Journal of Machine Learning Research 23 (2022) 1-33 Submitted 9/21; Revised 7/22; Published 8/22

A Computationally Efficient Framework for Vector Representation of Persistence Diagrams

Kit C Chan KCHAN@BGSU.EDU
Department of Mathematics and Statistics, Bowling Green State University
Bowling Green, OH 43403, USA

Umar Islambekov IUMAR@BGSU.EDU
Department of Mathematics and Statistics, Bowling Green State University
Bowling Green, OH 43403, USA

Alexey Luchinsky ALUCHI@BGSU.EDU
Department of Mathematics and Statistics, Bowling Green State University
Bowling Green, OH 43403, USA

Rebecca Sanders REBECCA.SANDERS@MARQUETTE.EDU
Department of Mathematical and Statistical Sciences, Marquette University
Milwaukee, WI 53233, USA

Editor: Sayan Mukherjee

Abstract

In Topological Data Analysis, a common way of quantifying the shape of data is to use a *persistence diagram* (PD). PDs are multisets of points in $\mathbb{R}^2$ computed using tools of algebraic topology. However, this multi-set structure limits the utility of PDs in applications. Therefore, in recent years efforts have been directed towards extracting informative and efficient summaries from PDs to broaden the scope of their use for machine learning tasks. We propose a computationally efficient framework to convert a PD into a vector in $\mathbb{R}^n$, called a *vectorized persistence block* (VPB). We show that our representation possesses many of the desired properties of vector-based summaries such as stability with respect to input noise, low computational cost and flexibility. Through simulation studies, we demonstrate the effectiveness of VPBs in terms of performance and computational cost for various learning tasks, namely clustering, classification and change point detection.

Keywords: Topological data analysis, Persistence diagram, Persistence block

1. Introduction

In recent years, topological data analysis (TDA) has become a popular approach to study data that have inherent shape structure [see Edelsbrunner and Harer (2010); Carlsson (2009)]. One of the primary tools of TDA is *persistent homology* (PH) [see Zomorodian and Carlsson (2005); Edelsbrunner et al. (2008)]. The idea behind PH is to construct a sequence of nested topological spaces, typically *simplicial complexes* (indexed by some scale parameter), on top of data points and keep a record of various topological features (such as connected components, one dimensional holes, two dimensional holes, etc.) that emerge and later disappear as the scale parameter increases. A simplicial complex acts as a bridge between discrete data points and the underlying object they are sampled from, and its combinatorial nature makes it

suitable for computational purposes. The topological features are detected using the concept of homology from algebraic topology (see e.g. Rotman (2013); Kaczynski et al. (2006) for a detailed treatment of the topic). The output of applying PH is a *persistence diagram* (PD) which is a multiset of points in $\mathbb{R}^2$ where each point (b, d) represents a topological feature that is "born" at scale value b and "dies" at scale value d. A PD can be viewed as a multiscale topological signature of the data from which it is computed.

In applications, PDs are often used within statistical and machine learning frameworks for further analysis of the data. A direct way in this regard is to endow the space of PDs with the bottleneck or Wasserstein metrics [see Mileyko et al. (2011); Cohen-Steiner et al. (2007)] and use them in distance-based learning settings such as k-medoids or k-nearest neighbors (kNN). Though stability results exist for PDs with respect to these metrics [see Cohen-Steiner et al. (2007); Chazal et al. (2014); Cohen-Steiner et al. (2010)], this approach has two limitations: the scope of applicability and the computational cost. Firstly, it is not suitable for machine learning methods requiring input features to be elements of a Hilbert space (the space of PDs can not be turned into a Hilbert space [see Bubenik and Vergili (2018); Mileyko et al. (2011)]). Secondly, as the number of points in PDs grows, so does the cost of computing distances between them. Therefore, in order to facilitate the use of PDs for a broader class of machine learning methods, much research in TDA centers around the problem of extracting useful and computationally efficient summaries from PDs that are identified as vectors of a Hilbert space. There are two common ways to achieve this. The first method to map PDs into a Hilbert space is by defining a kernel function which encodes similarity between PDs and using it in kernel-based machine learning settings [see Chen et al. (2015); Kusano et al. (2016); Li et al. (2014); Reininghaus et al. (2015)]. The second approach is to extract a vector from a PD which is an element of $\mathbb{R}^n$ [see Bubenik (2015); Rieck et al. (2017); Adams et al. (2017); Berry et al. (2020); Atienza et al. (2020); Richardson and Werman (2014); Islambekov et al. (2020); Chung and Lawson (2019)]. According to Adams et al. (2017), the desired properties of vectorization-based summaries are stability with respect to input noise, computational efficiency, interpretable connection to the PD and flexibility that allows different treatments of points of the PD based on their relative importance.

In the present paper, we follow the approach introduced in Adams et al. (2017) with modifications that allow us to have a computationally efficient framework to extract vectorizaton-based summaries that have all of the above characteristics. First, we switch from the birth-death to birth-persistence coordinates using the map $(b, d) \mapsto (b, p)$, where $p = d - b$. Then, instead of using the 2-dimensional Gaussian distribution centered at (b, d) as is Adams et al. (2017), we use a characteristic function of a square centered at (b, p). In other words, for a given persistence diagram D in the new coordinates, we construct a *persistence block* (PB) which is a function $\widetilde{f}(D) : \mathbb{R}^2 \to \mathbb{R}$ defined as a linear combination of characteristic functions of squares centered at the points of D (see Figure 1):

$$\widetilde{f}(D) = \sum_{u=(b,p)\in D} f(u)\chi_{E(u)} \tag{1}$$

where $f : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ and $E(u)$ is a square centered at $u = (b, p) \in D$ with width depending on the persistence value p and the scale-free parameter $\tau \in [0, 1]$. The value of τ can be set in advance or optimally tuned within the machine learning method in question (see Section

4 for more details). Finally, we superimpose a rectangular grid $\mathcal{P}$ on the domain of the PB $\tilde{f}(D)$ and compute the integral of $\tilde{f}(D)$ with respect to some chosen weight function w over each grid cell S:

$$I_{f,S}(D) = \int_S \widetilde{f}(D)\, w dA,$$

By arranging these integral values into a vector in $\mathbb{R}^n$, where n in the number of grid cells, we obtain a *vectorized persistence block* (VPB):

$$\widetilde{f}_{\mathcal{P}}(D) = (I_{f,S_1}(D), I_{f,S_2}(D), \ldots, I_{f,S_n}(D))$$

The idea behind a VPB is motivated by the work on *persistence images* (PI) in Adams et al. (2017). In Section 2, we discuss in more detail the similarities and differences between the two approaches.

The rest of the paper is organized as follows. In Section 3, we formally introduce PBs and VPBs and their continuity and stability results with respect to the Wasserstein distance. The results of the simulation studies are presented in Section 4. Finally, in Section 5, we summarize the contributions of our work and highlight potential directions for future research. Our `R` code to compute VPBs can be found at `https://github.com/uislambekov/TDA-VPB.git`.

2. Related Work

There are several vectorization-based summaries of PDs found in the TDA literature such as persistence landscapes Bubenik (2015), persistence images Adams et al. (2017), persistence curves Chung and Lawson (2019), persistence indicator functions Rieck et al. (2017) and persistent entropy Atienza et al. (2020), to name a few. Of these different summaries, the work on persistence images (PI) in Adams et al. (2017) serves as a motivation for our method. In this section we primarily focus on presenting a detailed discussion about the similarities and differences of the two approaches.

