Computers & Graphics: X 2 (2019) 100005

Contents lists available at ScienceDirect

Computers & Graphics: X

journal homepage: www.elsevier.com/locate/cagx

Special Section on SMI 2019

A kernel for multi-parameter persistent homology

René Corbet[a], Ulderico Fugacci[a], Michael Kerber[a], Claudia Landi[b], Bei Wang[c,*]

[a] Graz University of Technology, Austria
[b] University of Modena and Reggio Emilia, Italy
[c] University of Utah, USA

ARTICLE INFO

Article history:
Received 13 March 2019
Revised 18 May 2019
Accepted 21 May 2019
Available online 31 May 2019

Keywords:
Topological data analysis
Machine learning
Persistent homology
Multivariate analysis

ABSTRACT

Topological data analysis and its main method, persistent homology, provide a toolkit for computing topological information of high-dimensional and noisy data sets. Kernels for one-parameter persistent homology have been established to connect persistent homology with machine learning techniques with applicability on shape analysis, recognition and classification. We contribute a kernel construction for multi-parameter persistence by integrating a one-parameter kernel weighted along straight lines. We prove that our kernel is stable and efficiently computable, which establishes a theoretical connection between topological data analysis and machine learning for multivariate data analysis.

1. Introduction

Topological data analysis (TDA) is an active area in data science with a growing interest and notable successes in a number of applications in science and engineering [1–8]. TDA extracts in-depth geometric information in amorphous solids [5], determines robust topological properties of evolution from genomic data sets [2] and identifies distinct diabetes subgroups [6] and a new subtype of breast cancer [9] in high-dimensional clinical data sets, to name a few. In the context of shape analysis, TDA techniques have been used in the recognition, classification [10,11], summarization [12], and clustering [13] of 2D/3D shapes and surfaces. Oftentimes, such techniques capture and highlight structures in data that conventional techniques fail to treat [11,13] or reveal properly [5].

TDA employs the mathematical notion of *simplicial complexes* [14] to encode higher order interactions in the system, and at its core uses the computational framework of *persistent homology* [15–19] to extract multi-scale topological features of the data. In particular, TDA extracts a rich set of topological features from high-dimensional and noisy data sets that complement geometric and statistical features, which offers a different perspective for machine learning. The question is, *how can we establish and enrich the theoretical connections between TDA and machine learning*?

Informally, *homology* was developed to classify topological spaces by examining their topological features such as connected components, tunnels, voids and holes of higher dimensions; *persistent homology* studies homology of a data set at multiple scales. Such information is summarized by the *persistence diagram*, a finite multi-set of points in the plane. A persistence diagram yields a complete description of the topological properties of a data set, making it an attractive tool to define features of data that take topology into consideration. Furthermore, a celebrated theorem of persistent homology is the *stability* of persistence diagrams [20] – small changes in the data lead to small changes of the corresponding diagrams, making it suitable for robust data analysis.

However, interfacing persistence diagrams directly with machine learning poses technical difficulties, because persistence diagrams contain point sets in the plane that do not have the structure of an inner product, which allows length and angle to be measured. In other words, such diagrams lack a Hilbert space structure for kernel-based learning methods such as kernel SVMs or PCAs [21]. Recent work proposes several variants of *feature maps* [21–23] that transform persistence diagrams into L^2-functions over $\mathbb{R}^2$. This idea immediately enables the application of topological features for kernel-based machine learning methods as establishing a kernel function implicitly defines a Hilbert space structure [21].

A serious limit of standard persistent homology and its initial interfacing with machine learning [21–25] is the restriction to only a single scale parameter, thereby confining its applicability to the univariate setting. However, in many real-world applications, such as data acquisition and geometric modeling, we often encounter richer information described by multivariate data sets [26–28]. Consider, for example, climate simulations where multiple physical parameters such as temperature and pressure are computed simultaneously; and we are interested in understanding

* Corresponding author.
E-mail addresses: maths@rene-corbet.de (R. Corbet), kerber@tugraz.at (M. Kerber), clandi@unimore.it (C. Landi), beiwang@sci.utah.edu (B. Wang).

https://doi.org/10.1016/j.cagx.2019.100005

the interplay between these parameters. Consider another example in multivariate shape analysis, various families of functions carry information about the geometry of 3D shape objects, such as mesh density, eccentricity [29] or Heat Kernel Signature [30]; and we are interested in creating multivariate signatures of shapes from such functions. Unlike the univariate setting, very few topological tools exist for the study of multivariate data [29,31,32], let alone the integration of multivariate topological features with machine learning.

The active area of *multi-parameter persistent homology* [26] studies the extension of persistence to two or more (independent) scale parameters. A complete discrete invariant such as the persistence diagram does not exist for more than one parameter [26]. To gain partial information, it is common to study *slices*, that is, one-dimensional affine subspaces where all parameters are connected by a linear equation. In this paper, we establish, for the first time, a theoretical connection between topological features and machine learning algorithms via the kernel approach for multi-parameter persistent homology. Such a theoretical underpinning is necessary for applications in multivariate data analysis.

Our contribution. We propose the first kernel construction for multi-parameter persistent homology. Our kernel is *generic*, *stable* and can be *approximated in polynomial time*. For simplicity, we formulate all our results for the case of two parameters, although they extend to more than two parameters.

Our input is a data set that is filtered according to two scale parameters and has a finite description size; we call this a *bi-filtration* and postpone its formal definition to Section 2. Our main contribution is the definition of a feature map that assigns to a bi-filtration $\mathcal{X}$ a function $\Phi_{\mathcal{X}} : \Delta^{(2)} \to \mathbb{R}$, where $\Delta^{(2)}$ is a subset of $\mathbb{R}^4$. Moreover, $\Phi^2_{\mathcal{X}}$ is integrable over $\Delta^{(2)}$, effectively including the space of bi-filtrations into the Hilbert space $L^2(\Delta^{(2)})$. Therefore, based on the standard scalar product in $L^2(\Delta^{(2)})$, a 2-parameter kernel is defined such that for two given bi-filtrations $\mathcal{X}$ and $\mathcal{Y}$ we have

$$\langle \mathcal{X}, \mathcal{Y} \rangle_{\Phi} := \int_{\Delta^{(2)}} \Phi_{\mathcal{X}} \Phi_{\mathcal{Y}} d\mu. \tag{1}$$

We construct our feature map by interpreting a point of $\Delta^{(2)}$ as a pair of (distinct) points in $\mathbb{R}^2$ that define a unique slice. Along this slice, the data simplifies to a *mono-filtration* (i.e., a filtration that depends on a single scale parameter), and we can choose among a large class of feature maps and kernel constructions of standard, one-parameter persistence. To make the feature map well-defined, we restrict our attention to a finite rectangle R.

Our inclusion into a Hilbert space induces a distance between bi-filtrations as

$$d_{\Phi}(\mathcal{X}, \mathcal{Y}) := \sqrt{\int (\Phi_{\mathcal{X}} - \Phi_{\mathcal{Y}})^2 d\mu}. \tag{2}$$

We prove a stability bound, relating this distance measure to the matching distance and the interleaving distance (see the paragraph on related work below). We also show that this stability bound is tight up to constant factors (see Section 4).

Finally, we prove that our kernel construction admits an efficient approximation scheme. Fixing an absolute error bound ϵ, we give a polynomial time algorithm in $1/\epsilon$ and the size of the bi-filtrations $\mathcal{X}$ and $\mathcal{Y}$ to compute a value r such that $r \leq \langle \mathcal{X}, \mathcal{Y} \rangle_{\Phi} \leq r + \epsilon$. On a high level, the algorithm subdivides the domain into boxes of smaller and smaller width and evaluates the integral of (1) by lower and upper sums within each subdomain, terminating the process when the desired accuracy has been achieved. The technical difficulty lies in the accurate and certifiable approximation of the variation of the feature map when moving the argument within a subdomain.

Related work. Our approach heavily relies on the construction of stable and efficiently computable feature maps for mono-filtrations. This line of research was started by Reininghaus et al. [21], whose approach we discuss in some detail in Section 2. Alternative kernel constructions appeared in [24,33]. Kernel constructions fit into the general framework of including the space of persistence diagrams in a larger space with more favorable properties. Other examples of this idea are persistent landscapes [22] and persistent images [34], which can be interpreted as kernel constructions as well. Kernels and related variants defined on mono-filtrations have been used to discriminate and classify shapes and surfaces [21,25]. An alternative approach comes from the definition of suitable (polynomial) functions on persistence diagrams to arrive at a fixed-dimensional vector in $\mathbb{R}^d$ on which machine learning tasks can be performed; see [35–38].

As previously mentioned, a persistence diagram for multi-parameter persistence does not exist [26]. However, bi-filtrations still admit meaningful distance measures, which lead to the notion of closeness of two bi-filtrations. The most prominent such distance is the *interleaving distance* [39], which, however, has recently been proved to be NP-complete to compute and approximate [40]. Computationally attractive alternatives are (multi-parameter) bottleneck distance [41] and the *matching distance* [42,43], which compares the persistence diagrams along all slices (appropriately weighted) and picks the worst discrepancy as the distance of the bi-filtrations. This distance can be approximated up to a precision ϵ using an appropriate subsample of the lines [42], and also computed exactly in polynomial time [43]. Our approach extends these works in the sense that not just a distance, but an inner product on bi-filtrations, is defined with our inclusion into a Hilbert space. In a similar spirit, the software library RIVET [44] provides a visualization tool to explore bi-filtrations by scanning through the slices.

2. Preliminaries

We introduce the basic topological terminology needed in this work. We restrict ourselves to the case of simplicial complexes as input structures for a clearer geometric intuition of the concepts, but our results generalize to more abstract input types (such as minimal representations of persistence modules) without problems.

Mono-filtrations. Given a vertex set V, an *(abstract) simplex* is a non-empty subset of V, and an *(abstract) simplicial complex* is a collection of such subsets that is closed under the operation of taking non-empty subsets. A *subcomplex* of a simplicial complex X is a simplicial complex Y with $Y \subseteq X$. Fixing a finite simplicial complex X, a *mono-filtration* $\mathcal{X}$ of X is a map that assigns to each real number α, a subcomplex $\mathcal{X}(\alpha)$ of X, with the property that whenever $\alpha \leq \beta$, $\mathcal{X}(\alpha) \subseteq \mathcal{X}(\beta)$. The *size* of $\mathcal{X}$ is the number of simplices of X. Since X is finite, $\mathcal{X}(\alpha)$ changes at only finitely many places when α grows continuously from $-\infty$ to $+\infty$; we call these values *critical*. More formally, α is *critical* if there exists no open neighborhood of α such that the mono-filtration assigns the identical subcomplex to each value in the neighborhood. For a simplex σ of X, we call the *critical value* of σ the infimum over all α for which $\sigma \in \mathcal{X}(\alpha)$. For simplicity, we assume that this infimum is a minimum, so every simplex has a unique critical value wherever it is included in the mono-filtration.

