On the Metric Distortion of Embedding Persistence Diagrams into Separable Hilbert Spaces

Mathieu Carrière
Department of Systems Biology, Columbia University, New York, USA
mc4660@cumc.columbia.edu

Ulrich Bauer
Department of Mathematics, Technical University of Munich (TUM), Germany
ulrich.bauer@tum.de

──── **Abstract** ────

Persistence diagrams are important descriptors in Topological Data Analysis. Due to the nonlinearity of the space of persistence diagrams equipped with their *diagram distances*, most of the recent attempts at using persistence diagrams in machine learning have been done through kernel methods, i.e., embeddings of persistence diagrams into Reproducing Kernel Hilbert Spaces, in which all computations can be performed easily. Since persistence diagrams enjoy theoretical stability guarantees for the diagram distances, the *metric properties* of the feature map, i.e., the relationship between the Hilbert distance and the diagram distances, are of central interest for understanding if the persistence diagram guarantees carry over to the embedding. In this article, we study the possibility of embedding persistence diagrams into separable Hilbert spaces with bi-Lipschitz maps. In particular, we show that for several stable embeddings into infinite-dimensional Hilbert spaces defined in the literature, any lower bound must depend on the cardinalities of the persistence diagrams, and that when the Hilbert space is finite dimensional, finding a bi-Lipschitz embedding is impossible, even when restricting the persistence diagrams to have bounded cardinalities.

1 Introduction

The increase of available data in both academia and industry has been exponential over the past few decades, making data analysis and machine learning ubiquitous in many different fields of science. Topological Data Analysis (TDA) [5] is one specific field of data science, which focuses more on *complex* rather than big data. The general assumption of TDA is that data is actually sampled from geometric or low-dimensional domains, whose topological features are relevant to the analysis. These topological features are usually encoded in a mathematical object called *persistence diagram*, which is roughly a set of points in the plane, each point representing a topological feature whose size is contained in the coordinates of the point. Persistence diagrams have been proved to bring complementary information to other traditional descriptors in many different applications, often leading to large result improvements. This is also due to the *stability* properties of the persistence diagrams, which state that persistence diagrams computed on similar data are also very close in the diagram distances [2, 8, 9].

Unfortunately, the use of persistence diagrams in machine learning methods is not straightforward, since many algorithms expect data to be Euclidean vectors, while persistence diagrams are sets of points with possibly different cardinalities. Moreover, the *diagram distances* used to compare persistence diagrams are computed by means of optimal matchings,
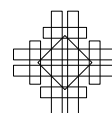
and thus are quite different from Euclidean metrics. The usual way to cope with such difficult data is to use *kernel methods*. A kernel is a symmetric function on the data whose evaluation on a pair of data points equals the scalar product of the images of these points under a *feature map* into a Hilbert space, called the *Reproducing Kernel Hilbert Space* of the kernel. Many algorithms can be *kernelized*, such as PCA and SVM, allowing one to handle non-Euclidean data as soon as either a kernel or a feature map is available.

Hence, the question of defining a feature map into a Hilbert space has been intensively studied in the past few years, and, as of today, various methods have been proposed and implemented, either into finite or infinite dimensional Hilbert spaces [4, 7, 21, 16, 1, 6, 13]. Since persistence diagrams enjoy stability properties, it is also natural to ask the same guarantee for their embeddings. Indeed, various feature maps defined in the literature satisfy a stability property stating that the Hilbert distance between the image of the persistence diagrams is upper bounded by some specific diagram distance, most commonly the 1-Wasserstein diagram distance. In many cases, this upper bound applies only to a restricted set of persistence diagrams with bounded number and bounded range of persistence pairs, and these bounds enter the constant in the stability estimate. However, some unconditional stability results exist as well, e.g., for the Persistence Scale Space feature map [21].

A more difficult question is to prove whether a lower bound also holds or not. As a first step in this direction, a lower bound for the *Sliced Wasserstein distance* was proved in [6], showing that this metric is equivalent to the first diagram distance. Moreover, since the Sliced Wasserstein distance is conditionally negative definite, a Gaussian kernel can be defined with it with Berg's theorem [3]. However, even in this case, the resulting Sliced Wasserstein kernel distance is not equivalent to the Sliced Wasserstein distance, and so the corresponding feature map is not guaranteed to be bi-Lipschitz. Thus, the question remained open in general.

Contributions

In this article, we consider the general question of the existence of bi-Lipschitz embeddings of persistence diagrams into separable Hilbert spaces. More precisely, we show the following results:

- For several stable feature maps defined in the literature, if such a bi-Lipschitz embedding exists for persistence diagrams with bounded number and range of points, then the ratio between upper and lower bound goes to ∞ as the bounds on the number of points in the persistence diagrams and on their range increase to ∞ (Theorem 3.5 and Proposition 3.9).
- Such a bi-Lipschitz embedding does not exist if the Hilbert space is finite dimensional (Theorem 4.4),

Finally, we also provide experimental evidence of this behavior by computing the metric distortions of various feature maps for persistence diagrams with increasing cardinalities.

Related work

Feature maps for persistence diagrams can be classified into two different classes, depending on whether the corresponding Hilbert space has finite or infinite dimension.

In the infinite dimensional case, the first attempt was that proposed in [4], in which persistence diagrams are turned into families of L^2 functions, called *landscapes*, by computing the homological rank functions given by the persistence diagram points. Another common way to define a feature map is to see the points of the persistence diagrams as centers of Gaussians with a fixed bandwidth, weighted by the distance of the point to the diagonal.

This is the approach originally advocated in [21], and later generalized in [17], leading to the so-called *Persistence Scale Space* and *Persistence Weighted Gaussian* feature maps. Another possibility is to define a Gaussian-like feature map by using the *Sliced Wasserstein distance* between persistence diagrams, which is conditionally negative definite. This implicit feature map, called the *Sliced Wasserstein* map, was defined in [6].

In the finite dimensional case, many different possibilities are available. One may consider evaluating a family of tropical polynomials onto the persistence diagram [14], taking the sorted vector of the pairwise distances between the persistence diagram points [7], or computing the coefficients of a complex polynomial whose roots are given by the persistence diagram points [10]. Another line of work has been proposed in [1] by discretizing the Persistence Scale Space feature map. The idea is to discretize the plane into a fixed grid, and then compute a value for each pixel by integrating Gaussian functions centered on the persistence diagram points. Finally, persistence diagrams have been incorporated in deep learning frameworks in [13], in which Gaussian functions (whose means and variances are optimized by the neural network during training) are integrated against persistence diagrams seen as discrete measures.