A PI is obtained by integrating a persistence surface over each cell of a superimposed grid. For a PD D, a persistence surface $\rho_D : \mathbb{R}^2 \to \mathbb{R}$ is a weighted sum of differentiable probability distributions $\phi_u : \mathbb{R}^2 \to \mathbb{R}$ with mean $u \in D$ (see Figure 2):

$$\rho_D = \sum_{u=(b,p)\in D} f(u)\phi_u. \tag{2}$$

In practice, ϕ_u is usually taken as a bivariate Gaussian distribution with mean u and covariance matrix $\sigma^2 I_{2\times 2}$. In this case, the parameter σ^2 of the Gaussians is the same for all points in the PD and is selected by a user. The weights $f(u)$, $u \in D$, encode a relative importance of the points in D. In the simulation studies of Adams et al. (2017), the weights are computed from a piecewise linear function depending only on the persistence value and all the grid cells are constructed to be of equal size. If the birth values are all zero in D, as it often happens for the homological dimension H_0, then one uses univariate Gaussian distributions instead.

Despite the similarity in construction, a PB has several key differences from a PI. Firstly, for the purposes of our simulation studies, we construct a PB as an unweighted sum of

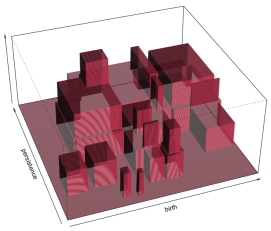

Figure 1: A sample persistence block constructed on top of a simulated PD with 15 points.

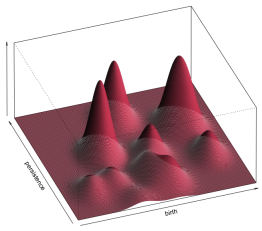

Figure 2: A sample persistence surface constructed on top of a simulated PD with 15 points.

characteristic functions of the squares centered at the points of a PD (i.e., we set $f(u) = 1$ for all $u \in D$ in (1); see also Section 3.2 for more details). Secondly, the distribution functions ϕ_u in (2) are assumed to be differentiable, whereas $\chi_{E(u)}$ in (1) are non-differentiable. This difference is crucial since the stability results for PIs critically rely on ϕ_u being differentiable. Thirdly, the sizes of the squares $\chi_{E(u)}$ are not the same, but rather change according to their center locations and the parameter τ. In contrast, the dispersion of ϕ_u remains the same for all points of a PD. When ϕ_u is chosen as a bivariate Guassian distribution, the τ parameter mimics the role of σ^2 which controls the spread of the Gaussians. However, while σ^2 is a scale-dependent parameter, τ is a scale-free parameter of proportion with values in $[0, 1]$ which makes it relatively easier to select. Moreover, a gain in computational cost with regard to PBs (see Section 4 for more details) facilitates the tuning of τ to be more affordable in applications via a parameter grid search. Fourthly, unlike with a PI, a weight function $w : \mathbb{R}^2 \to \mathbb{R}$ is introduced in the vectorization step of a PB via integration (see Section 3.3 for more details). Due to simplicity of construction, one can derive a closed-form formula to compute the corresponding integrals for a large class of weight functions which enables a fast computation of VPBs in practice. Finally, as opposed to the grid cells in Adams et al. (2017) which have the same dimensions, in our simulations, they are of different sizes and selected in a data-driven way based on sample percentiles of the birth and persistence values.

3. Theory

In this section, we provide a mathematical background for our analysis of persistence diagrams. We set forth a foundation and establish continuity and stability results for representing a persistence diagram as a function in an infinite dimensional Hilbert space L^2 and as a vector in $\mathbb{R}^n$.

3.1 Persistence Block

An important aspect of topological data analysis is its ability to track various topological features appearing and disappearing within a filtration built upon data. One approach to code this information is to record the birth b which is the time a topological feature appears, and the death d which is time the exact same feature disappears. In this way, every

topological feature corresponds to exactly one point (b, d) in $\mathbb{R}^2$. Since a topological feature disappears after it appears, we have $0 \leq b \leq d$, and consequently the point (b, d) lies on or above the line $y = x$. An alternative approach is to record the birth b of a topological feature and its persistence p, or in other words the life span (i.e. $p = d - b$), of the same feature. Again, each topological feature corresponds to exactly one point (b, p) in $\mathbb{R}^2$. Since the persistence p can be any nonnegative number and if necessary by shifting the birth coordinate values by an appropriate constant, we can assume that the point (b, p) lies in the first quadrant of $\mathbb{R}^2$. For the framework of the present paper, we take the second approach and assume that a persistence diagram (PD) is a collection of points in the first quadrant of $\mathbb{R}^2$, with the interpretation that a point (b, p) in a PD represents the birth and persistence of some topological feature.

To facilitate the computations following the ideas in Adams et al. (2017), we need to identify a point in a PD with a function in some Hilbert space L^2. In Adams et al. (2017), a point (b, d) in a PD is identified with a function given by a 2-dimensional Gaussian distribution centered at (b, d). In our present paper, we identify a point (b, p) in a PD with a characteristic function of a square centered at (b, p). In Section 4, we show this identification leads to a computationally efficient framework.

Keeping in mind that in practice we only deal with a finite number of PDs and topological features, we can assume all persistence diagrams are finite and there are positive real numbers α, β such that all persistence diagrams are subsets of the compact rectangle $\Omega = [0, \alpha] \times [0, \beta]$. The set of all finite persistence diagrams is denoted by $P(\Omega)$ and the set of all persistence diagrams with at most N points is denoted by $P_N(\Omega)$. Technically speaking we can assume all persistence diagrams are in $P_N(\Omega)$ for some integer N. However, in a careful treatment of our theory dealing with persistence diagrams with different numbers of points, we consider $P_N(\Omega)$ as a subset of $P(\Omega)$ and $P(\Omega) = \bigcup_N P_N(\Omega)$.

To identify a point in a persistence diagram D in $P(\Omega)$ with a function in some Hilbert space L^2, we define a *length function* $\lambda : \Omega \to [0, \infty)$, which is a continuous function such that

$$0 < \lambda(x, y) < 2y \quad \text{whenever } (x, y) \in \Omega \text{ with } y > 0, \text{ and} \tag{3}$$

$$\lambda(x, 0) = 0 \quad \text{whenever } (x, 0) \in \Omega. \tag{4}$$

For any point $(x, y) \in \Omega$, let $E(x, y)$ be the square centered at (x, y) with side length $\lambda(x, y)$. That is,

$$E(x, y) = \left[x - \frac{\lambda}{2}, x + \frac{\lambda}{2}\right] \times \left[y - \frac{\lambda}{2}, y + \frac{\lambda}{2}\right], \text{ where } \lambda = \lambda(x, y). \tag{5}$$

Note part of the square $E(x, y)$ may be in the second quadrant, although that part of the square may or may not be included in our integration in Lemma 2 depending on how we apply our theory in the applications. Regardless, we first study the properties of the whole square $E(x, y)$ in the following lemma, which explains the relationship between the area of overlapping squares and the length function λ.