Bi-filtrations. For points in $\mathbb{R}^2$, we write $(a, b) \leq (c, d)$ if $a \leq c$ and $b \leq d$. Similarly, we say $(a, b) < (c, d)$ if $a < c$ and $b < d$. For a finite simplicial complex X, a *bi-filtration* $\mathcal{X}$ of X is a map that assigns to each point $p \in \mathbb{R}^2$ a subcomplex $\mathcal{X}(p)$ of X, such that whenever $p \leq q$, $\mathcal{X}(p) \subseteq \mathcal{X}(q)$. Again, a point $p = (p_1, p_2)$ is called *critical* for $\mathcal{X}$ if, for any $\epsilon > 0$, both $\mathcal{X}(p_1 - \epsilon, p_2)$ and $\mathcal{X}(p_1, p_2 - \epsilon)$

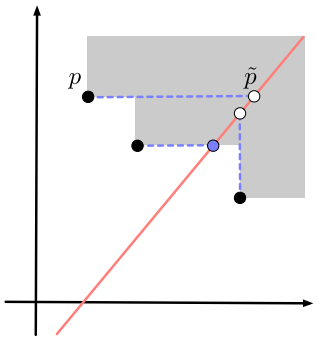

Fig. 1. The three black points mark the three critical points of some simplex σ in X. The shaded area denotes the positions at which σ is present in the bi-filtration. Along the given slice (red line), the dashed lines denote the first position where the corresponding critical point "affects" the slice. This position is either the upper-vertical, or right-horizontal projection of the critical point onto the slice, depending on whether the critical point is below or above the line. For σ, we see that it enters the slice at the position marked by the blue point.

are not identical to $\mathcal{X}(p)$. Note that unlike in the mono-filtration case, the set of critical points might not be finite. We call a bi-filtration *tame* if it has only finitely many such critical points. For a simplex σ, a point $p \in \mathbb{R}^2$ is *critical* for σ if, for any $\epsilon > 0$, σ is neither in $\mathcal{X}(p_1 - \epsilon, p_2)$ nor in $\mathcal{X}(p_1, p_2 - \epsilon)$, whereas σ is in both $\mathcal{X}(p_1 + \epsilon, p_2)$ and $\mathcal{X}(p_1, p_2 + \epsilon)$. Again, for simplicity, we assume that $\sigma \in \mathcal{X}(p)$ in this case. A consequence of tameness is that each simplex has a finite number of critical points. Therefore, we can represent a tame bi-filtration of a finite simplicial complex X by specifying the set of critical points for each simplex in X. The sum of the number of critical points over all simplices of X is called the *size* of the bi-filtration. We henceforth assume that bi-filtrations are always represented in this form; in particular, we assume tameness throughout this paper.

A standard example to generate bi-filtrations is by an arbitrary function $F : X \to \mathbb{R}^2$ with the property that if $\tau \subset \sigma$ are two simplices of X, $F(\tau) \leq F(\sigma)$. We define the *sublevel set* $\mathcal{X}^F(p)$ as

$$\mathcal{X}^F(p) := \{\sigma \in X \mid F(\sigma) \leq p\},$$

and let $\mathcal{X}^F$ denote its corresponding *sublevel set bi-filtration*. It is easy to verify that $\mathcal{X}^F$ yields a (tame) bi-filtration and $F(\sigma)$ is the unique critical value of σ in the bi-filtration.

Slices of a bi-filtration. A bi-filtration $\mathcal{X}$ contains an infinite collection of mono-filtrations. Let $\mathcal{L}$ be the set of all non-vertical lines in $\mathbb{R}^2$ with positive slope. Fixing any line $\ell \in \mathcal{L}$, we observe that when traversing this line in positive direction, the subcomplexes of the bi-filtration are nested in each other. Note that ℓ intersects the anti-diagonal $x = -y$ in a unique base point b. Parameterizing ℓ as $b + \lambda \cdot a$, where a is the (positive) unit direction vector of ℓ, we obtain the mono-filtration

$$\mathcal{X}_\ell(\alpha) := \mathcal{X}(b + \alpha \cdot a).$$

We will refer to this mono-filtration $\mathcal{X}_\ell$ as a *slice* of $\mathcal{X}$ along ℓ (and sometimes also call ℓ itself the slice, abusing notation). The critical values of a slice can be inferred by the critical points of the bi-filtration in a computationally straightforward way. Instead of a formal description, we refer to Fig. 1 for a graphical description. Also, if the bi-filtration is of size n, each of its slices is of size at most n.

Persistent homology. A mono-filtration $\mathcal{X}$ gives rise to a persistence diagram. Formally, we obtain this diagram by applying the homology functor to $\mathcal{X}$, yielding a sequence of vector spaces and linear maps between them, and splitting this sequence into indecomposable parts using representation theory. Instead of rolling out the entire theory (which is explained, for instance, in [45]), we give an intuitive description here.

Persistent homology measures how the topological features of a data set evolve when considered across a varying scale parameter α. The most common example involves a point cloud in $\mathbb{R}^d$, where considering a fixed scale α means replacing the points by balls of radius α. As α increases, the data set undergoes various topological configurations, starting as a disconnected point cloud for $\alpha = 0$ and ending up as a topological ball when α approaches ∞; see Fig. 2(a) for an example in $\mathbb{R}^2$.

The topological information of this process can be summarized as a finite multi-set of points in the plane, called the *persistence diagram*. Each point of the diagram corresponds to a topological feature (i.e., connected components, tunnels, voids, etc.), and its coordinates specify at which scales the feature appears and disappears in the data. As illustrated in Fig. 2(a), all five (connected) components are born (i.e., appear) at $\alpha = 0$. The green component dies (i.e., disappears) when it merges with the red component at $\alpha = 2.5$; similarly, the orange, blue and pink components die at scales 3, 3.2 and 3.7, respectively. The red component never dies as α goes to ∞. The 0-dimensional persistence diagram is defined to have one point per component with birth and death value as its coordinates (Fig. 2(c)). The *persistence* of a feature is then merely its distance from the diagonal. While we focus on the components, the concept generalizes to higher dimensions, such as tunnels (1-dimensional homology) and voids (2-dimensional homology). For instance, in Fig. 2(a), a tunnel appears at $\alpha = 4.2$ and disappears at $\alpha = 5.6$, which gives rise to a purple point (4.2, 5.6) in the 1-dimensional persistence diagram (Fig. 2(c)).

From a computational point of view, the nested sequence of spaces formed by unions of balls (Fig. 2(a)) can be replaced by a nested sequence of simplicial complexes by taking their nerves, thereby forming a mono-filtration of simplicial complexes that captures the same topological information but has a much smaller footprint (Fig. 2(b)).

In the context of shape analysis, we apply persistent homology to capture the topological information of 2D and 3D shape objects by employing various types of mono-filtrations. A simple example is illustrated in Fig. 3: we extract point clouds sampled from the boundary of 2D shape objects and compute the persistence diagrams using Vietoris-Rips complex filtrations.

Stability of persistent homology. Bottleneck distance represents a similarity measure between persistence diagrams. Let D, D' be two persistence diagrams. Without loss of generality, we can assume that both contain infinitely many copies of the points on the diagonal. The *bottleneck distance* between D and D' is defined as

$$d_B(D, D') := \inf_{\gamma} \sup_{x \in D} \|x - \gamma(x)\|_\infty, \tag{3}$$

where γ ranges over all bijections from D to D'. We will also use the notation $d_B(\mathcal{X}, \mathcal{Y})$ for two mono-filtrations instead of $d_B(D(\mathcal{X}), D(\mathcal{Y}))$

A crucial result for persistent homology is the *stability theorem* proven in [46] and re-stated in our notation as follows. Given two functions $f, g : X \to \mathbb{R}$ whose sublevel sets form two mono-filtrations of a finite simplicial complex X, the induced persistence diagrams satisfy

$$d_B(D_f, D_g) \leq \|f - g\|_\infty := \sup_{\sigma \in X} |f(\sigma) - g(\sigma)|. \tag{4}$$

Feature maps for mono-filtrations. Several feature maps aimed at the construction of a kernel for mono-filtrations have been proposed in the literature [21–23]. We discuss one example: the persistence scale-space kernel [21] assigns to a mono-filtration $\mathcal{X}$ an L^2-function $\phi_{\mathcal{X}}$ defined on $\Delta^{(1)} := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 < x_2\}$.

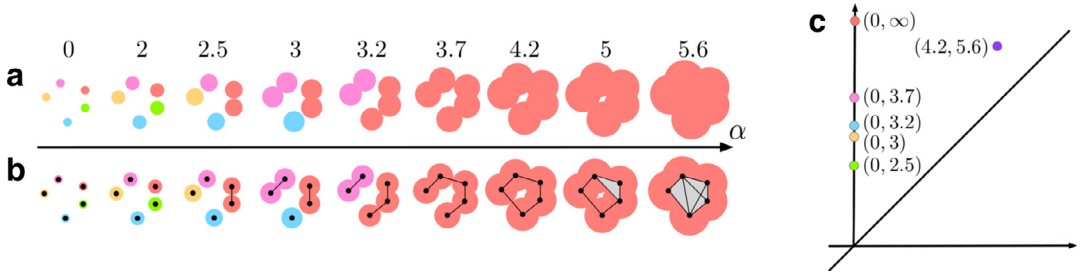

Fig. 2. Computing persistent homology of a point cloud in $\mathbb{R}^2$. (a) A nested sequence of topological spaces formed by unions of balls at increasing parameter values. (b) A mono-filtration of simplicial complexes that captures the same topological information as in (a). (c) 0-dimensional and 1-dimensional persistence diagrams combined.

The main idea behind the definition of $\phi_{\mathcal{X}}$ is to define a sum of Gaussian peaks, all of the same height and width, with each peak centered at one finite off-diagonal point of the persistence diagram $D(\mathcal{X})$ of $\mathcal{X}$. To make the construction robust against perturbations, the function has to be equal to 0 across the diagonal (the boundary of $\Delta^{(1)}$). This is achieved by adding negative Gaussian peaks at the reflections of the off-diagonal points along the diagonal. Writing $\bar{z}$ for the reflection of a point z, we obtain the formula,

$$\phi_{\mathcal{X}}(x) := \frac{1}{4\pi t} \sum_{z \in D(\mathcal{X})} e^{\frac{\|x-z\|_2^2}{4t}} - e^{\frac{\|x-\bar{z}\|_2^2}{4t}}, \tag{5}$$

where t is the width of the Gaussian, which is a free parameter of the construction. See Fig. 4(b) and (c) for an illustration of a transformation of a persistence diagram to the function $\phi_{\mathcal{X}}$. The induced kernel enjoys several stability properties and can be evaluated efficiently without explicit construction of the feature map; see [21] for details.