2　Background

2.1　Persistence Diagrams

Persistent homology is a technique of TDA, using concepts from algebraic topology, which allows the user to compute and encode topological information of datasets in a compact descriptor called the *persistence diagram*. Given a space X and a continuous and real-valued function $f : X \to \mathbb{R}$, the persistence diagram of f can be computed under mild conditions (the function has to be *tame*, see [8] for more details), and consists in a finite set of points with multiplicities in the upper-diagonal half-plane $\mathrm{Dg}(f) = \{(x_i, y_i)\} \subset \{(x, y) \in \mathbb{R}^2 : y > x\}$. This set of points is computed from the family of *sublevel sets* of f, that is the sets of the form $f^{-1}((-\infty, \alpha])$, for some $\alpha \in \mathbb{R}$. More precisely, persistence diagrams encode the different *topological events* that occur as α increases from $-\infty$ to ∞. Such topological events include creation and merging of connected components and cycles in every dimension. For example, when dealing with a point cloud $\hat{X} \subset \mathbb{R}^n$, a common strategy to obtain a persistence diagram from $\hat{X}$ is to set $X = \mathbb{R}^n$ and to use the Euclidean distance to $\hat{X}$ as the function f. See also Figure 1 for another example.

Intuitively, persistent homology records, for each topological feature that appears in the family of sublevel sets, the value α_b at which the feature appears, called the *birth value*, and the value α_d at which it gets merged or filled in, called the *death value*. These values are then used as coordinates for a corresponding point in the persistence diagram. In general, depending on the space X, there will also be topological features for which there is no finite death value. Such features are however not considered in the context of the present paper. Note that several features may have the same birth and death values, so points in the persistence diagram have multiplicities. Moreover, since $\alpha_d \geq \alpha_b$, these points are always located above the diagonal $\Delta = \{(x, x) : x \in \mathbb{R}\}$. A general intuition about persistence diagrams is that the distance of a point to Δ is a direct measure of its relevance: if a point is close to Δ, it means that the corresponding cycle got filled in right after its appearance, thus suggesting that it is likely due to noise in the dataset. On the contrary, points that are far away from Δ represent cycles with a significant life span, and are more likely to be relevant for the analysis. We refer the interested reader to [11, 19] for more details about persistent homology.
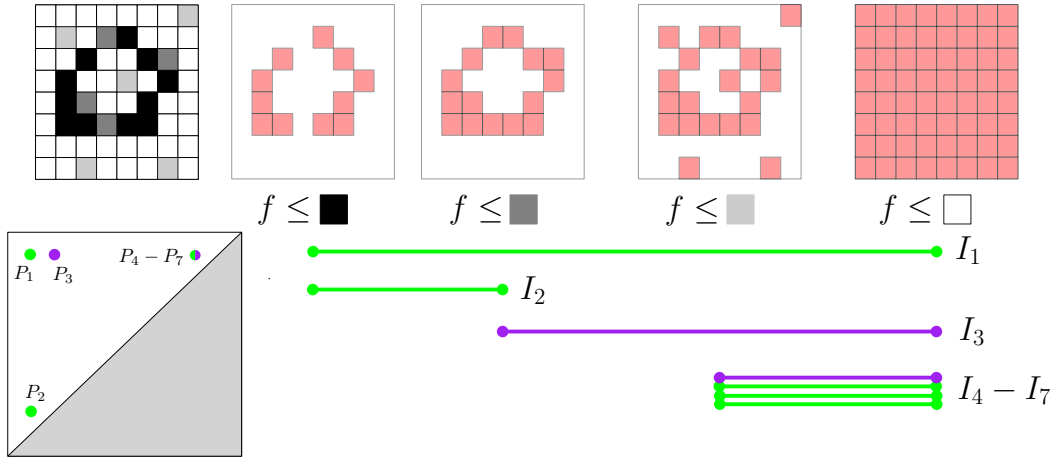
Figure 1 Example of persistence diagram computation. The space we consider is a blurry image of a zero, and the function f that we use is the grey level value on each pixel. We show four different sublevel sets of f. For each sublevel set, the corresponding pixels are displayed in pink color. In the first sublevel set, two connected components are present in the sublevel set, so we start two intervals I_1 and I_2. In the second one, one connected component got merged to the other, so we stop the corresponding interval I_2, and a cycle (loop) is created, so we start a third interval I_3. In the third sublevel set, a new small cycle is created, as well as three more connected components. In the fourth sublevel set, all pixels belong to the set: all cycles are filled in and all connected components are merged together, so all intervals end here. Finally, each interval I_k is represented as a point P_k in the plane (using the endpoints as coordinates).

Notation

Let $\mathcal{D}$ be the set of persistence diagrams with at most a countable number of points. More formally, $\mathcal{D}$ can be equivalently defined as a set of multiplicity functions $\{m : \mathbb{R}^2 \setminus \Delta \to \mathbb{N} : \text{supp}(m) \text{ is countable}\}$, where each point $q \in \text{supp}(m)$ is a point in the corresponding persistence diagram with multiplicity $m(q)$. Let $\mathcal{D}_N$ be the set of persistence diagrams with less than N points, i.e., $\mathcal{D}_N = \{m : \mathbb{R}^2 \setminus \Delta \to \mathbb{N} : \sum_q m(q) < N\}$. Let $\mathcal{D}^L$ be the space of persistence diagrams included in $[-L, L]^2$, i.e., $\mathcal{D}^L = \{m : \mathbb{R}^2 \setminus \Delta \to \mathbb{N} : \text{supp}(m) \subset [-L, L]^2\}$. Finally, let $\mathcal{D}_N^L$ be the space of persistence diagrams with less than N points included in $[-L, L]^2$, i.e., $\mathcal{D}_N^L = \mathcal{D}_N \cap \mathcal{D}^L$. Obviously, we have the following sequences of (strict) inclusions: $\mathcal{D}_N^L \subset \mathcal{D}_N \subset \mathcal{D}$, and $\mathcal{D}_N^L \subset \mathcal{D}^L \subset \mathcal{D}$.

Diagram distances

Persistence diagrams can be efficiently compared using the *diagram distances*, which is a family of distances parametrized by an integer p that rely on the computation of *partial matchings*. Recall that two persistence diagrams Dg_1 and Dg_2 may have different number of points. A *partial matching* Γ between Dg_1 and Dg_2 is a subset of $\text{Dg}_1 \times \text{Dg}_2$. It comes along with Γ_1 (resp. Γ_2), which is the set of points of Dg_1 (resp. Dg_2) that are not matched to a point of Dg_2 (resp. Dg_1) by Γ. The p-cost of Γ is given as:

$$c_p(\Gamma) = \sum_{(p,q)\in\Gamma} \|p - q\|_\infty^p + \sum_{p\in\Gamma_1} \|p - \Delta\|_\infty^p + \sum_{q\in\Gamma_2} \|q - \Delta\|_\infty^p.$$

The p-diagram distance is then defined as the cost of the best partial matching:

▶ **Definition 2.1.** *Given two persistence diagrams* Dg_1 *and* Dg_2, *the p-diagram distance* d_p *is defined as:*

$$d_p(\mathrm{Dg}_1, \mathrm{Dg}_2) = \inf_\Gamma \sqrt[p]{c_p(\Gamma)}.$$

Note that in the literature, these distances are often called the *Wasserstein distances* between persistence diagrams. Here, we follow the denomination of [6]. In particular, taking a maximum instead of a sum in the definition of the cost,

$$c_\infty(\Gamma) = \max_{(p,q)\in\Gamma} \|p - q\|_\infty + \max_{p\in\Gamma_1} \|p - \Delta\|_\infty + \max_{q\in\Gamma_2} \|q - \Delta\|_\infty.$$

allows to add one more distance in the family, the *bottleneck distance* $d_\infty(\mathrm{Dg}_1, \mathrm{Dg}_2) = \inf_\Gamma c_\infty(\Gamma)$.