Lemma 1 *For any two points $(x, y), (a, b) \in \Omega$, let $E = E(x, y)$ and $E_0 = E(a, b)$ be two squares with side lengths $\lambda = \lambda(x, y)$ and $\lambda_0 = \lambda(a, b)$, centered at (x, y) and (a, b) respectively. We have the following statements:*

(1) $\text{Area}(E \cap E_0) > 0$ *if and only if* $\|(x,y)-(a,b)\|_\infty < \frac{\lambda+\lambda_0}{2}$;

(2) $\text{Area}(E \cup E_0) = \lambda^2 + \lambda_0^2$ *whenever* $\|(x,y)-(a,b)\|_\infty \geq \frac{\lambda+\lambda_0}{2}$;

(3) If $\text{Area}(E \cap E_0) > 0$, *then*

$$\text{Area}(E \cap E_0) = \left(\frac{\lambda+\lambda_0}{2} - |x-a|\right)\left(\frac{\lambda+\lambda_0}{2} - |y-b|\right) \leq \left(\frac{\lambda+\lambda_0}{2}\right)^2;$$

(4) If $\text{Area}(E \cap E_0) > 0$, *then*

$$\text{Area}(E \setminus E_0) + \text{Area}(E_0 \setminus E) \leq \frac{|\lambda-\lambda_0|^2}{2} + 2(\lambda+\lambda_0)\|(x,y)-(a,b)\|_\infty.$$

Proof The proofs of Statements (1), (2) and (3) follow directly from visualizing two overlapping and two non-overlapping squares. To establish Statement (4), observe that whenever $\text{Area}(E \cap E_0) > 0$, we have

$$\begin{aligned}
\text{Area}(E \setminus E_0) &= \text{Area}(E) - \text{Area}(E \cap E_0) \\
&= \lambda^2 - \left(\frac{\lambda+\lambda_0}{2} - |x-a|\right)\left(\frac{\lambda+\lambda_0}{2} - |y-b|\right) \\
&\leq \lambda^2 - \left(\frac{\lambda+\lambda_0}{2}\right)^2 + \frac{\lambda+\lambda_0}{2}(|x-a| + |y-b|) \\
&\leq \lambda^2 - \left(\frac{\lambda+\lambda_0}{2}\right)^2 + (\lambda+\lambda_0)\|(x,y)-(a,b)\|_\infty.
\end{aligned}$$

By interchanging the roles of E and E_0 in the above inequality, we have

$$\text{Area}(E_0 \setminus E) \leq \lambda_0^2 - \left(\frac{\lambda+\lambda_0}{2}\right)^2 + (\lambda+\lambda_0)\|(x,y)-(a,b)\|_\infty.$$

Therefore,

$$\begin{aligned}
\text{Area}(E \setminus E_0) + \text{Area}(E_0 \setminus E) &\leq \lambda^2 + \lambda_0^2 - 2\left(\frac{\lambda+\lambda_0}{2}\right)^2 + 2(\lambda+\lambda_0)\|(x,y)-(a,b)\|_\infty \\
&= \frac{(\lambda-\lambda_0)^2}{2} + 2(\lambda+\lambda_0)\|(x,y)-(a,b)\|_\infty,
\end{aligned}$$

which concludes the proof of Lemma 1. ■

Based on the definition of the length function $\lambda(x,y)$, each square $E(x,y)$ is contained in $[-\beta, \alpha+\beta] \times [0, 2\beta]$, and hence the union $\bigcup_{(x,y)\in\Omega} E(x,y) \subseteq [-\beta, \alpha+\beta] \times [0, 2\beta]$. By letting $\Omega' = [-\beta, \alpha+\beta] \times [0, 2\beta]$ and letting $w : \Omega' \longrightarrow [0, \infty)$ be a continuous nonnegative *weight function* such that

$$w(x,y) > 0, \text{ whenever } y > 0, \tag{6}$$

we consider the weighted area measure wdA on Ω' and its corresponding Hilbert space $L^2(wdA)$. Any function $f : \Omega \longrightarrow \mathbb{R}$ induces a map $\widetilde{f} : P(\Omega) \longrightarrow L^2(wdA)$ by

$$\widetilde{f}(D) = \sum_{(a,b)\in D} f(a,b)\chi_{E(a,b)} \tag{7}$$

where $\chi_{E(a,b)}$ is the characteristic function of the square $E(a,b)$ centered at (a,b) as defined in (5). Since $P(\Omega) = \bigcup_{N=1}^{\infty} P_N(\Omega)$, the induced map $\widetilde{f}$ can be restricted to $P_N(\Omega)$ for any integer $N \geq 1$. The vector $\widetilde{f}(D)$ in the Hilbert space $L^2(wdA)$ is called a *persistence block*, as we 3-dimensionally visualize a linear combination of $\chi_{E(a,b)}$ in (7) as a block. The map $\widetilde{f}$ is called the *persistence block transformation induced by* f, or simply the *induced persistence block map* if the function f is understood. Taking $\widetilde{f}(D)$ as a function in $L^2(wdA)$, we generally integrate the persistence block $\widetilde{f}(D)$ over the rectangle Ω'. However, there are instances within the theory or within the applications where one may choose to integrate the persistence block $\widetilde{f}(D)$ over the smaller rectangle Ω. In these instances, we only require the nonnegative weight function w to be defined on the rectangle Ω with w continuously extended to Ω' as necessary.

In Theorem 3 we establish a necessary and sufficient condition for an induced persistence block map $\widetilde{f}$ to be continuous, in the case when f is bounded. For that we need the following lemma, noting that Ω and Ω' are compact and $\lambda : \Omega \longrightarrow [0,\infty)$ and $w : \Omega' \longrightarrow [0,\infty)$ are continuous, and hence the functions λ and w are necessarily bounded.

Lemma 2 *Let $\Omega = [0,\alpha] \times [0,\beta]$ and $\Omega' = [-\beta, \alpha+\beta] \times [0, 2\beta]$ with $\alpha, \beta > 0$. Suppose $f : \Omega \longrightarrow \mathbb{R}$ is bounded, and M, C are a positive scalars such that*

$$\|f\|_\infty \leq M, \ \|\lambda\|_\infty \leq M, \ \text{and } \|w\|_\infty = C. \tag{8}$$

Then we have the following estimations for any two points $(x,y),(a,b)$ in Ω.

(1) If $\text{Area}(E(x,y) \cap E(a,b)) > 0$, *then*

$$\begin{aligned}
&\int_\Omega |f(x,y)\chi_{E(x,y)} - f(a,b)\chi_{E(a,b)}|^2 wdA \\
\leq\ &\int_{\Omega'} |f(x,y)\chi_{E(x,y)} - f(a,b)\chi_{E(a,b)}|^2 wdA \\
\leq\ &4CM^3\Big(|f(x,y) - f(a,b)| + |\lambda(x,y) - \lambda(a,b)| + \|(x,y) - (a,b)\|_\infty\Big).
\end{aligned}$$

(2) If $\text{Area}(E(x,y) \cap E(a,b)) = 0$, *then*

$$\begin{aligned}
\int_\Omega |f(x,y)\chi_{E(x,y)} - f(a,b)\chi_{E(a,b)}|^2 wdA &\leq \int_{\Omega'} |f(x,y)\chi_{E(x,y)} - f(a,b)\chi_{E(a,b)}|^2 wdA \\
&\leq 2CM^3\|(x,y) - (a,b)\|_\infty.
\end{aligned}$$

Proof For convenience, let $E = E(x,y)$, $\lambda = \lambda(x,y)$ and $E_0 = E(a,b)$, $\lambda_0 = \lambda(a,b)$. To establish the first inequalities in Statements (1) and (2), note that $\Omega \subseteq \Omega'$ and the integrands of both integrals over Ω are nonnegative.