More generally, in this paper, we look at the class of all feature maps that assign to a mono-filtration $\mathcal{X}$ a function in $L^2(\Delta^{(1)})$. For such a feature map $\phi_{\mathcal{X}}$, we define the following properties:

- *Absolutely boundedness.* There exists a constant $v_1 > 0$ such that, for any mono-filtration $\mathcal{X}$ of size n and any $x \in \Delta^{(1)}$, $0 \leq \phi_{\mathcal{X}}(x) \leq v_1 \cdot n$.
- *Lipschitzianity.* There exists a constant $v_2 > 0$ such that, for any mono-filtration $\mathcal{X}$ of size n and any $x, x' \in \Delta^{(1)}$, $|\phi_{\mathcal{X}}(x) - \phi_{\mathcal{X}}(x')| \leq v_2 \cdot n \cdot \|x - x'\|_2$.
- *Internal stability.* There exists a constant $v_3 > 0$ such that, for any pair of mono-filtrations $\mathcal{X}, \mathcal{Y}$ of size n and any $x \in \Delta^{(1)}$, $|\phi_{\mathcal{X}}(x) - \phi_{\mathcal{Y}}(x)| \leq v_3 \cdot n \cdot d_B(\mathcal{X}, \mathcal{Y})$.
- *Efficiency.* For any $x \in \Delta^{(1)}$, $\phi_{\mathcal{X}}(x)$ can be computed in polynomial time in the size of $\mathcal{X}$, that is, in $O(n^k)$ for some $k \geq 0$.

It can be verified easily that the scale-space feature map from above satisfies all these properties. The same is true, for instance, if the Gaussian peaks are replaced by linear peaks (that is, replacing the Gaussian kernel in (5) by a triangle kernel).

3. A feature map for multi-parameter persistent homology

Let ϕ be a feature map (such as the scale-space kernel) that assigns to a mono-filtration a function in $L^2(\Delta^{(1)})$. Starting from ϕ, we construct a feature map Φ on the set of all bi-filtrations Ω that has values in a Hilbert space.

The feature map Φ assigns to a bi-filtration $\mathcal{X}$ a function $\Phi_{\mathcal{X}} : \Delta^{(2)} \to \mathbb{R}$. We set

$$\Delta^{(2)} := \left\{(p, q) \mid p \in \mathbb{R}^2, q \in \mathbb{R}^2, p < q\right\}$$

as the set of all pairs of points where the first point is smaller than the second one. $\Delta^{(2)}$ can be interpreted naturally as a subset of $\mathbb{R}^4$, but we will usually consider elements of $\Delta^{(2)}$ as pairs of points in $\mathbb{R}^2$.

Fixing $(p, q) \in \Delta^{(2)}$, let ℓ denote the unique slice through these two points. Along this slice, the bi-filtration gives rise to a mono-filtration $\mathcal{X}_\ell$, and consequently a function $\phi_{\mathcal{X}_\ell} : \Delta^{(1)} \to \mathbb{R}$ using the considered feature map for mono-filtrations. Moreover, using the parameterization of the slice ℓ as $b + \lambda \cdot a$ from Section 2, there exist real values λ_p, λ_q such that $b + \lambda_p a = p$ and $b + \lambda_q a = q$. Since $p < q$ and $\lambda_p < \lambda_q$, hence $(\lambda_p, \lambda_q) \in \Delta^{(1)}$. We define $\Phi_{\mathcal{X}}(p, q)$ to be the weighted function value of $\phi_{\mathcal{X}_\ell}$ at (λ_p, λ_q) (see also Fig. 4), that is,

$$\Phi_{\mathcal{X}}(p, q) := w(p, q) \cdot \phi_{\mathcal{X}_\ell}(\lambda_p, \lambda_q), \tag{6}$$

where $w(p, q)$ is a weight function $w : \Delta^{(2)} \to \mathbb{R}$ defined below.

The weight function w has two components. First, let R be a bounded axis-aligned rectangle in $\mathbb{R}^2$; its bottom-left corner coincides with the origin of the coordinate axes. We define w such that its weight is 0 if p or q is outside of R. Second, for pairs of points within $R \times R$, we assign a weight depending on the slope of the induced slices. Formally, let ℓ be parameterized as $b + \lambda \cdot a$ as above, and recall that a is a unit vector with non-negative coordinates. Write $a = (a_1, a_2)$ and set $\hat{\ell} := \min\{a_1, a_2\}$. Then, we define

$$w(p, q) := \chi_R(p) \cdot \chi_R(q) \cdot \hat{\ell},$$

where χ_R is the characteristic function of R, mapping a point x to 1 if $x \in R$ and 0 otherwise.

The factor $\hat{\ell}$ ensures that slices that are close to being horizontal or vertical attain less importance in the feature map. The same weight is assigned to slices in the matching distance [42]. $\hat{\ell}$ is not important for obtaining an L^2-function, but its meaning will become clear in the stability results of Section 4. We also remark that the largest weight is attained for the diagonal slice with a value of $1/\sqrt{2}$. Consequently, w is a non-negative function upper bounded by $1/\sqrt{2}$.

To summarize, our map Φ depends on the choice of an axis-aligned rectangle R and a choice of feature map for mono-filtrations, which itself might have associated parameters. For instance, using the scale-space feature map requires the choice of the width t (see (5)). It is only left to argue that the image of the feature map Φ is indeed an L^2-function.

Theorem 1. *If ϕ is absolutely bounded, then $\Phi_{\mathcal{X}}$ is in $L^2(\Delta^{(2)})$.*

Proof. Let $\mathcal{X}$ be a bi-filtration of size n. As mentioned earlier, each slice $\mathcal{X}_\ell$ is of a size at most n. By absolute boundedness and the fact that the weight function is upper bounded by $\frac{1}{\sqrt{2}}$, it follows that $|\Phi_{\mathcal{X}}(p, q)| \leq \frac{v_1 n}{\sqrt{2}}$ for all (p, q). Since the support of $\Phi_{\mathcal{X}}$ is compact $(R \times R)$, the integral of $\Phi_{\mathcal{X}}^2$ over $\Delta^{(2)}$ is finite, being absolutely bounded and compactly supported. □

Note that Theorem 1 remains true even without restricting the weight function to R, provided we consider a weight

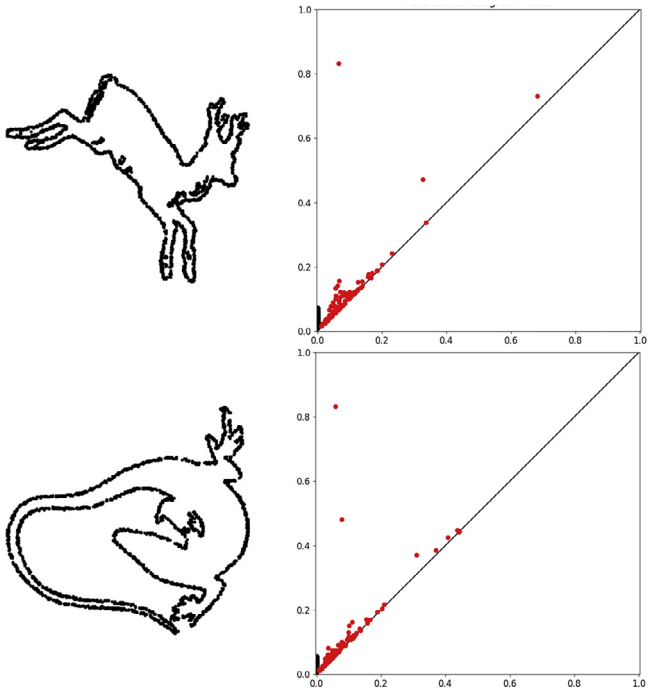

Fig. 3. The persistence diagrams of 2D shape objects. Black and red points are 0-dimensional and 1-dimensional features respectively (ignoring points with ∞ persistence).

function that is square-integrable over $\Delta^{(2)}$. We skip the (easy) proof.

4. Stability

An important and desirable property for a kernel is its stability. In general, stability means that small perturbations in the input data imply small perturbations in the output data. In our setting, small changes between multi-filtrations (with respect to matching distance) should not induce large changes in their corresponding feature maps (with respect to L^2 distance).

Adopted to our notation, the matching distance is defined as

$$d_{match}(\mathcal{X},\mathcal{Y}) = \sup_{\ell\in\mathcal{L}} \big(\hat{\ell}\cdot d_B(\mathcal{X}_\ell,\mathcal{Y}_\ell)\big),$$

where $\mathcal{L}$ is the set of non-vertical lines with positive slope [47].

Theorem 2. *Let $\mathcal{X}$ and $\mathcal{Y}$ be two bi-filtrations. If ϕ is absolutely bounded and internally stable, we have*

$$\|\Phi_{\mathcal{X}} - \Phi_{\mathcal{Y}}\|_{L^2} \le C\cdot n\cdot \mathrm{area}(R)\cdot d_{match}(\mathcal{X},\mathcal{Y}),$$

for some constant C.

Proof. Absolute boundedness ensures that the left-hand side is well-defined by Theorem 1. Now we use the definition of $\|\cdot\|_{L^2}$ and the internal stability of ϕ to obtain

$$\begin{aligned}\|\Phi_{\mathcal{X}}-\Phi_{\mathcal{Y}}\|_{L^2}^2 &= \int_{\Delta^{(2)}} \big|w(p,q)\cdot\phi_{\mathcal{X}_\ell}(\lambda_p,\lambda_q) - w(p,q)\cdot\phi_{\mathcal{Y}_\ell}(\lambda_p,\lambda_q)\big|^2 d\mu\\ &\le \int_{\Delta^{(2)}} (w(p,q)\cdot v_3\cdot n\cdot d_B(\mathcal{X}_\ell,\mathcal{Y}_\ell))^2 d\mu\\ &= (v_3\cdot n)^2 \int_{\Delta^{(2)}} (w(p,q)\cdot d_B(\mathcal{X}_\ell,\mathcal{Y}_\ell))^2 d\mu\end{aligned}$$

Since $w(p,q)$ is zero outside $R\times R$, the integral does not change when restricted to $\Delta^{(2)}\cap(R\times R)$. Within this set, $w(p,q)$ simplifies to $\hat{\ell}$, with ℓ the line through p and q. Hence, we can further bound

$$\begin{aligned}&= (v_3\cdot n)^2 \int_{\Delta^{(2)}\cap(R\times R)} (\hat{\ell}\cdot d_B(\mathcal{X}_\ell,\mathcal{Y}_\ell))^2 d\mu\\ &\le (v_3\cdot n)^2 \int_{\Delta^{(2)}\cap(R\times R)} \underbrace{\sup_{\ell\in\mathcal{L}} \big(\hat{\ell}\cdot d_B(\mathcal{X}_\ell,\mathcal{Y}_\ell)\big)^2}_{=d_{match}(\mathcal{X},\mathcal{Y})} d\mu\\ &= (v_3\cdot n\cdot d_{match}(\mathcal{X},\mathcal{Y}))^2 \int_{\Delta^{(2)}\cap(R\times R)} 1 d\mu.\end{aligned}$$

The claimed inequality follows by noting that the final integral is equal to $\frac{1}{4}\mathrm{area}(R)^2$. □

As a corollary, we get the the same stability statement with respect to interleaving distance instead of matching distance [48, Thm.1]. Furthermore, we obtain a stability bound for sublevel set bi-filtrations of functions $X\to\mathbb{R}^2$ [47, Thm.4]:

Corollary 3. *Let $F, G : X\to\mathbb{R}^2$ be two functions that give rise to sublevel set bi-filtrations $\mathcal{X}$ and $\mathcal{Y}$, respectively. If ϕ is absolutely bounded and internally stable, we have*

$$\|\Phi_{\mathcal{X}} - \Phi_{\mathcal{Y}}\|_{L^2} \le C\cdot n\cdot \mathrm{area}(R)\cdot \|F-G\|_\infty,$$

for some constant C.