Stability

A useful property of persistence diagrams is their *stability* in terms of the data generating the diagrams. Indeed, it is well known in the literature that persistence diagrams computed from close functions are close themselves in the bottleneck distance:

▶ **Theorem 2.2** ([8, 9]). *Given two tame functions* $f, g : X \to \mathbb{R}$, *one has the following inequality:*

$$d_\infty(\mathrm{Dg}(f), \mathrm{Dg}(g)) \leq \|f - g\|_\infty. \tag{1}$$

In other words, the map Dg is 1-Lipschitz. Note that stability results exist as well for the other diagram distances, but these results are weaker than the above Lipschitz condition, and they require more conditions – see [19].

2.2 Bi-Lipschitz embeddings.

The main question that we adress in this article is the one of preserving the persistence diagram metric properties when using embeddings into Hilbert spaces. For instance, one may ask the images of persistence diagrams under a feature map into a Hilbert space to be stable as well. A natural question is then whether a lower bound also exists, i.e., whether the feature map Φ is a *bi-Lipschitz embedding* between $(\mathcal{D}, d_p)$ and $\mathcal{H}$.

▶ **Definition 2.3.** *Let* (X, d_X) *and* (Y, d_Y) *be two metric spaces. A* bi-Lipschitz embedding *between* (X, d_X) *and* (Y, d_Y) *is a map* $\Phi : X \to Y$ *such that there are constants* $0 < A, B < \infty$ *such that*

$$A\, d_X(x, x') \leq d_Y(\Phi(x), \Phi(x')) \leq B\, d_X(x, x')$$

for any $x, x' \in X$. *The metrics* d_X *and* d_Y *are called* strongly equivalent, *and the constants* A *and* B *are called the* lower *and* upper *metric distortion bounds, respectively. If* $A = B = 1$, Φ *is called an* isometric *embedding.*

Note that this definition is equivalent to the commonly used definition, which additionally requires $A = \frac{1}{B}$.

▶ Remark 2.4. Finding an isometric embedding of persistence diagrams into a Hilbert space is impossible since geodesics are unique in a Hilbert space, while this is not the case for persistence diagrams, as shown in the proof of Proposition 2.4 in [23].

▶ **Remark 2.5.** For feature maps that are *bounded*, i.e., those maps Φ such that there exists a constant $C > 0$ for which $\|\Phi(\mathrm{Dg})\| \leq C$ for all Dg, it is obviously impossible to find a bi-Lipschitz embedding. This involves for instance the Sliced Wasserstein (SW) feature map [6], which is defined implicitly from a Gaussian-like function.

3   Mapping into separable Hilbert spaces

In our first main result, we use *separability* to determine whether a bi-Lipschitz embedding can exist between the space of persistence diagrams and a Hilbert space.

▶ **Definition 3.1.** *A metric space is called* separable *if it has a dense countable subset.*

For instance, the following three Hilbert spaces (equipped with their canonical metrics) are separable: $\mathbb{R}^n$, ℓ_2 and $L_2(\Omega)$, where Ω is separable. The two following results describe well-known properties of separable spaces.

▶ **Proposition 3.2.** *Any subspace of a separable metric space is separable as well.*

▶ **Proposition 3.3.** *Let (X, d_X) and (Y, d_Y) be two metric spaces, and assume there is a bi-Lipschitz embedding $\Phi : X \to Y$, with Lipschitz constants A and B. Then X is separable if and only if $\mathrm{im}(\Phi)$ is separable.*

The following lemma shows that for a feature map Φ which is bi-Lipschitz when restricted to $\mathcal{D}_N^L$, the limits of the corresponding constants can actually be used to study the general metric distortion in $\mathcal{D}$.

▶ **Lemma 3.4.** *Let $p \in \mathbb{R}_+^*$ and let d be a continuous metric on $(\mathcal{D}, d_p)$. Let*

$$R_N^L = \left\{ \frac{d_p(\mathrm{Dg}, \mathrm{Dg}')}{d(\mathrm{Dg}, \mathrm{Dg}')} : \mathrm{Dg} \neq \mathrm{Dg}' \in \mathcal{D}_N^L \right\},$$
$$A_N^L = \inf\ R_N^L \quad and \quad B_N^L = \sup\ R_N^L.$$

Since A_N^L is nonincreasing and B_N^L is nondecreasing with respect to N and L, we define

$$A_N = \lim_{L \to \infty} A_N^L, \qquad\qquad A^L = \lim_{N \to \infty} A_N^L, \qquad\qquad A = \liminf_{N,L \to \infty} A_N^L,$$
$$B_N = \lim_{L \to \infty} B_N^L, \qquad\qquad B^L = \lim_{N \to \infty} B_N^L, \qquad\qquad B = \limsup_{N,L \to \infty} B_N^L,$$

where the limit superior and inferior for $N, L \to \infty$ are taken for the nets A_N^L and B_N^L over the directed set $\mathbb{N} \times \mathbb{R}$. Then the following inequalities hold:

$$A^L\, d(\mathrm{Dg}, \mathrm{Dg}') \leq d_p(\mathrm{Dg}, \mathrm{Dg}') \leq B^L\, d(\mathrm{Dg}, \mathrm{Dg}') \qquad\qquad \textit{for all } \mathrm{Dg}, \mathrm{Dg}' \in \mathcal{D}^L,$$
$$A_N\, d(\mathrm{Dg}, \mathrm{Dg}') \leq d_p(\mathrm{Dg}, \mathrm{Dg}') \leq B_N\, d(\mathrm{Dg}, \mathrm{Dg}') \qquad\qquad \textit{for all } \mathrm{Dg}, \mathrm{Dg}' \in \mathcal{D}_N,$$
$$A\, d(\mathrm{Dg}, \mathrm{Dg}') \leq d_p(\mathrm{Dg}, \mathrm{Dg}') \leq B\, d(\mathrm{Dg}, \mathrm{Dg}') \qquad\qquad \textit{for all } \mathrm{Dg}, \mathrm{Dg}' \in \mathcal{D}.$$

Note that A, A_N, A^L, B, B_N and B^L may be equal to 0 or ∞, so it does not necessarily hold that d and d_p are strongly equivalent on either $\mathcal{D}_N$, $\mathcal{D}^L$, or $\mathcal{D}$.