Focusing on the integrals over Ω' in both statements (1) and (2), we observe that

$$\begin{aligned} |f(x,y)-f(a,b)|^2 &\le (|f(x,y)|+|f(a,b)|)\,|f(x,y)-f(a,b)| \\ &\le 2M\,|f(x,y)-f(a,b)|\,. \end{aligned} \tag{9}$$

By repeated applications of the inequalities in (8), we have

$$\begin{aligned} &\int_{\Omega'} |f(x,y)\chi_E - f(a,b)\chi_{E_0}|^2\, w dA \qquad (10)\\ &= \int_{E\cap E_0} |f(x,y)-f(a,b)|^2\, w dA + \int_{E\setminus E_0} |f(x,y)|^2\, w dA + \int_{E_0\setminus E} |f(a,b)|^2\, w dA \\ &\le |f(x,y)-f(a,b)|^2 \int_{E\cap E_0} w dA + M^2 \int_{E\setminus E_0} w dA + M^2 \int_{E_0\setminus E} w dA \\ &\le C\,|f(x,y)-f(a,b)|^2\,\mathrm{Area}(E\cap E_0) + CM^2\mathrm{Area}(E\setminus E_0) + CM^2\mathrm{Area}(E_0\setminus E) \\ &\le 2CM\,|f(x,y)-f(a,b)|\,\mathrm{Area}(E\cap E_0) + CM^2\Big(\mathrm{Area}(E\setminus E_0)+\mathrm{Area}(E_0\setminus E)\Big),\ \text{by (9)}. \end{aligned}$$

If $\mathrm{Area}(E\cap E_0)>0$, by Statement (3) of Lemma 1, we have

$$\mathrm{Area}(E\cap E_0) \le \left(\frac{\lambda+\lambda_0}{2}\right)^2 \le M^2,\ \text{by (8)}. \tag{11}$$

Similarly, by Statement (4) of Lemma 1, we have

$$\begin{aligned} \mathrm{Area}(E\setminus E_0)+\mathrm{Area}(E_0\setminus E) &\le \frac{|\lambda-\lambda_0|^2}{2} + 2(\lambda+\lambda_0)\|(x,y)-(a,b)\|_\infty \qquad (12)\\ &\le \frac{(\lambda+\lambda_0)|\lambda-\lambda_0|}{2} + 2(\lambda+\lambda_0)\|(x,y)-(a,b)\|_\infty \\ &\le M|\lambda-\lambda_0| + 4M\|(x,y)-(a,b)\|_\infty,\ \text{by (8)} \\ &= 4M\Big(|\lambda-\lambda_0| + \|(x,y)-(a,b)\|_\infty\Big). \end{aligned}$$

Substituting (11) and (12) into (10) yields

$$\begin{aligned} &\int_{\Omega'} |f(x,y)\chi_E - f(a,b)\chi_{E_0}|^2\, w dA \qquad (13)\\ &\le 2CM^3\,|f(x,y)-f(a,b)| + 4CM^3\Big(|\lambda-\lambda_0| + \|(x,y)-(a,b)\|_\infty\Big) \\ &\le 4CM^3\Big(\,|f(x,y)-f(a,b)| + |\lambda-\lambda_0| + \|(x,y)-(a,b)\|_\infty\Big). \end{aligned}$$

If $\mathrm{Area}(E\cap E_0)=0$, then

$$\begin{aligned} \mathrm{Area}(E\setminus E_0)+\mathrm{Area}(E_0\setminus E) &= \mathrm{Area}(E)+\mathrm{Area}(E_0) \qquad (14)\\ &= \lambda^2+\lambda_0^2 \\ &\le M\lambda + M\lambda_0,\ \text{by (8)} \\ &= M(\lambda+\lambda_0) \\ &\le 2M\|(x,y)-(a,b)\|_\infty,\ \text{by Statement (2) of Lemma 1.} \end{aligned}$$

Combining the above inequality with (10), we get

$$\begin{aligned}&\int_{\Omega'} |f(x,y)\chi_E - f(a,b)\chi_{E_0}|^2 \, w dA \\ &\leq CM^2(\text{Area}(E \setminus E_0) + \text{Area}(E_0 \setminus E)) \\ &\leq 2CM^3 \|(x,y) - (a,b)\|_\infty,\end{aligned}$$

which concludes the proof of the lemma. ■

In order to establish a necessary and sufficient condition to ensure the continuity of an induced persistence block map, we utilize the p-Wasserstein distance for the metric on the space of persistence diagrams. For $1 \leq p < \infty$, the *p-Wasserstein distance* $W_p(D, D_0)$ between two persistence diagrams D, D_0 in $P_N(\Omega)$ is defined by

$$W_p(D, D_0) = \inf \left\{ \left(\sum_{(a,b) \in D} \|\gamma(a,b) - (a,b)\|_\infty^p \right)^{1/p} \middle| \; \gamma : D \longrightarrow D_0 \text{ is a bijection} \right\}. \quad (15)$$

where additional points along the x-axis may be added to D or D_0 in order to define a bijection γ. For $0 < p < 1$, we similarly define the *p-Wasserstein distance* $W_p(D, D_0)$ by

$$W_p(D, D_0) = \inf \left\{ \sum_{(a,b) \in D} \|\gamma(a,b) - (a,b)\|_\infty^p \middle| \; \gamma : D \longrightarrow D_0 \text{ is a bijection} \right\}, \quad (16)$$

To explain the difference between the two formulas in (15) and (16), let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ be two vectors in $\mathbb{R}^n$. The formula $d_p(x,y) = \left(\sum_{j=1}^n |x_j - y_j|^p \right)^{1/p}$ defines a metric on $\mathbb{R}^n$ only when $1 \leq p < \infty$. However, in the case when $0 < p < 1$, the above formula fails to define a metric. On the positive side, by modifying the formula to $d_p(x,y) = \sum_{j=1}^n |x_j - y_j|^p$, we generate a metric on $\mathbb{R}^n$ (see (Rudin, 1991), Section 1.47, page 36). Thus in both (15) and (16), we take the infimum of a metric.

Before we state the main theorem of the current subsection, we remark that the weight function w is defined on Ω', and hence the norm of the Hilbert space $L^2(wdA)$ is given by

$$\|f\|_2 = \left(\int_{\Omega'} |f|^2 w dA \right)^{1/2} \text{ for any } f \in L^2(wdA). \quad (17)$$

Theorem 3 *Let $0 < p < \infty$ and $N \geq 1$. For $\alpha, \beta > 0$, let $\Omega = [0, \alpha] \times [0, \beta]$ and $\Omega' = [-\beta, \alpha + \beta] \times [0, 2\beta]$. Suppose $\lambda : \Omega \longrightarrow [0, \infty)$ is a continuous length function satisfying (3), (4) and $w : \Omega' \longrightarrow [0, \infty)$ is a continuous weight function satisfying (6). A bounded function $f : \Omega \longrightarrow \mathbb{R}$ is continuous on $\Omega \setminus \{(x, 0) : 0 \leq x \leq \alpha\}$ if and only if the induced persistence block map $\widetilde{f} : P_N(\Omega) \longrightarrow L^2(wdA)$ is continuous with respect to the p-Wasserstein distance W_p on $P_N(\Omega)$ and the norm $\|\cdot\|_2$ on the Hilbert space $L^2(wdA)$.*

Proof Suppose the induced persistence block map $\widetilde{f} : P_N(\Omega) \longrightarrow L^2(wdA)$ is continuous. To show the underlying function $f : \Omega \longrightarrow \mathbb{R}$ is continuous on $\Omega \setminus \{(x, 0) : 0 \leq x \leq \alpha\}$, let $(a, b) \in \Omega$ with $b > 0$ and let $\epsilon > 0$. We are to show there is a $\delta > 0$ such that

$$|f(x,y) - f(a,b)| < \epsilon, \quad \text{whenever } (x,y) \in \Omega \text{ with } \|(x,y) - (a,b)\|_\infty < \delta.$$

To this end, by the continuity of the length function λ and the assumption that $b > 0$, there exists a real number δ_1 with $0 < \delta_1 < \lambda(a,b)/4$ such that

$$\text{Area}(E(x,y) \cap E(a,b)) \geq \frac{1}{2}\text{Area}(E(a,b)) > 0, \text{ whenever } \|(x,y)-(a,b)\|_\infty < \delta_1. \tag{18}$$