We remark that the appearance of n in the stability bound is not desirable as the bound worsens when the complex size increases (unlike, for instance, the bottleneck stability bound in (4), which is independent of n). The factor of n comes from the internal stability property of ϕ, so we have to strengthen this condition on ϕ. However, we show that such an improvement is impossible for a large class of "reasonable" feature maps.

For two bi-filtrations $\mathcal{X},\mathcal{Y}$ we define $\mathcal{X}\oplus\mathcal{Y}$ by setting $(\mathcal{X}\oplus\mathcal{Y})(p) := \mathcal{X}(p)\sqcup\mathcal{Y}(p)$ for all $p\in\mathbb{R}^2$. A feature map Φ is *additive* if $\Phi_{\mathcal{X}\oplus\mathcal{Y}} = \Phi(\mathcal{X})+\Phi(\mathcal{Y})$ for all bi-filtrations $\mathcal{X},\mathcal{Y}$. Φ is called *non-trivial* if there is a bi-filtration $\mathcal{X}$ such that $\|\Phi\|_{L^2}\neq 0$. Additivity and non-triviality for feature maps ϕ on mono-filtrations is defined in the analogous way. Note that, for instance, the scale space feature map is additive. Moreover, because $(\mathcal{X}\oplus\mathcal{Y})_\ell = \mathcal{X}_\ell\oplus\mathcal{Y}_\ell$ for every slice ℓ, a feature map Φ is additive if the underlying ϕ is additive.

For mono-filtrations, no additive, non-trivial feature map ϕ can satisfy

$$\|\phi_{\mathcal{X}} - \phi_{\mathcal{Y}}\| \le C\cdot n^\delta\cdot d_B(\mathcal{X},\mathcal{Y})$$

with $\mathcal{X},\mathcal{Y}$ mono-filtrations and $\delta\in[0,1)$; the proof of this statement is implicit in [21, Thm 3]. With similar ideas, we show that the same result holds in the multi-parameter case.

Theorem 4. *If Φ is additive and there exists $C>0$ and $\delta\in[0,1)$ such that*

$$\|\Phi_{\mathcal{X}} - \Phi_{\mathcal{Y}}\|_{L^2} \le C\cdot n^\delta\cdot d_{match}(\mathcal{X},\mathcal{Y})$$

for all bi-filtrations $\mathcal{X}$ and $\mathcal{Y}$, then Φ is trivial.

Proof. Assume to the contrary that there exists a bi-filtration $\mathcal{X}$ such that $\|\Phi_{\mathcal{X}}\|_{L^2}>0$. Then, writing $\mathcal{O}$ for the empty bi-filtration, by additivity we get $\|\Phi_{\sqcup_{i=1}^n\mathcal{X}} - \Phi_{\mathcal{O}}\|_{L^2} = n\|\Phi_{\mathcal{X}}-\Phi_{\mathcal{O}}\|_{L^2} > 0$. On the other hand, $d_{match}(\sqcup_{i=1}^n\mathcal{X},\mathcal{O}) = d_{match}(\mathcal{X},\mathcal{O})$. Hence, with C and δ as in the statement of the theorem,

$$\begin{aligned}\frac{\|\Phi_{\sqcup_{i=1}^n\mathcal{X}} - \Phi_{\mathcal{O}}\|_{L^2}}{C\cdot n^\delta\cdot d_{match}(\sqcup_{i=1}^n\mathcal{X},\mathcal{O})} &= \frac{n\|\Phi_{\mathcal{X}}-\Phi_{\mathcal{O}}\|_{L^2}}{C\cdot n^\delta\cdot d_{match}(\mathcal{X},\mathcal{O})}\\ &= n^{1-\delta}\frac{\|\Phi_{\mathcal{X}}-\Phi_{\mathcal{O}}\|_{L^2}}{C\cdot d_{match}(\mathcal{X},\mathcal{O})} \xrightarrow{n\to\infty} \infty,\end{aligned}$$

a contradiction. □

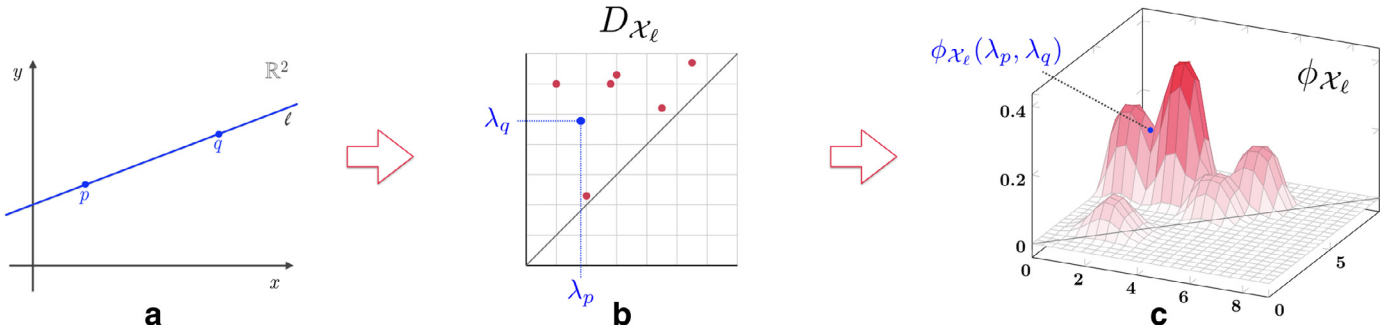

Fig. 4. An illustration of the construction of a feature map for multi-parameter persistent homology. (a) Given a bi-filtration $\mathcal{X}$ and a point $(p, q) \in \Delta^{(2)}$, the line ℓ passing through them is depicted and the parameter λ_p and λ_q computed. (b) The point (λ_p, λ_q) is embedded in the persistence diagram of the mono-filtration $\mathcal{X}_\ell$ obtained as the slice of $\mathcal{X}$ along ℓ. (c) The point (λ_p, λ_q) is assigned the value $\phi_{\mathcal{X}_\ell}(\lambda_p, \lambda_q)$ via the feature map ϕ.

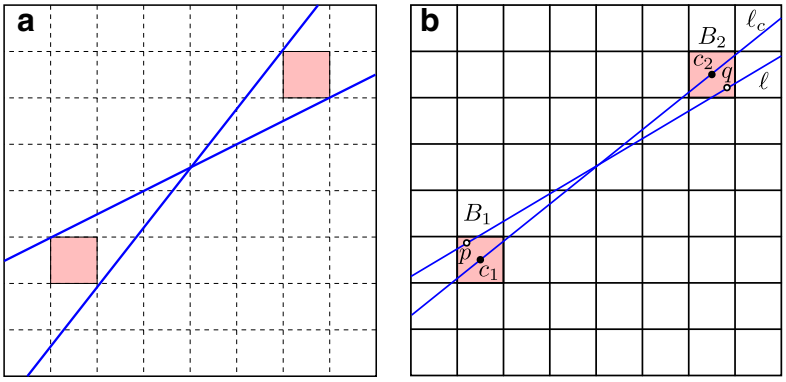

Fig. 5. (a) The two given slices realize the largest and smallest possible slope among all slices traversing the pink box pair. It can be easily seen that the difference of the unit vector of the center line to one of the unit vectors of these two lines realizes A for the given box pair. (b) Computing variations for the center slice and a traversing slice of a box pair.

5. Approximability

We provide an approximation algorithm to compute the kernel of two bi-filtrations $\mathcal{X}$ and $\mathcal{Y}$ up to any absolute error $\epsilon > 0$. Recall that our feature map Φ depends on the choice of a bounding box R. In this section, we assume R to be the unit square $[0, 1] \times [0, 1]$ for simplicity. We prove the following theorem that shows our kernel construction admits an efficient approximation scheme that is polynomial in $1/\epsilon$ and the size of the bi-filtrations.

Theorem 5. *Assume ϕ is absolutely bounded, Lipschitz, internally stable and efficiently computable. Given two bi-filtrations $\mathcal{X}$ and $\mathcal{Y}$ of size n and $\epsilon > 0$, we can compute a number r such that $r \le \langle \mathcal{X}, \mathcal{Y} \rangle_\Phi \le r + \epsilon$ in polynomial time in n and $1/\epsilon$.*

The proof of Theorem 5 will be illustrated in the following paragraphs, postponing most of the technical details to Appendix A.

Algorithm. Given two bi-filtrations $\mathcal{X}$ and $\mathcal{Y}$ of size n and $\epsilon > 0$, our goal is to efficiently approximate $\langle \mathcal{X}, \mathcal{Y} \rangle_\Phi$ by some number r. On the highest level, we compute a sequence of approximation intervals (with decreasing lengths) $J_1, J_2, J_3, \ldots$, each containing the desired kernel value $\langle \mathcal{X}, \mathcal{Y} \rangle_\Phi$. The computation terminates as soon as we find some J_i of width at most ϵ, in which case we return the left endpoint as an approximation to r.

For $s \in \mathbb{N}$ ($\mathbb{N}$ being the set of natural numbers), we compute J_s as follows. We split R into $2^s \times 2^s$ congruent squares (each of side length 2^{-s}) which we refer to as *boxes*. See Fig. 5(a) for an example when $s = 3$. We call a pair of such boxes a *box pair*. The integral from (1) can then be split into a sum of integrals over all 2^{4s} box pairs. That is,

$$\langle \mathcal{X}, \mathcal{Y} \rangle_\Phi = \int_{\Delta^{(2)}} \Phi_{\mathcal{X}} \Phi_{\mathcal{Y}} d\mu = \sum_{(B_1, B_2)} \int_{\Delta^{(2)} \cap (B_1 \times B_2)} \Phi_{\mathcal{X}} \Phi_{\mathcal{Y}} d\mu.$$

For each box pair, we compute an approximation interval for the integral, and sum them up using interval arithmetic to obtain J_s.

We first give some (almost trivial) bounds for $\langle \mathcal{X}, \mathcal{Y} \rangle_\Phi$. Let (B_1, B_2) be a box pair with centers located at c_1 and c_2, respectively. By construction, $\mathrm{vol}(B_1 \times B_2) = 2^{-4s}$. By the absolute boundedness of ϕ, we have

$$\int_{\Delta^{(2)} \cap (B_1 \times B_2)} \Phi_{\mathcal{X}} \Phi_{\mathcal{Y}} d\mu \le \int_{(B_1 \times B_2)} \left(\frac{1}{\sqrt{2}} \nu_1 n \cdot \frac{1}{\sqrt{2}} \nu_1 n \right) d\mu \tag{7}$$

$$= \frac{\nu_1^2 n^2}{2} \mathrm{vol}(B_1 \times B_2) = \frac{\nu_1^2 n^2}{2^{4s+1}}, \tag{8}$$

where $1/\sqrt{2}$ is the maximal weight. Let $U := \frac{\nu_1^2 n^2}{2^{4s+1}}$. If $c_1 \le c_2$, then we can choose $[0, U]$ as approximation interval. Otherwise, if $c_1 \not\le c_2$, then $\Delta^{(2)} \cap (B_1 \times B_2) = \emptyset$; we simply choose [0,0] as approximation interval.