Proof. We only prove the last inequality, since the proof extends verbatim to the other two. Pick any two persistence diagrams $\mathrm{Dg}, \mathrm{Dg}' \in \mathcal{D}$. Let $\Gamma = \{(p_i, q_i)\}_{i \in \mathbb{N}}$ be an optimal partial matching achieving $d_p(\mathrm{Dg}, \mathrm{Dg}')$, where p_i (resp. q_i) is either in Dg (resp. Dg') or in $\pi_\Delta(\mathrm{Dg}')$

(resp. $\pi_\Delta(\mathrm{Dg})$), and where π_Δ is the projection (in the Euclidean norm) onto the diagonal Δ. We now define two sequences of persistence diagrams $\{\mathrm{Dg}_n\}_{n\in\mathbb{N}}$ and $\{\mathrm{Dg}'_n\}_{n\in\mathbb{N}}$ recursively with $\mathrm{Dg}_0 = \mathrm{Dg}'_0 = \emptyset$ and

$$\mathrm{Dg}_{n+1} = \begin{cases} \mathrm{Dg}_n \text{ if } p_{n+1} \in \pi_\Delta(\mathrm{Dg}'), \\ \mathrm{Dg}_n \cup \{p_{n+1}\} \text{ otherwise,} \end{cases} \qquad \mathrm{Dg}'_{n+1} = \begin{cases} \mathrm{Dg}'_n \text{ if } q_{n+1} \in \pi_\Delta(\mathrm{Dg}), \\ \mathrm{Dg}'_n \cup \{q_{n+1}\} \text{ otherwise.} \end{cases}$$

Note that Γ might have only a finite number of elements if both Dg and Dg' have finite cardinalities. In this case, we set $\{\mathrm{Dg}_n\}_{n\in\mathbb{N}}$ (resp. $\{\mathrm{Dg}'_n\}_{n\in\mathbb{N}}$) to be constant and equal to Dg (resp. Dg') for sufficiently large n. Moreover, define

$$l_n = \max\{\max\{\|p\|_\infty : p \in \mathrm{Dg}_n\}, \max\{\|q\|_\infty : q \in \mathrm{Dg}'_n\}\},$$
$$s_n = \max\{\mathrm{card}(\mathrm{Dg}_n), \mathrm{card}(\mathrm{Dg}'_n)\},$$

Note that both $\{l_n\}_{n\in\mathbb{N}}$ and $\{s_n\}_{n\in\mathbb{N}}$ are nondecreasing. We have $\mathrm{Dg}_n, \mathrm{Dg}'_n \in \mathcal{D}_{s_n}^{l_n}$ and thus

$$A_{s_n}^{l_n} \, d(\mathrm{Dg}_n, \mathrm{Dg}'_n) \leq d_p(\mathrm{Dg}_n, \mathrm{Dg}'_n) \leq B_{s_n}^{l_n} \, d(\mathrm{Dg}_n, \mathrm{Dg}'_n). \tag{2}$$

Now, since $d_p(\mathrm{Dg}_n, \mathrm{Dg}) \to 0$ when $n \to \infty$, we have $d(\mathrm{Dg}_n, \mathrm{Dg}) \to 0$ by continuity of d, and similarly $d(\mathrm{Dg}'_n, \mathrm{Dg}') \to 0$. Hence, we have $d_p(\mathrm{Dg}_n, \mathrm{Dg}'_n) \to d_p(\mathrm{Dg}, \mathrm{Dg}')$ and $d(\mathrm{Dg}_n, \mathrm{Dg}'_n) \to d(\mathrm{Dg}, \mathrm{Dg}')$ with the triangle inequality. We finally obtain the desired inequality by letting $n \to \infty$ in (2). ◀

A corollary of the previous results is that even if a feature map taking values in a separable Hilbert space might be bi-Lipschitz when restricted to $\mathcal{D}_N^L$, the ratio of upper and lower bound has to go to ∞ as soon as the domain of the feature map is not separable.

▶ **Theorem 3.5.** *Let $\Phi : \mathcal{D}_\Phi \to \mathcal{H}$ be a feature map defined on a non-separable subspace $\mathcal{D}_\Phi$ of persistence diagrams containing every $\mathcal{D}_N^L$, i.e., $\mathcal{D}_N^L \subset \mathcal{D}_\Phi$ for each N, L. Assume Φ takes values in a separable Hilbert space $\mathcal{H}$, and that Φ is bi-Lipschitz on each $\mathcal{D}_N^L$ with constants A_N^L, B_N^L. Then $B_N^L/A_N^L \to \infty$ as $N, L \to \infty$.*

Note that, by Theorem 12 in [18], if the p-total persistence, i.e., the p-diagram distance to the empty diagram, of each element of $\mathcal{D}_\Phi$ is finite, then $\mathcal{D}_\Phi$ becomes separable w.r.t. d_p. In particular, this means that $\mathcal{D}_N^L$ is separable for each N, L. Moreover, this shows that Theorem 3.5 applies only to domains $\mathcal{D}_\Phi$ containing at least one diagram whose p-total persistence is infinite, in particular, a diagram with an infinite number of points.

However, many feature maps defined in the literature, such as the Persistence Weighted Gaussian feature map [17] or the Landscape feature map [4], are actually defined on such domains, and take values in separable spaces, such as the function space $L^2(\Omega)$, where Ω is the upper half-plane $\{(x,y) : x \leq y\}$. Hence, to illustrate how Theorem 3.5 applies to these feature maps, we now provide two lemmata. In the first one, we define a set $\mathcal{S}$ which is not separable with respect to d_1, and in the second one, we show that, nevertheless, $\mathcal{S}$ is actually included in the domain $\mathcal{D}_\Phi$ of these feature maps.

▶ **Lemma 3.6.** *Consider the set of points $P = \left\{ \left(k, k + \frac{1}{k}\right) \in \mathbb{R}^2 : k \in \mathbb{N} \right\}$, and define the set $\mathcal{S}$ of persistence diagrams as the power set $\mathcal{S} = \mathcal{P}(P)$. Then $(\mathcal{S}, d_1)$ is not separable.*

Proof. Let $p_k = \left(k, k + \frac{1}{k}\right) \in \mathbb{R}^2$ for all $k \in \mathbb{N}$. Then we have $\mathcal{S} = \{\mathrm{Dg}_u\}_{u\in\mathcal{U}}$, where $\mathcal{U} = \{0,1\}^\mathbb{N}$ is the set of sequences with values in $\{0,1\}$, and where $\mathrm{Dg}_u = \{p_i : i \in \mathrm{supp}(u)\}$. First note that since the sequences $u \in \mathcal{U}$ can have infinite support, the spaces $\mathcal{U}$ and $\mathcal{S} = \{\mathrm{Dg}_u\}_{u\in\mathcal{U}}$ are not countable.

Let $\sim$ be the equivalence relation on $\mathcal{S}$ defined as

$$\mathrm{Dg}_u \sim \mathrm{Dg}_v \iff |\operatorname{supp}(u) \triangle \operatorname{supp}(v)| < \infty,$$

where $\triangle$ denotes the symmetric difference of sets. Since the set of sequences with finite support is countable, it follows that each equivalence class $[\mathrm{Dg}_u]_\sim$ is countable as well. In particular, this means that the set of equivalence classes $\mathcal{S}/\sim$ is uncountable, since otherwise $\mathcal{S}$ would be countable as a countable union of countable equivalence classes.