Consider the compact square K centered at (a,b) given by

$$K = \{(x,y) \in \Omega : \|(x,y)-(a,b)\|_\infty \leq \delta_1\}.$$

Since $0 < \delta_1 < \lambda(a,b)/4$, the compact square K lies entirely above the x-axis, and hence $0 < \lambda(x,y) < 2y$ for all $(x,y) \in K$, by our definition of the length function λ. Furthermore, since λ is continuous, the function $(x,y) \mapsto y - \frac{\lambda(x,y)}{2}$ is continuous and stays positive on the compact set K. Thus, the set $\bigcup_{(x,y)\in K} E(x,y)$ is contained in a compact set that lies entirely above the x-axis. Thus, by the continuity the functions w and λ, there is a real number $\eta > 0$ such that

$$\eta < \min\{\lambda(x,y) : (x,y) \in K\}, \text{ and} \tag{19}$$

$$\eta < \min\left\{ w(u,v) : (u,v) \in \bigcup_{(x,y)\in K} E(x,y) \right\}. \tag{20}$$

By assumption, the induced persistence block map $\widetilde{f} : P_N(\Omega) \longrightarrow L^2(wdA)$ is continuous at the persistence diagram $D_0 = \{(a,b)\}$ in $P_N(\Omega)$ that consists of the single point (a,b). Thus, there exists a $\delta_2 > 0$ such that

$$\|\widetilde{f}(D) - \widetilde{f}(D_0)\|_2 < \frac{\eta^{3/2}\epsilon}{\sqrt{2}}, \text{ whenever } W_p(D,D_0) < \delta_2. \tag{21}$$

Select a $\delta > 0$ such that $0 < \max\{\delta, \delta^p\} < \min\{\delta_1, \delta_2\}$. For any $(x,y) \in \Omega$ with

$$\|(x,y)-(a,b)\|_\infty < \delta, \tag{22}$$

consider the persistence diagram $D = \{(x,y)\} \in P_N(\Omega)$. Observe that

$$\begin{aligned} W_p(D,D_0) &\leq \begin{cases} \|(x,y)-(a,b)\|_\infty, & \text{if } 1 \leq p < \infty \\ \|(x,y)-(a,b)\|_\infty^p, & \text{if } 0 < p < 1 \end{cases} \\ &< \max\{\delta, \delta^p\} \\ &< \delta_2. \end{aligned}$$

Hence, the upper estimation for $\|\widetilde{f}(D) - \widetilde{f}(D_0)\|_2$ in (21) holds for any $(x,y) \in \Omega$ satisfying (22). For a lower estimation, observe that

$$\begin{aligned} \|\widetilde{f}(D) - \widetilde{f}(D_0)\|_2^2 &= \int_{\Omega'} \left| f(x,y)\chi_{E(x,y)} - f(a,b)\chi_{E(a,b)} \right|^2 w dA \qquad (23) \\ &\geq \int_{E(x,y)\cap E(a,b)} |f(x,y) - f(a,b)|^2 \, w dA \\ &\geq \eta \text{Area}(E(x,y) \cap E(a,b)) \, |f(x,y) - f(a,b)|^2, \text{ by (20)} \\ &\geq \frac{\eta}{2}\text{Area}(E(a,b)) \, |f(x,y) - f(a,b)|^2, \text{ by (18)} \\ &\geq \frac{\eta^3}{2} \, |f(x,y) - f(a,b)|^2, \text{ by (19)}. \end{aligned}$$

Combining estimations in (21) and (23) yields

$$|f(x,y)-f(a,b)| \leq \frac{\sqrt{2}}{\eta^{3/2}}\|\widetilde{f}(D)-\widetilde{f}(D_0)\|_2 < \frac{\sqrt{2}}{\eta^{3/2}}\cdot\frac{\epsilon\eta^{3/2}}{\sqrt{2}} = \epsilon,$$

which completes the proof of the continuity of the function f.

Conversely, suppose the bounded function $f:\Omega\longrightarrow\mathbb{R}$ is continuous on $\Omega\setminus\{(x,0):0\leq x\leq\alpha\}$. To prove the induced persistence block map $\widetilde{f}:P_N(\Omega)\longrightarrow L^2(wdA)$ is continuous, select a persistence diagram $D_0\in P_N(\Omega)$ and an $\epsilon>0$. We are to show there is a real number $\delta>0$ such that

$$\|\widetilde{f}(D)-\widetilde{f}(D_0)\|_2<\epsilon, \quad \text{whenever } D\in P_N(\Omega) \text{ with } W_p(D,D_0)<\delta. \tag{24}$$

To this end, since the function f is bounded and the functions λ, w are continuous on the compact sets $\Omega=[0,\alpha]\times[0,\beta]$ and $\Omega'=[-\beta,\alpha+\beta]\times[0,2\beta]$ respectively, we may select constants $M,C>0$ such that

$$\|f\|_\infty = \sup\{|f(x,y)|:(x,y)\in\Omega\}\leq M, \tag{25}$$
$$\|\lambda\|_\infty = \sup\{\lambda(x,y):(x,y)\in\Omega\}\leq M, \tag{26}$$
$$\|w\|_\infty = \sup\left\{w(x,y):(x,y)\in\Omega'\right\} = C. \tag{27}$$

Furthermore, by the continuity of the length function λ on the compact set Ω, the function λ is uniformly continuous on Ω. Thus, there exists a real number $\delta_1>0$ such that

$$|\lambda(x,y)-\lambda(a,b)|<\frac{\epsilon^2}{192CM^3N^2}, \text{ whenever } \|(x,y)-(a,b)\|_\infty<\delta_1. \tag{28}$$

Next, by the definition of the length function in (3), select a real number c such that

$$0<c<\min\{\lambda(a,b):(a,b)\in D_0 \text{ with } b>0\}<2\beta. \tag{29}$$

Likewise by the continuity on a compact set, the function f is uniformly continuous on the compact set $[0,\alpha]\times[\frac{c}{2},\beta]$, and so there exists a real number $\delta_2>0$ such that

$$|f(x,y)-f(a,b)|<\frac{\epsilon^2}{192CM^3N^2}. \tag{30}$$

whenever $(x,y),(a,b)\in[0,\alpha]\times[\frac{c}{2},\beta]$ with $\|(x,y)-(a,b)\|_\infty<\delta_2$. Select a real number $\delta>0$ such that

$$0<\max\{\delta,\delta^{1/p}\}<\min\left\{\frac{c}{2},\,\delta_1,\,\delta_2,\,\frac{\epsilon^2}{192CM^3N^2}\right\}. \tag{31}$$

To show this is the desired δ for (24) to hold, let D be any persistence diagram in $P_N(\Omega)$ with

$$W_p(D,D_0)<\delta. \tag{32}$$

Since D, D_0 are in $P_N(\Omega)$, the p-Wasserstein distance $W_p(D,D_0)$ is attained for a specific bijection γ. Furthermore, in order to have a bijection γ between D and D_0, we may need to

add at most N points along the x-axis to each persistence diagram. Thus, we may assume the persistence diagrams D, D_0 each has m points with $1 \leq m \leq 2N$, and without loss of generality, we can assume the points in D and D_0 are arranged as

$$D_0 = \{(a_j, b_j) : 1 \leq j \leq m\} \text{ and } D = \{(x_j, y_j) : 1 \leq j \leq m\} \tag{33}$$

so that we use (15) and (16) to rewrite (32) as

$$\begin{aligned} W_p(D, D_0) &= \begin{cases} \left(\sum_{j=1}^{m} \|(x_j, y_j) - (a_j, b_j)\|_\infty^p\right)^{1/p}, & \text{if } 1 \leq p < \infty \\ \sum_{j=1}^{m} \|(x_j, y_j) - (a_j, b_j)\|_\infty^p, & \text{if } 0 < p < 1 \end{cases} \\ &< \delta. \end{aligned} \tag{34}$$

Moreover, we can assume, by the symmetry of the roles of D and D_0 in (34), there exists an integer m_0 with $1 \leq m_0 \leq m$ and

$$b_j > 0 \text{ if and only if } 1 \leq j \leq m_0.$$

In order to show $\|\widetilde{f}(D) - \widetilde{f}(D_0)\|_2 < \epsilon$, we first establish the following claim.