We can derive a second lower and upper bound for $\langle \mathcal{X}, \mathcal{Y} \rangle_\Phi$ as follows. We evaluate $\Phi_{\mathcal{X}}$ and $\Phi_{\mathcal{Y}}$ at the pair of centers (c_1, c_2), which is possible due to the efficiency hypothesis of ϕ. Let $\nu_{\mathcal{X}} = \Phi_{\mathcal{X}}(c_1, c_2)$ and $\nu_{\mathcal{Y}} = \Phi_{\mathcal{Y}}(c_1, c_2)$. Then, we compute *variations* $\delta_{\mathcal{X}}, \delta_{\mathcal{Y}} \ge 0$ relative to the box pair, with the property that, for any pair $(p, q) \in B_1 \times B_2$, $\Phi_{\mathcal{X}}(p, q) \in [\nu_{\mathcal{X}} - \delta_{\mathcal{X}}, \nu_{\mathcal{X}} + \delta_{\mathcal{X}}]$, and $\Phi_{\mathcal{Y}}(p, q) \in [\nu_{\mathcal{Y}} - \delta_{\mathcal{Y}}, \nu_{\mathcal{Y}} + \delta_{\mathcal{Y}}]$. In other words, variations describe how far the value of $\Phi_{\mathcal{X}}$ (or $\Phi_{\mathcal{Y}}$) deviates from its value at (c_1, c_2) within $B_1 \times B_2$. Combined with the derivations starting in (7), we have for any pair $(p, q) \in B_1 \times B_2$,

$$\max\{0, (\nu_{\mathcal{X}} - \delta_{\mathcal{X}})(\nu_{\mathcal{Y}} - \delta_{\mathcal{Y}})\} \tag{9}$$

$$\leq \Phi_{\mathcal{X}}(p,q)\Phi_{\mathcal{Y}}(p,q) \tag{10}$$

$$\leq \min\left\{\frac{v_1^2 n^2}{2}, (v_{\mathcal{X}}+\delta_{\mathcal{X}})(v_{\mathcal{Y}}+\delta_{\mathcal{Y}})\right\}. \tag{11}$$

By multiplying the bounds obtained in (9) by the volume of $\Delta^{(2)} \cap (B_1 \times B_2)$, we get a lower and an upper bound for the integral of $\Phi_{\mathcal{X}}\Phi_{\mathcal{Y}}$ over a box pair (B_1, B_2). By summing over all possible box pairs, the obtained lower and upper bounds are the endpoints of J_s.

Variations. We are left with computing the variations relative to a box pair. For simplicity, we set $\delta := \delta_{\mathcal{X}}$ and explain the procedure only for $\mathcal{X}$; the treatment of $\mathcal{Y}$ is similar.

We say that a slice ℓ *traverses* (B_1, B_2) if it intersects both boxes in at least one point. One such slice is the *center slice* ℓ_c, which is the slice through c_1 and c_2. See Fig. 5(b) for an illustration. We set D to be the maximal bottleneck distance of the center slice and every other slice traversing the box pair (to be more precise, of the persistence diagrams along the corresponding slices). We set W as the maximal difference between the weight of the center slice and any other slice traversing the box pair, where the weight w is defined as in Section 3. Write λ_{c_1} for the parameter value of c_1 along the center slice. For every slice ℓ traversing the box pair and any point $p \in \ell \cap B_1$, we have a value λ_p, yielding the parameter value of p along ℓ. We define L_1 as the maximal difference of λ_p and λ_{c_1} among all choices of p and ℓ. We define L_2 in the same way for B_2 and set $L := \max\{L_1, L_2\}$. With these notations, we obtain Lemma 6 below.

Lemma 6. *For all* $(p, q) \in B_1 \times B_2$,

$$|\Phi_{\mathcal{X}}(p,q) - \Phi_{\mathcal{X}}(c_1,c_2)| \leq \frac{v_3 n}{\sqrt{2}} D + v_1 nW + v_2 nL.$$

Proof. Plugging in (6) and using triangle inequality, we obtain

$$\begin{aligned}
&|\Phi_{\mathcal{X}}(p,q) - \Phi_{\mathcal{X}}(c_1,c_2)| \\
&= \left|\hat{\ell}\phi_{\mathcal{X}_\ell}(\lambda_p,\lambda_q) - \hat{\ell}_c\phi_{\mathcal{X}_{\ell_c}}(\lambda_{c_1},\lambda_{c_2})\right| \\
&\leq \hat{\ell}\left|\phi_{\mathcal{X}_\ell}(\lambda_p,\lambda_q) - \phi_{\mathcal{X}_{\ell_c}}(\lambda_p,\lambda_q)\right| + \phi_{\mathcal{X}_{\ell_c}}(\lambda_p,\lambda_q)\left|\hat{\ell} - \hat{\ell}_c\right| \\
&\quad + \hat{\ell}_c\left|\phi_{\mathcal{X}_{\ell_c}}(\lambda_p,\lambda_q) - \phi_{\mathcal{X}_{\ell_c}}(\lambda_{c_1},\lambda_{c_2})\right|
\end{aligned}$$

and bound the three parts separately. The first summand is upper bounded by $\frac{v_3 nD}{\sqrt{2}}$ because of internal stability of the feature map ϕ and because $\hat{\ell} \leq \frac{1}{\sqrt{2}}$ for any slice ℓ. The second summand is upper bounded by $v_1 nW$ by the absolute boundedness of ϕ. The third summand is bounded by $v_2 nL$, because $\|(\lambda_p,\lambda_q) - (\lambda_{c_1},\lambda_{c_2})\|_2 \leq \sqrt{2}\|(\lambda_p,\lambda_q) - (\lambda_{c_1},\lambda_{c_2})\|_\infty \leq L$ and by ϕ being Lipschitz, $\left|\phi_{\mathcal{X}_{\ell_c}}(\lambda_p,\lambda_q) - \phi_{\mathcal{X}_{\ell_c}}(\lambda_{c_1},\lambda_{c_2})\right| \leq \sqrt{2}v_2 nL$, and $\hat{\ell} \leq \frac{1}{\sqrt{2}}$. The result follows. □

Next, we bound D by simple geometric quantities. We use the following lemma, whose proof appeared in [48]:

Lemma 7. *[48] Let ℓ and ℓ' be two slices with parameterizations $b + \lambda a$ and $b' + \lambda a'$, respectively. Then, the bottleneck distance of the two persistence diagrams along these slices is upper bounded by*

$$\frac{2\|a - a'\|_\infty + \|b - b'\|_\infty}{\hat{\ell}\hat{\ell}'}.$$

We define A as the maximal infinity distance of the directional vector of the center slice ℓ_c and any other slice ℓ traversing the box pair. We define B as the maximal infinity distance of the base point of ℓ_c and any other ℓ. Finally, we set M as the minimal weight among all slices traversing the box pair. Using Lemma 7, we see that

$$D \leq \frac{2A + B}{M\hat{\ell}_c},$$

and we set

$$\delta := \frac{v_3 n(2A+B)}{\sqrt{2}M\hat{\ell}_c} + v_1 nW + v_2 nL. \tag{12}$$

It follows from Lemmas 6 and 7 that δ indeed satisfies the required variation property.

We remark that δ might well be equal to ∞, if the box pair admits a traversing slice that is horizontal or vertical, in which case the lower and upper bounds derived from the variation are vacuous. While (12) looks complicated, the values v_1, v_2, v_3 are constants coming from the considered feature map ϕ, and all the remaining values can be computed in constant time using elementary geometric properties of a box pair. We only explain the computation of A in Fig. 5(a) and skip the details of the other values.

Analysis. At this point, we have not made any claim that the algorithm is guaranteed to terminate. However, its correctness follows at once because J_s indeed contains the desired kernel value. Moreover, handling one box pair has a complexity that is polynomial in n, because the dominant step is to evaluate $\Phi_{\mathcal{X}}$ at the center (c_1, c_2). Hence, if the algorithm terminates at iteration s_0, its complexity is

$$\sum_{s=1}^{s_0} O\left(2^{4s}\, poly(n)\right).$$

This is because in iteration s, 2^{4s} box pairs need to be considered. Clearly, the geometric series above is dominated by the last iteration, so the complexity of the method is $O(2^{4s_0}\, poly(n))$. The last (and technically most challenging) step is to argue that $s_0 = O(\log n + \log \frac{1}{\epsilon})$, which implies that the algorithm indeed terminates and its complexity is polynomial in n and $1/\epsilon$.

To see that we can achieve any desired accuracy for the value of the kernel, i.e., that the interval width tends to 0, we observe that, if the two boxes B_1, B_2 are sufficiently far away and the resolution s is sufficiently large, the magnitudes A, B, W, and L in (12) are all small, because the parameterizations of two slices traversing the box pair are similar (see Lemmas 11–14 in Appendix A). Moreover, if every slice traversing the box pair has a sufficiently large weight (i.e., the slice is close to the diagonal), the value M in (12) is sufficiently large. These two properties combined imply that the variation of such a box pair (which we refer to as the *good* type) tends to 0 as s goes to ∞. Hence, the bound based on the variation tends to the correct value for good box pairs.

However, no matter how high the resolution, there are always *bad* box pairs for which either B_1, B_2 are close, or are far but close to horizontal and vertical, and hence yield a very large variation. For each of these box pairs, the bounds derived from the variation are vacuous, but we still have the trivial bounds $[0, U]$ based on the absolute boundedness of ϕ. Moreover, the total volume of these bad box pairs goes to 0 when s goes to ∞ (see Lemmas 9 and10 in Appendix A). So, the contribution of these box pairs tends to 0. These two properties complete the proof of Theorem 5.

A more careful investigation of our proof shows that the complexity of our algorithm is $O(n^{80+k}(1/\epsilon)^{40})$, where k is the efficiency constant of the feature map (Section 2). We made little effort to optimize the exponents in this bound.

6. Conclusions and future developments

We restate our main results for the case of a multi-filtration $\mathcal{X}$ with d parameters: there is a feature map that associates to $\mathcal{X}$

a real-valued function $\Phi_{\mathcal{X}}$ whose domain is of dimension $2d$, and introduces a kernel between a pair of multi-filtrations with a stable distance function, where the stability bounds depend on the ($2d$-dimensional) volume of a chosen bounding box. The proofs of these generalized results carry over from the results of this paper. Moreover, assuming that d is a constant, we claim that the kernel can be approximated in polynomial time to any constant (with the polynomial exponent depending on d). A proof of this statement requires to adapt the definitions and proofs of Appendix A to the higher-dimensional case; we omit details.

Other generalizations include replacing filtrations of simplicial complexes with persistence modules (with a suitable finiteness condition), passing to sublevel sets of a larger class of (tame) functions and replacing the scale-space feature map with a more general family of single-parameter feature maps. All these generalizations will be discussed in subsequent work.