We now prove the result by contradiction. Assume that $\mathcal{S}$ is separable, and let $\mathcal{S}' \subset \mathcal{S}$ be the corresponding dense countable subset of $\mathcal{S}$. Let $\epsilon > 0$. Then for each $u \in \mathcal{U}$, there is at least one sequence $u' \in \mathcal{U}$ such that $\mathrm{Dg}_{u'} \in \mathcal{S}'$ and $d_1(\mathrm{Dg}_u, \mathrm{Dg}_{u'}) \leq \epsilon$. We now claim that every such u' satisfies $\mathrm{Dg}_{u'} \in [\mathrm{Dg}_u]_\sim$. Indeed, assume $\mathrm{Dg}_{u'} \notin [\mathrm{Dg}_u]_\sim$ and let $\mathcal{I} = \operatorname{supp}(u') \triangle \operatorname{supp}(u)$. Then, since $|\mathcal{I}| = \infty$, we would have

$$d_1(\mathrm{Dg}_u, \mathrm{Dg}_{u'}) = \sum_{k \in \mathcal{I}} \frac{1}{k} = \infty > \epsilon,$$

which is not possible. Hence, this means that $|\mathcal{S}'| \geq |\mathcal{S}/\sim|$. However, we showed that $\mathcal{S}/\sim$ is uncountable, meaning that $\mathcal{S}'$ is uncountable as well, which leads to a contradiction, since $\mathcal{S}'$ is countable by assumption.   ◀

We now show that the Persistence Weighted Gaussian and the Landscape feature maps are well-defined on the set $\mathcal{S}$. Let us first formally define these feature maps.

▶ **Definition 3.7.** *Given $p = (u, v) \in \mathbb{R}^2$, $u \leq v$, let ϕ_p be the piecewise linear hat function defined as*

$$\phi_p(t) = \begin{cases} \frac{v-u}{2}(1 - \frac{2}{v-u}|t - \frac{u+v}{2}|) & \text{if } x \leq t \leq y, \\ 0 & \text{otherwise.} \end{cases}$$

Then, given a persistence diagram Dg, let $\lambda_k : t \mapsto \max_k \{\phi_p(t)\}_{p \in \mathrm{Dg}}$, where $\max_k$ denotes the k-th largest element. The Landscape feature map *is defined as:*

$$\Phi_{\mathrm{L}} : \mathrm{Dg} \mapsto \bar{\lambda}, \quad \text{where} \quad \bar{\lambda}(x, y) = \begin{cases} \lambda_{\lceil x \rceil}(y) & x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

▶ **Definition 3.8.** *Let $\omega : \mathbb{R}^2 \to \mathbb{R}$ be a weight function and $\sigma > 0$. The* Persistence Weighted Gaussian *feature map is defined as:*

$$\Phi_{\mathrm{PWG}}^\omega : \mathrm{Dg} \mapsto \sum_{p \in \mathrm{Dg}} \omega(p) e^{-\frac{\|\cdot - p\|_2^2}{2\sigma^2}}.$$

▶ **Proposition 3.9.** *Let ω be the weight function $(x, y) \mapsto (y - x)^2$. Let $\mathcal{S}$ be the set of persistence diagrams defined in Lemma 3.6. Then:*

$$\mathcal{S} \subset \mathcal{D}_{\Phi_{\mathrm{PWG}}^\omega} \quad \text{and} \quad \mathcal{S} \subset \mathcal{D}_{\Phi_{\mathrm{L}}}.$$

Proof. Let $u_k \in \mathcal{U}$ be the sequence defined with $u_n = 1$ if $n \leq k$ and $u_n = 0$ otherwise. To show the desired result, it suffices to show that $\{\Phi_{\mathrm{PWG}}^\omega(\mathrm{Dg}_{u_k})\}_{k \in \mathbb{N}}$ and $\{\Phi_{\mathrm{L}}(\mathrm{Dg}_{u_k})\}_{k \in \mathbb{N}}$ are Cauchy sequences in $L^2(\mathbb{R}^2)$. Let $q \geq p \geq 1$, and let us study $\|\Phi(\mathrm{Dg}_{u_q}) - \Phi(\mathrm{Dg}_{u_p})\|_{L^2(\mathbb{R}^2)}^2$ for each feature map.

- Case $\Phi^\omega_{\mathrm{PWG}}$. We have the following inequalities:

$$\|\Phi^\omega_{\mathrm{PWG}}(\mathrm{Dg}_{u_q}) - \Phi^\omega_{\mathrm{PWG}}(\mathrm{Dg}_{u_p})\|^2_{L^2(\mathbb{R}^2)}$$

$$= \int_{\mathbb{R}^2} \left(\sum_{k=p}^q \frac{1}{k^2} \mathrm{e}^{-\frac{\|x-p_k\|_2^2}{2\sigma^2}} \right)^2 \mathrm{d}x = \sum_{k=p}^q \sum_{l=p}^q \frac{1}{k^2 l^2} \int_{\mathbb{R}^2} \mathrm{e}^{-\frac{\|x-p_k\|_2^2 + \|x-p_l\|_2^2}{2\sigma^2}} \mathrm{d}x$$

$$= \pi\sigma^2 \sum_{k=p}^q \sum_{l=p}^q \frac{1}{k^2 l^2} \mathrm{e}^{-\frac{\|p_k-p_l\|_2^2}{4\sigma^2}} \quad \text{(cf Appendix C in [20])}$$

$$\leq \pi\sigma^2 \left(\sum_{k=p}^q \frac{1}{k^2} \right) \left(\sum_{l=p}^q \frac{1}{l^2} \right)$$

The result simply follows from the fact that $\{\sum_{k=1}^n \frac{1}{k^2}\}_{n\in\mathbb{N}}$ is convergent and Cauchy.

- Case Φ_{L}. Since all triangular functions, as defined in Definition 3.7, have disjoint support, it follows that the only non-zero lambda function is $\lambda_1 = \sum_{n=1}^k \phi_n$, where ϕ_n is a triangular function defined with $\phi_n(t) = \frac{1}{2n}(1 - |2n(t - (n + \frac{1}{2n}))|)$ if $n \leq t \leq n + \frac{1}{n}$ and 0 otherwise. See Figure 2.
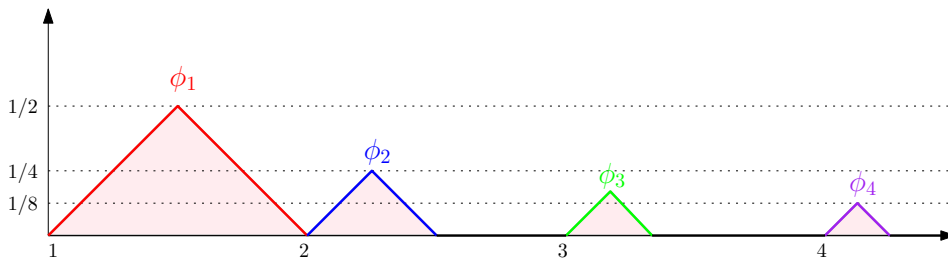

Figure 2 Image of Dg_{u_4} under Φ_{L}.