Claim. For integers j with $1 \leq j \leq m$, we have

$$\int_{\Omega'} \left| f(x_j, y_j)\chi_{E(x_j,y_j)} - f(a_j, b_j)\chi_{E(a_j,b_j)} \right|^2 w dA < \frac{\epsilon^2}{16N^2}.$$

Proof of Claim. In the case that $1 \leq m_0 < m$, we see that if an integer j satisfies $m_0 < j \leq m$, then we have $b_j = 0$ and $\lambda(a_j, b_j) = 0$, which in turn implies $\text{Area}(E(x_j, y_j) \cap E(a_j, b_j)) = 0$. Thus, if $m_0 < j \leq m$, by Lemma 2, Statement (2), we have

$$\begin{aligned} \int_{\Omega'} \left| f(x_j, y_j)\chi_{E(x_j,y_j)} - f(a_j, b_j)\chi_{E(a_j,b_j)} \right|^2 w dA &\leq 2CM^3 \|(x_j, y_j) - (a_j, b_j)\|_\infty \\ &\leq 2CM^3 \max\{\delta, \delta^{1/p}\}, \text{ by (34)} \\ &< 2CM^3 \cdot \frac{\epsilon^2}{192CM^3N^2}, \text{ by (31)} \\ &< \frac{\epsilon^2}{16N^2}, \end{aligned}$$

which establishes the claim for the integers j satisfying $m_0 < j < m$.

For an integer j with $1 \leq j \leq m_0$, we have $b_j > 0$ and by (29), we have $\lambda(a_j, b_j) \geq c > 0$. It follows from (3) that $(a_j, b_j) \in [0, \alpha] \times [\frac{c}{2}, \beta]$. Also, observe that by (34)

$$\begin{aligned} \|(x_j, y_j) - (a_j, b_j)\|_\infty &\leq \begin{cases} W_p(D, D_0), & \text{if } 1 \leq p < \infty \\ W_p(D, D_0)^{1/p}, & \text{if } 0 < p < 1 \end{cases} \\ &< \max\{\delta, \delta^{1/p}\} \\ &\leq \frac{c}{2}, \text{ by (31)} \\ &\leq \frac{\lambda(a_j, b_j)}{2}, \text{ by (29)} \\ &\leq \frac{\lambda(x_j, y_j) + \lambda(a_j, b_j)}{2}, \text{ because } \lambda(x_j, y_j) \geq 0. \end{aligned}$$

Hence, we have $(x_j, y_j), (a_j, b_j) \in [0, \alpha] \times [\frac{c}{2}, \beta]$ with $\|(x_j, y_j) - (a_j, b_j)\|_\infty < \max\{\delta, \delta^{1/p}\}$. Furthermore, by Lemma 1, we have $\text{Area}(E(x_j, y_j) \cap E(a_j, b_j)) > 0$.

Returning to the integral in our claim, observe that Statement (1) of Lemma 2 gives us

$$\begin{aligned}
&\int_{\Omega'} |f(x_j, y_j)\chi_{E(x_j,y_j)} - f(a_j, b_j)\chi_{E(a_j,b_j)}|^2 w dA \qquad (35)\\
&\leq 4CM^3\Big(|f(x_j, y_j) - f(a_j, b_j)| + |\lambda(x_j, y_j) - \lambda(a_j, b_j)| + \|(x_j, y_j) - (a_j, b_j)\|_\infty\Big)\\
&< 4CM^3\left(\frac{\epsilon^2}{192CM^3N^2} + \frac{\epsilon^2}{192CM^3N^2} + \frac{\epsilon^2}{192CM^3N^2}\right), \text{ by (30), (28), (31)}\\
&= \frac{\epsilon^2}{16N^2},
\end{aligned}$$

which concludes the proof of our claim. □

Using the Claim, we now finish the proof of the theorem by showing $\|\widetilde{f}(D) - \widetilde{f}(D_0)\|_2 < \epsilon$. Observe that

$$\begin{aligned}
\|\widetilde{f}(D) - \widetilde{f}(D_0)\|_2 &= \left\|\sum_{j=1}^m f(x_j, y_j)\chi_{E(x_j,y_j)} - \sum_{j=1}^m f(a_j, b_j)\chi_{E(a_j,b_j)}\right\|_2\\
&\leq \sum_{j=1}^m \|f(x_j, y_j)\chi_{E(x_j,y_j)} - f(a_j, b_j)\chi_{E(a_j,b_j)}\|_2\\
&= \sum_{j=1}^m \left(\int_{\Omega'} |f(x_j, y_j)\chi_{E(x_j,y_j)} - f(a_j, b_j)\chi_{E(a_j,b_j)}|^2 w dA\right)^{1/2}.
\end{aligned}$$

In the case that $m_0 < m$, we use our claim to get

$$\begin{aligned}
\|\widetilde{f}(D) - \widetilde{f}(D_0)\|_2 &< \sum_{j=1}^{m_0} \frac{\epsilon}{4N} + \sum_{j=m_0+1}^{m} \frac{\epsilon}{4N} \qquad (36)\\
&= m_0\frac{\epsilon}{4N} + (m - m_0)\frac{\epsilon}{4N}\\
&\leq 2N\frac{\epsilon}{4N} + 2N\frac{\epsilon}{4N}\\
&= \epsilon.
\end{aligned}$$

Lastly, in the case that $m_0 = m$, we do not have the second summand on the right-hand side of (36), but the same argument gives $\|\widetilde{f}(D) - \widetilde{f}(D_0)\|_2 < \epsilon$. ■

A continuous function $f : [0, \infty) \times [0, \infty) \longrightarrow \mathbb{R}$ is always bounded on the compact set $\Omega = [0, \alpha] \times [0, \beta]$ with $\alpha, \beta > 0$ and continuous on the set $\Omega \setminus \{(x, 0) : 0 \leq x \leq \alpha\}$. Thus, Theorem 3 yields the following corollary.

Corollary 4 *Let $0 < p < \infty$, let $f : [0, \infty) \times [0, \infty) \longrightarrow \mathbb{R}$ be a continuous function and let λ, w be as given in Theorem 3. For any integer $N \geq 1$ and $\Omega = [0, \alpha] \times [0, \beta]$ with $\alpha, \beta > 0$, the induced persistence block map $\widetilde{f} : P_N(\Omega) \longrightarrow L^2(wdA)$ is continuous with respect to the p-Wasserstein distance W_p on $P_N(\Omega)$ and the norm $\|\cdot\|_2$ of $L^2(wdA)$.*

The corollary provides a theoretical result on the continuity of the induced persistence block map. On a practical note, we infer from the corollary that the computation of the vector $\widetilde{f}(D)$ for any given persistence diagram D is robust with respect to the noises that incur in D.

3.2 Identification Map

We continue the study of induced persistence block maps $\widetilde{f} : P_N(\Omega) \longrightarrow L^2(wdA)$ in the special case that the underlying function f is simply the constant function 1. In this situation, the induced persistence block map becomes what we call the *identification map* $\rho : P_N(\Omega) \longrightarrow L^2(wdA)$ given by

$$\rho(D) = \sum_{(a,b)\in D} \chi_{E(a,b)}. \tag{37}$$

Since the constant function is continuous on Ω, by Theorem 3, the identification map $\rho : P_N(\Omega) \longrightarrow L^2(wdA)$ is continuous with respect to the p-Wasserstein distance W_p for any p with $0 < p < \infty$. When the length function λ is Lipschitz, the identification map ρ is stable with respect to the 1/2-Wasserstein distance.