The next step is an efficient implementation of our kernel approximation algorithm. We have implemented a prototype in C++, realizing a more adaptive version of the described algorithm. We have observed rather poor performance due to the sheer number of box pairs to be considered. Some improvements under consideration are to precompute all combinatorial persistence diagrams (cf. the barcode templates from [44]), to refine the search space adaptively using a quad-tree instead of doubling the resolution and to use techniques from numerical integration to handle real-world data sizes. We hope that an efficient implementation of our kernel will validate the assumption that including more than a single parameter will attach more information to the data set and improve the quality of machine learning algorithms using topological features.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was initiated at the Dagstuhl Seminar 17292 “Topology, Computation and Data Analysis”. We thank all members of the breakout session on multi-dimensional kernel for their valuable suggestions. RC, UF, and MK acknowledge support by the Austrian Science Fund (FWF) grant P 29984-N35. CL is partially supported by ARCES (University of Bologna), and FAR 2017 (University of Modena and Reggio Emilia). BW is partially supported by NSF grant DBI-1661375, IIS-1513616 and NIH grant R01-1R01EB022876-01. Fig. 3 is generated by Lin Yan.

Appendix A. Details on the Proof of Theorem 5

Overview. Recall that our approximation algorithm produces an approximation interval J_s for $s \in \mathbb{N}$ by splitting the unit square into $2^s \times 2^s$ boxes. For notational convenience, we write $u := 2^{-s}$ for the side length of these boxes.

We would like to argue that the algorithm terminates after $O(\log n + \log \frac{1}{\epsilon})$ iterations, which means that after that many iterations, an interval of width ϵ has been produced. The following Lemma 8 gives an equivalent criterion in terms of u and n.

Lemma 8. *Assume that there are constants e_1, $e_2 > 0$, such that* $\operatorname{width}(J_s) = O(n^{e_1} u^{e_2})$. *Then,* $\operatorname{width}(J_{s_0}) \le \epsilon$ *for some* $s_0 = O\left(\log n + \log \frac{1}{\epsilon}\right)$.

Proof. Assume that $\operatorname{width}(J_s) \le cn^{e_1} u^{e_2}$ for constants c and s sufficiently large. Since $u = 2^{-s}$, a simple calculation shows that $cn^{e_1} u^{e_2} \le \epsilon$ if and only if $s \ge \frac{\log c + e_1 \log n + \log \frac{1}{\epsilon}}{e_2}$. Hence, choosing

$$s := \left\lceil \frac{\log c + e_1 \log n + \log \frac{1}{\epsilon}}{e_2} \right\rceil = O\left(\log n + \log \frac{1}{\epsilon}\right)$$

ensures that $\operatorname{width}(J_{s_0}) \le \epsilon$. □

In the rest of this section, we will show that $\operatorname{width}(J_s) = O(n^2 u^{0.1})$.

Classifying box pairs. For the analysis, we partition the box pairs considered by the algorithm into 4 disjoint classes. We call a box pair (B_1, B_2):

- *null* if $c_1 \not\le c_2$,
- *close* if $c_1 \le c_2$ such that $\|c_1 - c_2\|_2 < \sqrt{u}$,
- *non-diagonal* if $c_1 \le c_2$ such that $\|c_1 - c_2\|_2 \ge \sqrt{u}$ and any line ℓ that traverses (B_1, B_2) satisfies $\hat{\ell} < u^{\frac{1}{5}}$,
- *good* if it is of neither of the previous three types.

According to this notation, the integral from (1) can then be split as

$$\langle \mathcal{X}, \mathcal{Y} \rangle_\Phi = \langle \mathcal{X}, \mathcal{Y} \rangle_{null} + \langle \mathcal{X}, \mathcal{Y} \rangle_{close} + \langle \mathcal{X}, \mathcal{Y} \rangle_{non-diag} + \langle \mathcal{X}, \mathcal{Y} \rangle_{good},$$

where, $\langle \mathcal{X}, \mathcal{Y} \rangle_{null}$ is defined as $\sum_{(B_1,B_2)\, null} \int_{\Delta^{(2)} \cap (B_1 \times B_2)} \Phi_{\mathcal{X}} \Phi_{\mathcal{Y}} d\mu$, and analogously for the other ones. We let $J_{s,null}$, $J_{s,close}$, $J_{s,non-diag}$, $J_{s,good}$ denote the four approximation intervals obtained from our algorithm when summing up the contributions of the corresponding box pairs. Then clearly, J_s is the sum of these four intervals. For simplicity, we will write J_{null} instead of $J_{s,null}$ when s is fixed, and likewise for the other three cases.

We observe first that the algorithm yields $J_{null} = [0, 0]$, so null box pairs can simply be ignored. Box pairs that are either close or non-diagonal are referred to as *bad* box pairs in Section 5. We proceed by showing that the width of J_{close}, $J_{non-diag}$, and J_{good} are all bounded by $O(n^2 u^{0.1})$.

Bad box pairs. We start with bounding the width of J_{close}. Let $\mathcal{B}_{\text{close}}$ be the union of all close box pairs. Note that our algorithm assigns to each box pair (B_1, B_2) an interval that is a subset of $[0, U]$. Recall that $U = \frac{\nu_1^2 n^2}{2^{4s+1}}$. U can be rewritten as $\frac{\nu_1^2 n^2}{2} \operatorname{vol}(B_1 \times B_2)$, where $\operatorname{vol}(B_1 \times B_2)$ is the 4-dimensional volume of the box pair (B_1, B_2). It follows that

$$\operatorname{width}(J_{close}) \le \frac{\nu_1^2 n^2}{2} \operatorname{vol}(\mathcal{B}_{\text{close}}). \tag{A.1}$$

Lemma 9. *For $u \le \frac{1}{2}$,* $\operatorname{vol}(\mathcal{B}_{\text{close}}) \le 4\pi u$.

Proof. Fixed a point $p \in R$, for each point $q \in R$ such that $(p, q) \in \mathcal{B}_{\text{close}}$ and $p < q$, there exists a unique close box pair (B_1, B_2) that contains (p, q). By definition of close box pair, we have that:

$$\|p - q\|_2 \le \|p - c_1\|_2 + \|c_1 - c_2\|_2 + \|c_2 - q\|_2 \le \sqrt{u} + \sqrt{2}u.$$

Moreover, for $u \le \frac{1}{2}$, $\sqrt{2}u \le \sqrt{u}$, and so $\|p - q\|_2 \le 2\sqrt{u}$. Equivalently, q belongs to the 2-ball $B(p, 2\sqrt{u})$ centered at p and of radius

$2\sqrt{u}$. Then,

$$\begin{aligned}\mathrm{vol}(\mathcal{B}_{\mathrm{close}}) &= \int_{\mathcal{B}_{\mathrm{close}}} 1 d\mu \leq \int_{p\in R}\int_{q\in B(p,2\sqrt{u})} 1 d\mu \\ &\leq \int_{p\in R} 4\pi u d\mu = 4\pi u.\end{aligned}$$

□

Consequently, combined with (A.1), we have

$$\mathrm{width}(J_{close}) \leq \frac{4\pi v_1^2 n^2}{2} u = O(n^2 u^{0.1}).$$

Note that $u < 1$ and hence, $u \leq u^{0.1}$.

For the width of $J_{non-diag}$, we use exactly the same reasoning, making use of the following Lemma 10. Let $\mathcal{B}_{\text{non-diag}}$ be the union of all non-diagonal box pairs.

Lemma 10. *For $u \leq 2^{-\frac{5}{2}}$, $\mathrm{vol}(\mathcal{B}_{\text{non-diag}}) \leq \sqrt{2}u^{\frac{1}{5}}$.*

Proof. Fixed a point $p \in R$, for each point $q \in R$ such that $(p, q) \in \mathcal{B}_{\text{non-diag}}$ and $p < q$, there exists a unique non-diagonal box pair (B_1, B_2) that contains (p, q). We have that q lies in:

- Triangle $T_1(p)$ of vertices $p = (p_1, p_2)$, $(1, p_2)$, and $(1, p_2 + (1 - p_1)\frac{a_2}{a_1})$, if the line ℓ of maximum slope passing through $B_1 \times B_2$ is such that $\hat{\ell} = a_2$ where $a = (a_1, a_2)$ is the (positive) unit direction vector of ℓ;
- Triangle $T_2(p)$ of vertices $p = (p_1, p_2)$, $(p_1, 1)$, and $(p_1 + (1 - p_2)\frac{a_1}{a_2}, 1)$, if the line ℓ of minimum slope passing through $B_1 \times B_2$ is such that $\hat{\ell} = a_1$ where $a = (a_1, a_2)$ is the (positive) unit direction vector of ℓ.

Let us bound the area of the two triangles. Since the calculations are analogous, let us focus on $T_1(p)$. By definition, the basis of $T_1(p)$ is smaller than 1 while its height is bounded by $\frac{a_2}{a_1}$. The maximum value for the height of $T_1(p)$ is achieved for $a_2 = u^{\frac{1}{5}}$. So, by exploiting the identity $a_1^2 + a_2^2 = 1$, we have

$$\left(\frac{a_2}{a_1}\right)^2 = \frac{u^{\frac{2}{5}}}{1 - u^{\frac{2}{5}}}.$$

Under the conditions $u \leq 2^{-\frac{5}{2}}$ and $\frac{1}{2}u^{-\frac{2}{5}} \geq 1$ we have

$$\frac{a_2}{a_1} \leq \sqrt{2}u^{\frac{1}{5}}.$$

Therefore, $\mathrm{area}(T_1(p)) \leq \frac{\sqrt{2}}{2}u^{\frac{1}{5}}$. Similarly, $\mathrm{area}(T_2(p)) \leq \frac{\sqrt{2}}{2}u^{\frac{1}{5}}$. Finally,

$$\begin{aligned}\mathrm{vol}(\mathcal{B}_{\text{non-diag}}) &= \int_{\mathcal{B}_{\text{non-diag}}} 1 d\mu \\ &\leq \int_{p\in R}\int_{q\in T_1(p)\cup T_2(p)} 1 d\mu \\ &\leq \int_{p\in R} \sqrt{2}u^{\frac{1}{5}} d\mu \leq \sqrt{2}u^{\frac{1}{5}}.\end{aligned}$$

□

Good box pairs. For good box pairs, we use the fact that the variation of a box pair yields a subinterval of $[(v_x - \delta_x)(v_y - \delta_y)\mathrm{vol}(B_1 \times B_2), (v_x + \delta_x)(v_y + \delta_y)\mathrm{vol}(B_1 \times B_2)]$ as an approximation, so the width is bounded by $2(v_x\delta_y + v_y\delta_x)\mathrm{vol}(B_1 \times B_2)$. Let $\mathcal{B}_{\text{good}}$ be the union of all good box pairs. Since the volumes of all good box pairs sum up to at most one, that is, $\mathrm{vol}(\mathcal{B}_{\text{good}}) \leq 1$, it follows that the width of J_{good} is bounded by $2(v_x\delta_y + v_y\delta_x)$. By absolute boundedness, v_x and v_y are in $O(n)$, and recall that by definition,

$$\delta_x = \frac{v_3 n(2A + B)}{\sqrt{2}M\hat{\ell}_c} + v_1 nW + v_2 nL = O\left(n\left(\frac{A + B}{M^2} + W + L\right)\right)$$

based on the fact that $\hat{\ell} \geq M$. The same bound holds for δ_y. Hence,

$$\mathrm{width}(J_{good}) = O\left(n^2\left(\frac{A + B}{M^2} + W + L\right)\right).$$

It remains to show that $\frac{A+B}{M^2} + W + L = O(u^{0.1})$. Note that $M \geq u^{\frac{1}{5}}$ because the box pair is assumed to be good. We will show in the next lemmas that A, B, W, and L are all in $O(\sqrt{u})$, proving that the term is indeed in $O(u^{0.1})$. This completes the proof of the complexity of the algorithm.