Hence, we have the following inequalities:

$$\|\Phi_{\mathrm{L}}(\mathrm{Dg}_{u_q}) - \Phi_{\mathrm{L}}(\mathrm{Dg}_{u_p})\|^2_{L^2(\mathbb{R}^2)}$$

$$= \int_{\mathbb{R}} \left(\sum_{k=p}^q \phi_k(x) \right)^2 \mathrm{d}x = \sum_{k=p}^q \sum_{l=p}^q \int_{\mathbb{R}} \phi_k(x)\phi_l(x)\mathrm{d}x$$

$$= \sum_{k=p}^q \int_{\mathbb{R}} \phi_k(x)^2 \mathrm{d}x \leq \sum_{k=p}^q \int_{\mathbb{R}} \phi_k(x)\mathrm{d}x = \sum_{k=p}^q \frac{1}{4k^2}$$

Again, the result follows from the fact that $\{\sum_{k=1}^n \frac{1}{k^2}\}_{n\in\mathbb{N}}$ is convergent and Cauchy. ◀

Proposition 3.9 shows that Theorem 3.5 applies (with the metric d_1 between persistence diagrams) to the Landscape feature map and to the Persistence Weighted Gaussian feature map with weight function $(x, y) \mapsto (y - x)^2$ – actually, any weight function that is equivalent to or dominated by $(y - x)^2$ when $(y - x)$ goes to 0. Note also that the authors in [16] suggest using weight functions of the form $(x, y) \mapsto \arctan(C|y - x|^\alpha)$, which, in this case, means that Theorem 3.5 applies if $\alpha \geq 2$. In particular, in this case, any lower bound for the Persistence Weighted Gaussian feature map has to go to 0 when $N, L \to \infty$, since an upper bound exists for this map due to its stability properties – see Corollary 3.5 in [17].

4 Mapping into finite-dimensional Hilbert spaces

In our second main result, we show that more can be said about feature maps into $\mathbb{R}^n$ (equipped with the Euclidean metric), using the so-called *Assouad dimension*. This involves all vectorization methods for persistence diagrams that we described in the related work.

Assouad dimension

▶ **Definition 4.1** (Paragraph 10.13 in [12]). *Let (X, d_X) be a metric space. Given a subset $E \subset X$ and $r > 0$, let $N_r(E)$ be the least number of open balls of radius less than or equal to r that can cover E. The* Assouad dimension *of X is:*

$$\dim_A(X, d_X) = \inf\{\alpha > 0 \,:\, \exists C > 0 \; s.t. \; \sup_{x \in X} N_{\beta r}(B(x, r)) \leq C\beta^{-\alpha}, \; \forall r > 0, \beta \in (0, 1]\}.$$

Intuitively, the Assouad dimension measures the number of open balls of radius βr needed to cover an open ball of radius r. For example, the Assouad dimension of $\mathbb{R}^n$ is n. Moreover, the Assouad dimension is preserved by bi-Lipschitz embeddings.

▶ **Proposition 4.2** (Lemma 9.6 in [22]). *Let (X, d_X) and (Y, d_Y) be metric spaces with a bi-Lipschitz embedding $\Phi : X \to Y$. Then $\dim_A(X, d_X) = \dim_A(\mathrm{im}(\Phi), d_Y)$.*

The Assouad dimension is closely related to the familiar notion of *doubling* metric space, where only the number of open balls of radius βr needed to cover an open ball of radius r is considered: A metric space (X, d_X) is *doubling* if there is a constant M such that $N_{\frac{r}{2}}(B(x, r)) \leq M$ for all $x \in X$ and $r > 0$. In terms of Assouad dimension, this is equivalent to $\dim_A(X, d_X) < \infty$ [12]. Hence, the property of being doubling is also preserved under bi-Lipschitz maps.

Non-embeddability

We now show that $\mathcal{D}_N^L$ cannot be embedded into $\mathbb{R}^n$ with bi-Lipschitz embeddings. The proof of this fact is a consequence of the following lemma:

▶ **Lemma 4.3.** *Let $p \in \mathbb{N} \cup \{\infty\}$, $N \in \mathbb{N}$, and $L > 0$. Then $\dim_A(\mathcal{D}_N^L, d_p) = \infty$.*

Proof. Let B_p denote an open ball with d_p. We want to show that, for any $\alpha > 0$ and $C > 0$, it is possible to find a persistence diagram $\mathrm{Dg} \in \mathcal{D}_N^L$, a radius $r > 0$ and a factor $\beta \in (0, 1]$ such that the number of open balls of radius at most βr needed to cover $B_p(\mathrm{Dg}, r)$ is strictly larger than $C\beta^{-\alpha}$. To this end, we pick arbitrary $\alpha > 0$ and $C > 0$. The idea of the proof is to define Dg as the empty diagram, and to derive a lower bound on the number of balls with radius βr needed to cover $B_p(\mathrm{Dg}, r)$ by considering a family of persistence diagrams $\{\mathrm{Dg}_j'\}$ with only one point each, evenly distributed on the line $\{(x, x + r) : x \in [-L, L]\}$ such that the distance between two consecutive points is r in the ℓ_∞-distance. Indeed, the pairwise distance between any two such persistence diagrams is sufficiently large so that they must belong to different balls. Then we can control the number of persistence diagrams, and thus the number of balls, by taking r sufficiently small.

More formally, choose $\beta = \frac{1}{4}$, $M = 1 + \lfloor C\beta^{-\alpha} \rfloor > C\beta^{-\alpha}$, and $r = 2L/M$. We want to show that we have at least M balls in any cover by βr, meaning that $|\{\mathrm{Dg}_i\}| \geq M$. To this end, assume we are given an arbitrary cover of $B_p(\mathrm{Dg}, r)$ with open balls of radius less than βr centered on a family $\{\mathrm{Dg}_i\}$ as follows:

$$B_p(\mathrm{Dg}, r) \subseteq \bigcup_i B_p(\mathrm{Dg}_i, \beta r). \tag{3}$$
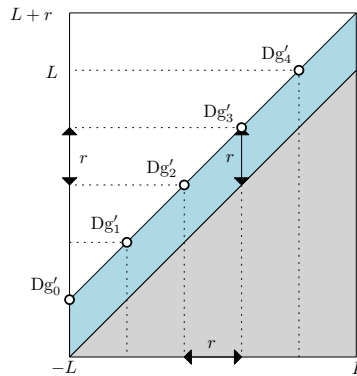
Figure 3 Persistence diagram used in the proof of Lemma 4.3. In this particular example, we have $M = 5$.

We now define a particular family of persistence diagrams which all have to lie in different elements of the cover (3). For any $0 \leq j \leq M - 1$, we let Dg'_j denote the persistence diagram containing only the point $(-L + jr, -L + (j+1)r)$, see Figure 3. It is clear that each Dg'_j is in $\mathcal{D}_N^L$.