While one may be more familiar with the 2-Wasserstein distance, the 1/2-Wasserstein distance also naturally arises due to the square root in the definition of the Hilbert space norm on $L^2(wdA)$. In some situations, the 1/2-Wasserstein distance provides a tighter bound than the 2-Wasserstein distance. For an illustration, take two persistence diagrams $D = \{(1,1),(20,5)\}$ and $D_0 = \{(2,2),(20,1)\}$. Observe that

$$\begin{aligned} W_{1/2}(D,D_0) &= \|(1,1)-(2,2)\|_\infty^{1/2} + \|(20,5)-(20,1)\|_\infty^{1/2} \\ &= \sqrt{1}+\sqrt{4} = 3, \\ W_2(D,D_0) &= \left(\|(1,1)-(2,2)\|_\infty^2 + \|(20,5)-(20,1)\|_\infty^2\right)^{1/2} \\ &= \sqrt{1^2+4^2} = \sqrt{17} \approx 4.12, \end{aligned}$$

and so $W_{1/2}(D,D_0) < W_2(D,D_0)$. Using the norm notation $\|\cdot\|_2$ of $L^2(wdA)$ given by (17), we state the main theorem of this subsection.

Theorem 5 *Let $N \geq 1$, and let $\Omega = [0,\alpha]\times[0,\beta]$ and $\Omega' = [-\beta,\alpha+\beta]\times[0,2\beta]$ for $\alpha,\beta > 0$. Suppose the length function $\lambda : \Omega \longrightarrow [0,\infty)$ given in (3), (4) is Lipschitz with respect to the sup-norm on Ω and the weight function $w : \Omega' \longrightarrow [0,\infty)$ satisfies (6). Then, the identification map $\rho : P_N(\Omega) \longrightarrow L^2(wdA)$ is stable with respect to the 1/2-Wasserstein distance. In particular, if we select scalars $M, C > 0$ such that*

$$|\lambda(x,y)-\lambda(a,b)| \leq M\|(x,y)-(a,b)\|_\infty \text{ whenever } (x,y),(a,b)\in\Omega, \text{ and} \tag{38}$$

$$\|\lambda\|_\infty \leq M \text{ and } \|w\|_\infty = C, \tag{39}$$

then we have

$$\left(\int_\Omega |\rho(D)-\rho(D_0)|^2 wdA\right)^{1/2} \leq \|\rho(D)-\rho(D_0)\|_2 \leq 2\sqrt{C}\sqrt{M(M+1)}\; W_{1/2}(D,D_0) \tag{40}$$

for any $D, D_0 \in P_N(\Omega)$.

Proof The first inequality in (40) is obvious because $\Omega \subseteq \Omega'$. To establish the second inequality (40), let D, D_0 be any two persistence diagrams in $P_N(\Omega)$. Following the same reasons as in the proof of Theorem 3, the 1/2-Wasserstein distance $W_{1/2}(D, D_0)$ is attained for a specific bijection $\gamma : D \longrightarrow D_0$. For this particular bijection γ to exist, we may need to add at most N points along the x-axis to each persistence diagram. Hence, we may assume the persistence diagrams D, D_0 each has m points with $1 \leq m \leq 2N$, and without loss of generality, the points in D and D_0 are arranged such that

$$D = \{(x_j, y_j) : 1 \leq j \leq m\} \text{ and } D_0 = \{(a_j, b_j) : 1 \leq j \leq m\},$$

and

$$W_{1/2}(D, D_0) = \sum_{j=1}^{m} \|(x_j, y_j) - (a_j, b_j)\|_\infty^{1/2}.$$

In order to estimate $\|\rho(D) - \rho(D_0)\|_2$, observe that

$$\begin{aligned}
\|\rho(D) - \rho(D_0)\|_2 &= \left\| \sum_{j=1}^{m} \chi_{E(x_j,y_j)} - \sum_{j=1}^{m} \chi_{E(a_j,b_j)} \right\|_2 \\
&\leq \sum_{j=1}^{m} \|\chi_{E(x_j,y_j)} - \chi_{E(a_j,b_j)}\|_2 \\
&= \sum_{j=1}^{m} \left(\int_{\Omega'} |\chi_{E(x_j,y_j)} - \chi_{E(a_j,b_j)}|^2 w dA \right)^{1/2} \\
&= \sum_{j=1}^{m} \left(\int_{E(x_j,y_j)\setminus E(a_j,b_j)} w dA + \int_{E(a_j,b_j)\setminus E(x_j,y_j)} w dA \right)^{1/2} \\
&\leq \sum_{j=1}^{m} \Big(C\text{Area}(E(x_j, y_j) \setminus E(a_j, b_j)) + C\text{Area}(E(a_j, b_j) \setminus E(x_j, y_j)) \Big)^{1/2} \\
&= \sum_{j=1}^{m} \sqrt{C} \Big(\text{Area}(E(x_j, y_j) \setminus E(a_j, b_j)) + \text{Area}(E(a_j, b_j) \setminus E(x_j, y_j)) \Big)^{1/2}.
\end{aligned} \tag{41}$$

For each integer j in the above summation, we either have $\text{Area}(E(x_j, y_j) \cap E(a_j, b_j)) > 0$ or $\text{Area}(E(x_j, y_j) \cap E(a_j, b_j)) = 0$. For the case when $\text{Area}(E(x_j, y_j) \cap E(a_j, b_j)) > 0$, by inequality (12) in the proof of Lemma 2, we have

$$\begin{aligned}
&\text{Area}(E(x_j, y_j) \setminus E(a_j, b_j)) + \text{Area}(E(a_j, b_j) \setminus E(x_j, y_j)) \\
&\leq 4M \Big(|\lambda(x_j, y_j) - \lambda(a_j, b_j)| + \|(x_j, y_j) - (a_j, b_j)\|_\infty \Big) \\
&\leq 4M \Big(M\|(x_j, y_j) - (a_j, b_j)\|_\infty + \|(x_j, y_j) - (a_j, b_j)\|_\infty \Big), \text{ by (38)} \\
&= 4M(M + 1)\|(x_j, y_j) - (a_j, b_j)\|_\infty.
\end{aligned} \tag{42}$$

For the case when $\text{Area}(E(x_j, y_j) \cap E(a_j, b_j)) = 0$, by inequality (14) in the proof of Lemma 2, we have

$$\begin{aligned} &\text{Area}(E(x_j, y_j) \setminus E(a_j, b_j)) + \text{Area}(E(a_j, b_j) \setminus E(x_j, y_j)) \\ &\leq 2M\|(x_j, y_j) - (a_j, b_j)\|_\infty \\ &\leq 4M(M+1)\|(x_j, y_j) - (a_j, b_j)\|_\infty. \end{aligned} \tag{43}$$

Returning to the estimation of $\|\rho(D) - \rho(D_0)\|_2$ in (41), the above inequalities (42) and (43) give us

$$\begin{aligned} \|\rho(D) - \rho(D_0)\|_2 &\leq \sum_{j=1}^{m} \sqrt{C}\Big(\text{Area}(E(x_j, y_j) \setminus E(a_j, b_j)) + \text{Area}(E(a_j, b_j) \setminus E(x_j, y_j))\Big)^{1/2} \\ &\leq \sum_{j=1}^{m} \sqrt{C}\Big(4M(M+1)\|(x_j, y_j) - (a_j, b_j)\|_\infty\Big)^{1/2} \\ &= 2\sqrt{C}\sqrt{M(M+1)}\sum_{j=1}^{m} \|(x_j, y_j) - (a_j, b_j)\|_\infty^{1/2} \\ &= 2\sqrt{C}\sqrt{M(M+1)}\; W_{1/2}(D, D_0), \end{aligned}$$

which finishes the whole proof of the Theorem. ■

The identification map ρ is defined on $P_N(\Omega)$, but the estimation in inequality (40) in Theorem 5 is independent of the integer N. From this observation, we can say the identification map $\rho : P(\Omega) \longrightarrow L^2(wdA)$ is stable with respect to the 1/2-Wasserstein distance. While we initially establish the stability of the identification map ρ relative to the 1/2-Wasserstein distance, we can extend our stability result using any p-Wasserstein distance for $\frac{1}{2} < p < \infty$ when we restrict to $P_N(\Omega)$ for any fixed integer $N \geq 1$. To achieve the estimation, we require Hölder's Inequality: for any vectors $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$, we have

$$\sum_{j=1}^{n} |x_j y_j| \leq \left(\sum_{j=1}^{n} |x_j|^{q_0}\right)^{1/q_0} \left(\sum_{j=1}^{n} |y_j|^{p_0}\right)^{1/p_0}, \tag{44}$$

where p_0, q_0 satisfy $1 < p_0, q_0 < \infty$ and $\frac{1}{p_0} + \frac{1}{q_0} = 1$.