Lemma 11. *Let (B_1, B_2) be a good box pair. Let a, a' be the unit direction vectors of two lines that pass through the box pair. Then, $\|a - a'\|_\infty \leq 2\sqrt{u}$. In particular, $A = O(\sqrt{u})$.*

Proof. Since (B_1, B_2) is a good box pair, the largest value for $\|a - a'\|_\infty$ is achieved when ℓ and ℓ' correspond to the lines passing through the box pair(B_1, B_2) with minimum and maximum slope, respectively. By denoting as $c_1 = (c_{1,x}, c_{1,y})$, $c_2 = (c_{2,x}, c_{2,y})$ the centers of B_1, B_2, we define ℓ to be the line passing through the points $c_1 + (-\frac{u}{2}, \frac{u}{2})$, $c_2 + (\frac{u}{2}, -\frac{u}{2})$. Similarly, let us call ℓ' the line passing through the points $c_1 + (\frac{u}{2}, -\frac{u}{2})$, $c_2 + (-\frac{u}{2}, \frac{u}{2})$. So, the unit direction vector a of ℓ can be expressed as

$$a = \frac{(c_2 + (\frac{u}{2}, -\frac{u}{2})) - (c_1 + (-\frac{u}{2}, \frac{u}{2}))}{\left\|(c_2 + (\frac{u}{2}, -\frac{u}{2})) - (c_1 + (-\frac{u}{2}, \frac{u}{2}))\right\|_2}.$$

Similarly, the unit direction vector a' of ℓ' is described by

$$a' = \frac{(c_2 + (-\frac{u}{2}, \frac{u}{2})) - (c_1 + (\frac{u}{2}, -\frac{u}{2}))}{\left\|(c_2 + (-\frac{u}{2}, \frac{u}{2})) - (c_1 + (\frac{u}{2}, -\frac{u}{2}))\right\|_2}.$$

Then, by denoting as $\langle \cdot, \cdot \rangle$ the scalar product,

$$\begin{aligned}\|a - a'\|_\infty^2 \leq \|a - a'\|_2^2 &= \|a\|_2^2 + \|a'\|_2^2 - 2\langle a, a'\rangle = 2(1 - \langle a, a'\rangle) \\ &= 2\left(1 - \left\langle \frac{c_2 - c_1 + (u, -u)}{\|c_2 - c_1 + (u, -u)\|_2}, \frac{c_2 - c_1 + (-u, u)}{\|c_2 - c_1 + (-u, u)\|_2} \right\rangle\right) \\ &= 2\left(1 - \frac{\|c_2 - c_1\|_2^2 - 2u^2}{\|c_2 - c_1 + (u, -u)\|_2 \|c_2 - c_1 + (-u, u)\|_2}\right).\end{aligned}$$

By an elementary calculation, one can prove that

$$\begin{aligned}&\|c_2 - c_1 + (u, -u)\|_2 \|c_2 - c_1 + (-u, u)\|_2 \\ &= \sqrt{4u^2\left(u^2 + 2(c_{2,x} - c_{1,x})(c_{2,y} - c_{1,y})\right) + \|c_2 - c_1\|_2^4}.\end{aligned}$$

Then,

$$\begin{aligned}&\|a - a'\|_\infty^2 \\ &\leq 2\left(1 - \frac{\|c_2 - c_1\|_2^2 - 2u^2}{\sqrt{4u^2\left(u^2 + 2(c_{2,x} - c_{1,x})(c_{2,y} - c_{1,y})\right) + \|c_2 - c_1\|_2^4}}\right) \\ &= 2\left(1 + \frac{2u^2 - \|c_2 - c_1\|_2^2}{\sqrt{4u^2\left(u^2 + 2(c_{2,x} - c_{1,x})(c_{2,y} - c_{1,y})\right) + \|c_2 - c_1\|_2^4}}\right).\end{aligned}$$

Since (B_1, B_2) is a good box pair, $\|c_2 - c_1\|_2 \geq \sqrt{u}$. So,

$$\|a - a'\|_\infty^2 \leq 2\left(1 + \frac{2u - 1}{\sqrt{4\left(u^2 + 2(c_{2,x} - c_{1,x})(c_{2,y} - c_{1,y})\right) + 1}}\right).$$

Since $\sqrt{4\left(u^2 + 2(c_{2,x} - c_{1,x})(c_{2,y} - c_{1,y})\right) + 1} \geq 1$, we have that

$$\|a - a'\|_\infty^2 \leq 2(1 + 2u - 1) = 4u.$$

Therefore,

$$\|a - a'\|_\infty \leq 2\sqrt{u}.$$

□

Lemma 12. *Let (B_1, B_2) be a good box pair. Let $\ell = a\lambda + b$, $\ell' = a'\lambda + b'$ be two lines that pass through the box pair such that a, a' are unit direction vectors and b, b' are the intersection points with the diagonal of the second and the fourth quadrant. Then $\|b - b'\|_\infty \leq 4\sqrt{u}$. In particular, $B = O(\sqrt{u})$.*

Proof. Since (B_1, B_2) is a good box pair, the largest value for $\|b - b'\|_\infty$ is achieved when ℓ and ℓ' correspond to the lines passing through the box pair(B_1, B_2) with minimum and maximum slope, respectively. By denoting the centers of B_1 and B_2 by c_1 and c_2, we define ℓ to be the line passing through the points $c_1 + (-\frac{u}{2}, \frac{u}{2})$, $c_2 + (\frac{u}{2}, -\frac{u}{2})$. Similarly, let us call ℓ' the line passing through the points $c_1 + (\frac{u}{2}, -\frac{u}{2})$, $c_2 + (-\frac{u}{2}, \frac{u}{2})$. So, ℓ can be expressed as

$$(x, y) = \frac{c_2 + (\frac{u}{2}, -\frac{u}{2}) - c_1 - (-\frac{u}{2}, \frac{u}{2})}{\left\|c_2 + (\frac{u}{2}, -\frac{u}{2}) - c_1 - (-\frac{u}{2}, \frac{u}{2})\right\|_2} t + c_1 + \left(-\frac{u}{2}, \frac{u}{2}\right),$$

where t is a parameter running on $\mathbb{R}$. By intersecting ℓ with the line $y = -x$, we get:

$$\frac{c_{2,x} + \frac{u}{2} - c_{1,x} + \frac{u}{2}}{\left\|c_2 + (\frac{u}{2}, -\frac{u}{2}) - c_1 - (-\frac{u}{2}, \frac{u}{2})\right\|_2} t + c_{1,x} - \frac{u}{2}$$
$$= \frac{-c_{2,y} + \frac{u}{2} + c_{1,y} + \frac{u}{2}}{\left\|c_2 + (\frac{u}{2}, -\frac{u}{2}) - c_1 - (-\frac{u}{2}, \frac{u}{2})\right\|_2} t - c_{1,y} - \frac{u}{2},$$

which can be written as

$$c_{1,x} + c_{1,y} = \frac{c_{1,x} + c_{1,y} - c_{2,x} - c_{2,y}}{\left\|c_2 + (\frac{u}{2}, -\frac{u}{2}) - c_1 - (-\frac{u}{2}, \frac{u}{2})\right\|_2} t,$$

letting us deduce that

$$t = \frac{(c_{1,x} + c_{1,y})\left\|c_2 + (\frac{u}{2}, -\frac{u}{2}) - c_1 - (-\frac{u}{2}, \frac{u}{2})\right\|_2}{c_{1,x} + c_{1,y} - c_{2,x} - c_{2,y}}.$$

So, by replacing t in the equation of ℓ we retrieve b:

$$b = \frac{c_2 + (\frac{u}{2}, -\frac{u}{2}) - c_1 - (-\frac{u}{2}, \frac{u}{2})}{\left\|c_2 + (\frac{u}{2}, -\frac{u}{2}) - c_1 - (-\frac{u}{2}, \frac{u}{2})\right\|_2}$$
$$\frac{(c_{1,x} + c_{1,y})\left\|c_2 + (\frac{u}{2}, -\frac{u}{2}) - c_1 - (-\frac{u}{2}, \frac{u}{2})\right\|_2}{c_{1,x} + c_{1,y} - c_{2,x} - c_{2,y}} + c_1 + \left(-\frac{u}{2}, \frac{u}{2}\right)$$
$$= \frac{(u, -u)(c_{1,x} + c_{1,y})}{c_{1,x} + c_{1,y} - c_{2,x} - c_{2,y}} + \frac{(c_2 - c_1)(c_{1,x} + c_{1,y})}{c_{1,x} + c_{1,y} - c_{2,x} - c_{2,y}}$$
$$+ c_1 + \left(-\frac{u}{2}, \frac{u}{2}\right).$$

Similarly,

$$b' = \frac{(-u, u)(c_{1,x} + c_{1,y})}{c_{1,x} + c_{1,y} - c_{2,x} - c_{2,y}} + \frac{(c_2 - c_1)(c_{1,x} + c_{1,y})}{c_{1,x} + c_{1,y} - c_{2,x} - c_{2,y}}$$
$$+ c_1 + \left(\frac{u}{2}, -\frac{u}{2}\right).$$

So,

$$\|b - b'\|_\infty = \left\|\left(2\frac{c_{1,x} + c_{1,y}}{c_{1,x} + c_{1,y} - c_{2,x} - c_{2,y}} - 1\right)(u, -u)\right\|_\infty$$
$$= \left|\frac{c_{1,x} + c_{1,y} + c_{2,x} + c_{2,y}}{c_{2,x} + c_{2,y} - c_{1,x} - c_{1,y}}\right| \|(u, -u)\|_\infty$$
$$\leq \frac{4r}{|c_{2,x} + c_{2,y} - c_{1,x} - c_{1,y}|} u.$$

Since (B_1, B_2) is a good box pair,

$$c_{2,x} + c_{2,y} - c_{1,x} - c_{1,y} = \|c_2 - c_1\|_1 \geq \|c_2 - c_1\|_2 \geq \sqrt{u}.$$

Finally,

$$\|b - b'\|_\infty \leq \frac{4}{\sqrt{u}} u = 4\sqrt{u}.$$

□

Lemma 13. *Let (B_1, B_2) be a good box pair. Let $\hat{\ell}$, $\hat{\ell}'$ be the weights of two lines ℓ and ℓ' that pass through the box pair. Then $|\hat{\ell} - \hat{\ell}'| \leq 4\sqrt{u}$. In particular, $W = O(\sqrt{u})$.*