Moreover, since $d_p(\mathrm{Dg}, \mathrm{Dg}'_j) = \frac{r}{2} < r$, it also follows that $\mathrm{Dg}'_j \in B_p(\mathrm{Dg}, r)$. Hence, according to (3), each Dg'_j is contained in some ball $B_p(\mathrm{Dg}_i, \beta r)$. Finally, note that no two $\mathrm{Dg}'_j \neq \mathrm{Dg}'_k$ can be contained in the same ball $B_p(\mathrm{Dg}_i, \beta r)$. Indeed, assuming that $\mathrm{Dg}'_j, \mathrm{Dg}'_k \in B_p(\mathrm{Dg}_i, \beta r)$, since the distance between Dg'_j and $\mathrm{Dg}'_{j'}$ is always obtained by matching their points to the diagonal, we reach a contradiction with the following application of the triangle inequality:

$$d_p(\mathrm{Dg}'_j, \mathrm{Dg}'_k) = 2^{\frac{1}{p}} \frac{r}{2} \leq d_p(\mathrm{Dg}'_j, \mathrm{Dg}_i) + d_p(\mathrm{Dg}_i, \mathrm{Dg}'_k) < 2\beta r = \frac{r}{2}.$$

This observation shows that there are at least M different open balls in the cover (3), which concludes the proof. ◀

The following theorem is then a simple consequence of Lemma 4.3 and Proposition 4.2:

▶ **Theorem 4.4.** *Let $p \in \mathbb{N} \cup \{\infty\}$ and $n \in \mathbb{N}$. Then, for any $N \in \mathbb{N}$ and $L > 0$, there is no bi-Lipschitz embedding between $(\mathcal{D}_N^L, d_p)$ and $\mathbb{R}^n$.*

Interestingly, the integers N and n are independent in Theorem 4.4: even if one restricts to persistence diagrams with only one point, it is still impossible to find a bi-Lipschitz embedding into $\mathbb{R}^n$, whatever n is.

5 Experiments

In this section, we illustrate our main results by computing the lower metric distortion bounds for the main stable feature maps in the literature. We use persistence diagrams with increasing number of points to experimentally observe the convergence of this bound to 0, as described in Theorem 3.5. More precisely, we generate 100 persistence diagrams for each cardinality in a range going from 10 to 1000 by uniformly sampling points in the unit upper half-square $\{(x, y) : 0 \leq x, y \leq 1, x \leq y\}$. See Figure 4 for an illustration.
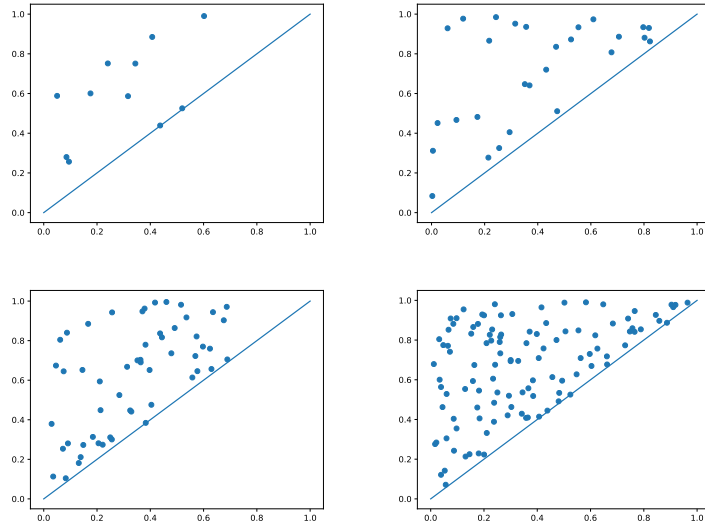
Figure 4 Example of synthetic persistence diagrams with cardinalities 10 (upper left), 30 (upper right), 60 (lower left) and 100 (lower right) generated for the experiment.

Then, we consider the following feature maps:
- the Persistence Weighted Gaussian with unit bandwidth (PWG) [17],
- the Persistence Scale Space with unit bandwidth (PSS) [21],
- the Landscape (LS) [4],
- the Persistence Image with resolution 10 x 10 and unit bandwidth (IM) [1]
- the Topological Vector with 10 dimensions (TV) [7],

Since most of these feature maps enjoy stability properties with respect to the first diagram distance d_1, we compute the ratios between the metrics in the Hilbert spaces corresponding to these feature maps and d_1. Moreover, we also look at the ratio induced by the square root of the Sliced Wasserstein distance SW between persistence diagrams [6]. Indeed, if the SW feature map is restricted to a set of persistence diagrams which are close to each other (w.r.t. the SW distance), then the distance in the corresponding Hilbert space is actually equivalent to the square root of the SW distance from the formula:

$$\|\Phi_{\mathrm{SW}}(\mathrm{Dg}) - \Phi_{\mathrm{SW}}(\mathrm{Dg}')\| = \sqrt{2\left(1 - \mathrm{e}^{-\mathrm{SW}(\mathrm{Dg},\mathrm{Dg}')}\right)}$$

Hence, we added the square root of the SW distance in our experiment. All feature maps were computed with the sklearn-tda library[1], which uses Hera[2] [15] as a backend to compute the first diagram distances d_1 between pairs of persistence diagrams. These ratios are then displayed as boxplots in Figure 5.

It is clear from Figure 5 that the extreme values of these ratios (the upper tail of the ratio distributions) increase with the cardinality of the persistence diagrams, as expected from Theorem 3.5. This is especially interesting in the case of the Sliced Wasserstein distance since the question whether the lower bound that was proved in [6], which increases with the number of points in the diagrams, was tight or not, i.e., if a lower bound which is oblivious

