mathrm{W}_2^2(D_\star, \bar{D})] \leq \mathbb{E}[\mathrm{C}_\star^2(\overline{\log_\star D}, o_\star)]$.

In Hilbert spaces, we know that the empirical mean $\overline{\log_\star D}$ converges to the population mean $\mathbb{E}[(\log_\star)_\# \boldsymbol{\mu}] = o_\star$ in the following sense

$$\mathbb{E}[\mathrm{C}_\star^2(\overline{\log_\star D}, o_\star)] = \frac{\sigma^2}{B}$$

where $\sigma^2 = \int \mathrm{C}_\star^2(\log_\star D, o_\star)]\,\mathrm{d}\mu(D) = \int \mathrm{W}_2^2(D, D_\star)\,\mathrm{d}\boldsymbol{\mu}(D)$, thus completing the proof. □

5 Discussion

In this paper, we introduced the notion of flat groupings for sets of persistence diagrams, which possess desirable geometric properties that have direct implications on statistical properties in the space of persistence diagrams. Flat groupings allow us to fill an important gap in the theory of statistical persistent homology over nearly the past decade. However, in practice, real data often generates persistence diagrams exhibiting persistence points near the diagonal which can make it difficult to construct flat groupings in practice. Given a set of persistence diagrams, a future direction of research is to determine an appropriate cropping of off-diagonal points so that a flat grouping of the persistence points may be constructed, and as a consequence, to approximate the Fréchet mean. Existing work by Fasy et al. (2014) proposes a statistical approach to crop off-diagonal points based on the construction of confidence regions around the diagonal; the driving assumption here is that topological features near the diagonal are considered "noise." However, more recent work by Reani and Bobrowski (2021) demonstrates that persistence diagrams with many persistence points near the diagonal may in fact correspond to datasets (point clouds) with very clear topological signal. The question of finding an appropriate cropping that preserves "true" signal and concurrently allows the construction of flat groupings is thus a different question than existing methods of topological signal processing or cropping.

An alternative measure of centrality of data is the median, which may also be defined for persistence diagrams and has a similar characterization as the Fréchet mean (Turner, 2020). However, understanding this measure entails an entirely different study, since the median is the minimum of the Fréchet function (2) with respect to the 1-Wasserstein distance, which would require studying the space $(\mathcal{D}_1, \mathrm{W}_1)$. Less is known about the geometry of $(\mathcal{D}_1, \mathrm{W}_1)$, since it is not any Alexandrov space of curvature bounded from below or above (Turner, 2013), thus none of the prior results established by Le Gouic et al. (2019) and used in this work are applicable; new tools and strategies would need to be constructed.

Acknowledgments

The authors wish to thank Katharine Turner for helpful discussions. Y.C. is funded by a President's PhD Scholarship at Imperial College London.

References

Alexander, S., V. Kapovitch, and A. Petrunin (2022). Alexandrov geometry: Foundations. *arXiv preprint arXiv:1903.08539*.

Burago, D., Y. Burago, and S. Ivanov (2001). *A course in metric geometry*, Volume 33. American Mathematical Society.

Cao, Y. and A. Monod (2022). Approximating Persistent Homology for Large Datasets. *arXiv preprint arXiv:2204.09155*.

Chazal, F., V. De Silva, M. Glisse, and S. Oudot (2016). *The structure and stability of persistence modules*, Volume 10. Springer.

Edelsbrunner, H. and J. Harer (2010). *Computational Topology: An Introduction*. American Mathematical Soc.

Edelsbrunner, H., J. Harer, et al. (2008). Persistent homology—A survey. *Contemporary mathematics 453*, 257–282.

Fasy, B. T., F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, and A. Singh (2014). Confidence sets for persistence diagrams. *The Annals of Statistics*, 2301–2339.

Lacombe, T., M. Cuturi, and S. Oudot (2018). Large scale computation of means and clusters for persistence diagrams using optimal transport. *Advances in Neural Information Processing Systems 31*.

Lang, U. and V. Schroeder (1997). Kirszbraun's theorem and metric spaces of bounded curvature. *Geometric & Functional Analysis GAFA 7*(3), 535–560.

Le Gouic, T. (2020). A note on flatness of non separable tangent cone at a barycenter. *Comptes Rendus. Mathématique 358*(4), 489–495.

Le Gouic, T., Q. Paris, P. Rigollet, and A. J. Stromme (2019). Fast convergence of empirical barycenters in Alexandrov spaces and the Wasserstein space. *arXiv preprint arXiv:1908.00828*.

Mileyko, Y., S. Mukherjee, and J. Harer (2011). Probability measures on the space of persistence diagrams. *Inverse Problems 27*(12), 124007.

Munch, E., K. Turner, P. Bendich, S. Mukherjee, J. Mattingly, and J. Harer (2015). Probabilistic Fréchet means for time varying persistence diagrams. *Electronic Journal of Statistics 9*(1), 1173–1204.

Reani, Y. and O. Bobrowski (2021). Cycle registration in persistent homology with applications in topological bootstrap. *arXiv preprint arXiv:2101.00698*.

Turner, K. (2013). Means and medians of sets of persistence diagrams. *arXiv preprint arXiv:1307.8300*.

Turner, K. (2020). Medians of populations of persistence diagrams. *Homology, Homotopy & Applications 22*(1).

Turner, K., Y. Mileyko, S. Mukherjee, and J. Harer (2014). Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry 52*(1), 44–70.