Proof. If $\hat{\ell} = a_1$ and $\hat{\ell}' = a'_1$, then, by applying Lemma 11,

$$|\hat{\ell} - \hat{\ell}'| = |a_1 - a'_1| \leq \|a - a'\|_\infty \leq 2\sqrt{u}.$$

On the other hand, if $\hat{\ell} = a_1$ and $\hat{\ell}' = a'_2$, then there exists a line ℓ'' passing through the box pair (B_1, B_2) such that $a'' = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$. By applying twice Lemma 11,

$$|\hat{\ell} - \hat{\ell}'| = |a_1 - a'_2| \leq |a_1 - \frac{\sqrt{2}}{2}| + |\frac{\sqrt{2}}{2} - a'_2|$$
$$= |a_1 - a''_1| + |a''_2 - a'_2| \leq \|a - a''\|_\infty + \|a'' - a'\|_\infty$$
$$\leq 4\sqrt{u}.$$

The cases $\hat{\ell} = a_2$, $\hat{\ell}' = a'_2$ and $\hat{\ell} = a_2$, $\hat{\ell}' = a'_1$ can be treated analogously to the previous ones. □

Lemma 14. *Let (p, q), (p', q') be two points in a good box pair (B_1, B_2) and let ℓ, ℓ' be the lines passing through p, q and p', q', respectively. In accordance with the usual parametrization, we have that $|\lambda_p - \lambda_{p'}| \leq \sqrt{2}u + 4\sqrt{u}$ and $|\lambda_q - \lambda_{q'}| \leq \sqrt{2}u + 4\sqrt{u}$. As a consequence, $L = O(\sqrt{u})$.*

Proof. Thanks to the definition of λ_p, the triangular inequality and Lemma 12, we have that:

$$\lambda_p = \|p - b\|_2 \leq \|p - p'\|_2 + \|p' - b'\|_2 + \|b' - b\|_2$$
$$\leq \sqrt{2}u + \lambda_{p'} + 4\sqrt{u}.$$

So, we have that $\lambda_p - \lambda_{p'} \leq \sqrt{2}u + 4\sqrt{u}$, and, similarly, $\lambda_{p'} - \lambda_p \leq \sqrt{2}u + 4\sqrt{u}$. Then,

$$|\lambda_p - \lambda_{p'}| \leq \sqrt{2}u + 4\sqrt{u}.$$

Analogously, it can be proven that

$$|\lambda_q - \lambda_{q'}| \leq \sqrt{2}u + 4\sqrt{u}.$$

□

References

[1] Carstens CJ, Horadam KJ. Persistent homology of collaboration networks. Math Probl Eng 2013;2013:7.

[2] Chan JM, Carlsson G, Rabadan R. Topology of viral evolution. Proc Natl Acad Sci USA 2013;110(46):18566–71.

[3] Giusti C, Pastalkova E, Curto C, Itskov V. Clique topology reveals intrinsic geometric structure in neural correlations. Proc Natl Acad Sci USA 2015;112(44):13455–60.

[4] Guo W, Banerjee AG. Toward automated prediction of manufacturing productivity based on feature selection using topological data analysis. In: Proceedings of the IEEE international symposium on assembly and manufacturing; 2016. p. 31–6.

[5] Hiraoka Y, Nakamura T, Hirata A, Escolar EG, Matsue K, Nishiura Y. Hierarchical structures of amorphous solids characterized by persistent homology. Proc Natl Acad Sci USA 2016;113(26):7035–40.

[6] Li L, Cheng WY, Glicksberg BS, Gottesman O, Tamler R, Chen R, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Sci Transl Med 2015;7(311):311ra174.

[7] Wang Y, Ombao H, Chung MK. Topological data analysis of single-trial electroencephalographic signals. Ann Appl Stat 2017;12(3):1506–34.

[8] Yoo J, Kim EY, Ahn YM, Ye JC. Topological persistence vineyard for dynamic functional brain connectivity during resting and gaming stages. J Neurosci Methods 2016;267(15):1–13.

[9] Nicolau M, Levine AJ, Carlsson G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. Proc Natl Acad Sci USA 2011;108(17):7265–70.

[10] Turner K, Mukherjee S, Boyer DM. Persistent homology transform for modeling shapes and surfaces. Inf Inference J IMA 2014;3(4):310–44.
[11] Li C, Ovsjanikov M, Chazal F. Persistence-based structural recognition. In: Proceedings of the ieee conference on computer vision and pattern recognition; 2014. p. 2003–10.
[12] Biasotti S, Falcidieno B, Spagnuolo M. Extended Reeb graphs for surface understanding and description. In: Proceedings of the ninth international conference on discrete geometry for computer imagery; 2000. p. 185–97.
[13] Skraba P, Ovsjanikov M, Chazal F, Guibas L. Persistence-based segmentation of deformable shapes. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition – Workshop on non-rigid shape analysis and deformable image alignment; 2010. p. 45–52.
[14] Munkres JR. Elements of algebraic topology. Redwood City, CA, USA: Addison-Wesley; 1984.
[15] Edelsbrunner H, Letscher D, Zomorodian AJ. Topological persistence and simplification. Discr Comput Geom 2002;28:511–33.
[16] Zomorodian A, Carlsson G. Computing persistent homology. Discr Comput Geom 2005;33(2):249–74.
[17] Edelsbrunner H, Harer J. Persistent homology – a survey. Contemp Math 2008;453:257–82.
[18] Edelsbrunner H, Harer J. Computational topology: An introduction. Providence, RI, USA: American Mathematical Society; 2010.
[19] Edelsbrunner H, Morozov D. Persistent homology: Theory and practice. In: Proceedings of the European congress of mathematics. European Mathematical Society Publishing House; 2012. p. 31–50.
[20] Cohen-Steiner D, Edelsbrunner H, Harer J. Stability of persistence diagrams. Discr Comput Geom 2007;37(1):103–20.
[21] Reininghaus J, Huber S, Bauer U, Kwitt R. A stable multi-scale kernel for topological machine learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 4741–8.
[22] Bubenik P. Statistical topological data analysis using persistence landscapes. J Mach Learn Res 2015;16(1):77–102.
[23] Kwitt R, Huber S, Niethammer M, Lin W, Bauer U. Statistical topological data analysis – a kernel perspective. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, editors. Proceedings of the advances in neural information processing systems 28. Curran Associates, Inc.; 2015. p. 3070–8.
[24] Kusano G, Fukumizu K, Hiraoka Y. Persistence weighted Gaussian kernel for topological data analysis 2016;48:2004–13.
[25] Hofer C, Kwitt R, Niethammer M, Uhl A. Deep learning with topological signatures. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, editors. Proceedings of the advances in neural information processing systems 30. Curran Associates, Inc.; 2017. p. 1634–44.
[26] Carlsson G, Zomorodian A. The theory of multidimensional persistence. Discr Comput Geom 2009;42(1):71–93.
[27] Carlsson G, Mémoli F. Multiparameter hierarchical clustering methods. In: Locarek-Junge H, Weihs C, editors. Classification as a tool for research. Springer Berlin Heidelberg; 2010. p. 63–70.
[28] Chazal F, Cohen-Steiner D, Guibas LJ, Mémoli F, Oudot SY. Gromov-Hausdorff stable signatures for shapes using persistence. In: Proceedings of the Eurographics symposium on geometry processing, 28; 2009. p. 1393–1403.
[29] Singh G, Mémoli F, Carlsson G. Topological methods for the analysis of high dimensional data sets and 3D object recognition. In: Proceedings of the Eurographics symposium on point-based graphics; 2007. p. 91–100.
[30] Sun J, Ovsjanikov M, Guibas L. A concise and provably informative multi-scale signature based on heat diffusion. Comput Graph Forum 2009;28(5):1383–92.
[31] Edelsbrunner H, Harer J. Jacobi sets of multiple Morse functions. In: Cucker F, DeVore R, Olver P, Süli E, editors. Proceedings of the foundations of computational mathematics. Cambridge University Press; 2002. p. 37–57. Minneapolis
[32] Edelsbrunner H, Harer J, Patel AK. Reeb spaces of piecewise linear mappings. In: Proceedings of the twenty-forth annual symposium on computational geometry; 2008. p. 242–50.
[33] Carriére M, Cuturi M, Oudot S. Sliced Wasserstein kernel for persistence diagrams. In: Proceedings of the thirty-forth international conference on machine learning, 70; 2017. p. 664–73.
[34] Adams H, Chepushtanova S, Emerson T, Hanson E, Kirby M, Motta F, et al. Persistence images: a stable vector representation of persistent homology. J Mach Learn Res 2017;18(1):218–52.
[35] Adcock A, Rubin D, Carlsson G. Classification of hepatic lesions using the matching metric. Comput Vis Image Understand 2014;121:36–42.
[36] Di Fabio B, Ferri M. Comparing persistence diagrams through complex vectors. In: M V, P E, editors. Proceedings of the international conference on image analysis and processing. Springer, Cham; 2015. p. 294–305.
[37] Adcock A, Carlsson E, Carlsson G. The ring of algebraic functions on persistence bar codes. Homol Homot Appl 2016;18:381–402.
[38] Kališnik S. Tropical coordinates on the space of persistence barcodes. Found Comput Math 2018;19(1):1–29.
[39] Lesnick M. The theory of the interleaving distance on multidimensional persistence modules. Found Comput Math 2015;15(3):613–50.
[40] Bjerkevik HB, Botnan MB, Kerber M. Computing the interleaving distance is NP-hard. arXiv:1811.09165; 2018.
[41] Dey TK, Xin C. Computing bottleneck distance for 2-d interval decomposable modules. In: Proceedings of the thirty-forth international symposium on computational geometry; 2018. p. 32:1–32:15.
[42] Biasotti S, Cerri A, Frosini P, Giorgi D. A new algorithm for computing the 2-dimensional matching distance between size functions. Pattern Recogn Lett 2011;32(14):1735–46.
[43] Kerber M, Lesnick M, Oudot S. Exact computation of the matching distance on 2-parameter persistence modules. In: Proceedings of the thirty-fifth international symposium on computational geometry; 2019. p. 46:1–46:15.
[44] Lesnick M, Wright M. Interactive visualization of 2-D persistence modules. arXiv:1512.00180; 2015.
[45] Oudot S. Persistence theory: From quiver representation to data analysis. Mathematical surveys and monographs, 209. American Mathematical Society; 2015.
[46] Cohen-Steiner D, Edelsbrunner H, Harer J. Stability of persistence diagrams. Discr Comput Geom 2007;37(1):103–20.
[47] Biasotti S, Cerri A, Frosini P, Giorgi D, Landi C. Multidimensional size functions for shape comparison. J Math Imaging Vis 2008;32(2):161.
[48] Landi C. The rank invariant stability via interleavings. In: Chambers EW, Fasy BT, Ziegelmeier L, editors. Research in computational topology, 13. Springer; 2018. p. 1–10.