---

[1] `https://github.com/MathieuCarriere/sklearn_tda`
[2] `https://bitbucket.org/grey_narn/hera`
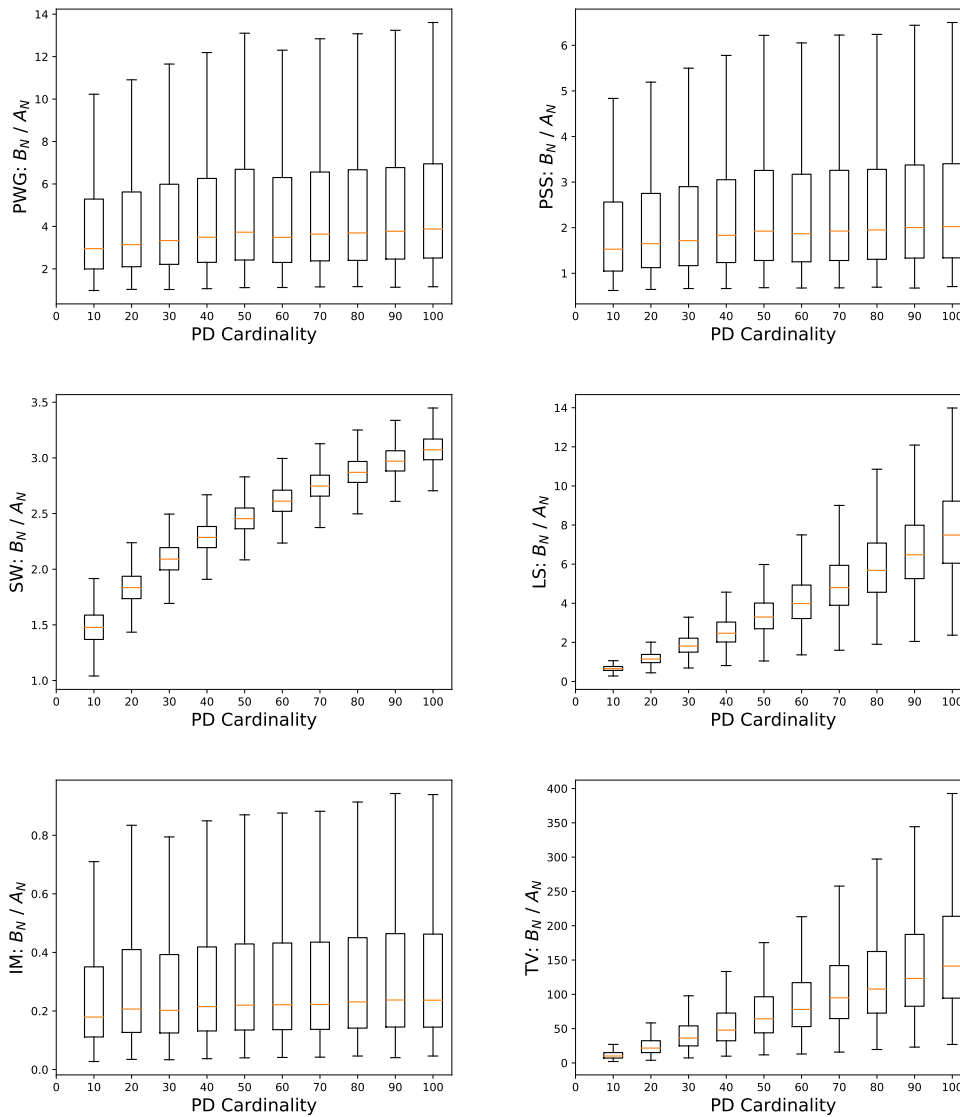
■ **Figure 5** Boxplots of the ratios between distances induced by various feature maps and the first diagram distance d_1.

to the number of points could be derived, is still open. Hence, it seems from Figure 5 that this is not the case empirically. It is also interesting to notice that the divergence speed of these ratios differ from a feature map to another. More precisely, it seems like the metric distortion bounds increase linearly with the cardinalities for the TV and LS feature maps and the Sliced Wasserstein distance, while it is increasing at a much lower speed for the other feature maps.

6   Conclusion

In this article, we provided two important theoretical results about the embedding of persistence diagrams in separable Hilbert spaces, which is a common technique in TDA to feed machine learning algorithms with persistence diagrams. Indeed, most of the recent

attempts have defined feature maps for persistence diagrams into Hilbert spaces and showed these maps were stable with respect to the first diagram distance, and conjectured whether a lower bound holds as well or not. In this work, we proved that this is never the case if the Hilbert space is finite dimensional, and that such a lower bound has to go to zero with the number of points for most other feature maps in the literature. We also provided experiments that confirm this result, by showing a clear increase of the metric distortion with the number of points for persistence diagrams generated uniformly in the unit upper half-square.

References

1   Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence Images: A Stable Vector Representation of Persistent Homology. *Journal of Machine Learning Research*, 18(8):1–35, 2017.

2   Ulrich Bauer and Michael Lesnick. Induced matchings and the algebraic stability of persistence barcodes. *Journal of Computational Geometry*, 6(2):162–191, 2015.

3   Christian Berg, Jens Christensen, and Paul Ressel. *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions.* Springer, 1984.

4   Peter Bubenik. Statistical Topological Data Analysis using Persistence Landscapes. *Journal of Machine Learning Research*, 16:77–102, 2015.

5   Gunnar Carlsson. Topology and Data. *Bulletin of the American Mathematical Society*, 46:255–308, 2009.

6   Mathieu Carrière, Marco Cuturi, and Steve Oudot. Sliced Wasserstein Kernel for Persistence Diagrams. In *Proceedings of the 34th International Conference on Machine Learning*, 2017.

7   Mathieu Carrière, Steve Oudot, and Maks Ovsjanikov. Stable Topological Signatures for Points on 3D Shapes. *Computer Graphics Forum*, 34, 2015.

8   Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. *The Structure and Stability of Persistence Modules.* Springer, 2016.

9   David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of Persistence Diagrams. *Discrete and Computational Geometry*, 37(1):103–120, 2007.

10   Barbara Di Fabio and Massimo Ferri. Comparing Persistence Diagrams Through Complex Vectors. In *Image Analysis and Processing – ICIAP 2015*, pages 294–305, 2015.

11   Herbert Edelsbrunner and John Harer. *Computational Topology: an introduction.* AMS Bookstore, 2010.

12   Juha Heinonen. *Lectures on Analysis on Metric Spaces.* Springer, 2001.

13   Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep Learning with Topological Signatures. In *Advances in Neural Information Processing Systems 30*, pages 1633–1643. Curran Associates, Inc., 2017.

14   Sara Kališnik. Tropical Coordinates on the Space of Persistence Barcodes. *Foundations of Computational Mathematics*, 2018.

15   Michael Kerber, Dmitriy Morozov, and Arnur Nigmetov. Geometry Helps to Compare Persistence Diagrams. *Journal of Experimental Algorithmics*, 22:1.4:1–1.4:20, September 2017.

16   Genki Kusano, Kenji Fukumizu, and Yasuaki Hiraoka. Persistence Weighted Gaussian Kernel for Topological Data Analysis. In *Proceedings of the 33rd International Conference on Machine Learning*, pages 2004–2013, 2016.

17   Genki Kusano, Kenji Fukumizu, and Yasuaki Hiraoka. Kernel Method for Persistence Diagrams via Kernel Embedding and Weight Factor. *Journal of Machine Learning Research*, 18(189):1–41, 2018.

18   Yuriy Mileyko, Sayan Mukherjee, and John Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12):124007, 2011.

19   Steve Oudot. *Persistence Theory: From Quiver Representations to Data Analysis.* Number 209 in Mathematical Surveys and Monographs. American Mathematical Society, 2015.

20   Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A Stable Multi-Scale Kernel for Topological Machine Learning. *CoRR*, abs/1412.6821, 2014. `arXiv:1412.6821`.

21   Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A Stable Multi-Scale Kernel for Topological Machine Learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

22   James C. Robinson. *Dimensions, Embeddings, and Attractors*, volume 186 of *Cambridge Tracts in Mathematics*. Cambridge University Press, 2010.

23   Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet Means for Distributions of Persistence Diagrams. *Discrete and Computational Geometry*, 52(1):44–70, 2014.