Before we state and prove a stability result in the next corollary, we need a clarification about the choice of a weight function. At first glance, a prescribed weight function W and a scaled version $w = tW$, where t is a positive constant, may appear to produce different results when we use them to obtain a finite dimensional vector summary from a given persistence block; that is, the vectorized persistence block (VPB) that was briefly discussed in the Introduction. However, on a second thought, the role of a weight function W in $L^2(WdA)$ is to put different weights on different parts of Ω', and so a weight function W and a scaled weight function $w = tW$ give the exact same relative weight on different parts of Ω'. In practice, when VPBs are standardized as input data, both W and $w = tW$ indeed produce

the same results. This point is illustrated in Section 4, which focuses on simulation and implementation.

In the following corollary, we show how to select the positive constant t, which depends on the maximum number N of data points that we are allowed to have and other parameters in our method to achieve the stability result.

Corollary 6 *Let $N \geq 1$, and let $\Omega = [0, \alpha] \times [0, \beta]$ and $\Omega' = [-\beta, \alpha + \beta] \times [0, 2\beta]$ for $\alpha, \beta > 0$. Suppose the length function $\lambda : \Omega \longrightarrow [0, \infty)$ satisfies the hypotheses of Theorem 5. Any continuous weight function $W : \Omega' \longrightarrow [0, \infty)$ can be scaled to a weight function $w : \Omega' \longrightarrow [0, \infty)$ so that the identification map $\rho : P_N(\Omega) \longrightarrow L^2(wdA)$ satisfies*

$$\left(\int_\Omega |\rho(D) - \rho(D_0)|^2 w dA\right)^{1/2} \leq \|\rho(D) - \rho(D_0)\|_2 \leq \begin{cases} W_p(D, D_0)^{1/2}, & \text{if } 1 \leq p < \infty \\ W_p(D, D_0)^{1/2p}, & \text{if } \frac{1}{2} < p < 1, \end{cases}$$

for any $D, D_0 \in P_N(\Omega)$. In particular for $p = 2$, we have

$$\left(\int_\Omega |\rho(D) - \rho(D_0)|^2 w dA\right)^{1/2} \leq \|\rho(D) - \rho(D_0)\|_2 \leq W_2(D, D_0)^{1/2}.$$

Proof Let the real number p satisfy $\frac{1}{2} < p < \infty$. As in the proof of Theorem 5, for any two persistence diagrams $D, D_0 \in P_N(\Omega)$, select the bijection $\gamma : D \longrightarrow D_0$ and the integer m with $1 \leq m \leq 2N$ so that we arrange the points in D and D_0 as

$$D = \{(x_j, y_j) : 1 \leq j \leq m\} \text{ and } D_0 = \{(a_j, b_j) : 1 \leq j \leq m\},$$

for which

$$W_p(D, D_0) = \begin{cases} \left(\sum_{j=1}^m \|(x_j, y_j) - (a_j, b_j)\|_\infty^p\right)^{1/p}, & \text{if } 1 \leq p < \infty \\ \sum_{j=1}^m \|(x_j, y_j) - (a_j, b_j)\|_\infty^p, & \text{if } \frac{1}{2} < p < 1. \end{cases} \tag{45}$$

Now, any given weight function W can be scaled by a positive constant t to obtain a weight function $w = tW$ whose norm satisfies

$$\|w\|_\infty = C, \quad \text{where} \quad C = \frac{1}{16N^2M(M+1)}. \tag{46}$$

It follows from (40) that

$$\begin{aligned} \|\rho(D) - \rho(D_0)\|_2 &\leq 2\sqrt{C}\sqrt{M(M+1)}\sum_{j=1}^m \|(x_j, y_j) - (a_j, b_j)\|_\infty^{1/2} \\ &= \sum_{j=1}^m 2\sqrt{C}\sqrt{M(M+1)}\|(x_j, y_j) - (a_j, b_j)\|_\infty^{1/2}. \end{aligned} \tag{47}$$

Applying Hölder's inequality in (44) to (47) with $p_0 = 2p > 1$ and $q_0 = \frac{1}{1-\frac{1}{p_0}} = \frac{2p}{2p-1} > 1$ yields

$$\begin{aligned}
&\|\rho(D) - \rho(D_0)\|_2 \\
&\leq \left(\sum_{j=1}^{m} \left(2\sqrt{C}\sqrt{M(M+1)}\right)^{q_0}\right)^{1/q_0} \left(\sum_{j=1}^{m} \|(x_j, y_j) - (a_j, b_j)\|_\infty^{p_0/2}\right)^{1/p_0} \\
&= 2m^{1/q_0}\sqrt{C}\sqrt{M(M+1)} \left(\sum_{j=1}^{m} \|(x_j, y_j) - (a_j, b_j)\|_\infty^{p}\right)^{1/2p}, \text{ because } p_0 = 2p \\
&\leq 4N\sqrt{C}\sqrt{M(M+1)} \left(\sum_{j=1}^{m} \|(x_j, y_j) - (a_j, b_j)\|_\infty^{p}\right)^{1/2p}, \text{ because } m^{1/q_0} \leq (2N)^{1/q_0} \leq 2N \\
&= \begin{cases} W_p(D, D_0)^{1/2}, & \text{if } 1 \leq p < \infty \\ W_p(D, D_0)^{1/2p}, & \text{if } \frac{1}{2} < p < 1, \end{cases} \text{ by (45) and (46).}
\end{aligned}$$

■

It follows from the above theorem and corollary that with respect to $W_p(D, D_0)$ for any $p \geq 1/2$, the computation of $\rho(D)$ for any persistence diagram D is robust with respect to the noises that incur in D, under an assumption on the length function λ and the appropriate scaled weight function w which serves our purpose.

3.3 Vectorization

Now, for a persistence diagram D, we would like to associate every persistence block $\widetilde{f}(D)$, which is some vector in the Hilbert space $L^2(wdA)$, with a vector in $\mathbb{R}^n$. There are two ways to do that, depending whether one wants to focus on $\Omega = [0, \alpha]\times[0, \beta]$ or $\Omega' = [-\beta, \alpha+\beta]\times[0, 2\beta]$ as discussed in the previous two subsections. Since both ways share the same mathematical theory, we develop them together as one in this subsection. To this end, select a partition $\mathcal{P}$ of either Ω or Ω', whichever one chooses to focus, into n grid cells $S_1, S_2, \ldots, S_n$. For any persistence diagram D in $P(\Omega)$, let

$$I_{f,S_i}(D) = \int_{S_i} \widetilde{f}(D)\, wdA,$$

and define the vector

$$\widetilde{f}_{\mathcal{P}}(D) = (I_{f,S_1}(D), I_{f,S_2}(D), \ldots, I_{f,S_n}(D))$$

in $\mathbb{R}^n$ to be the *vectorized persistence block* (VPB). Thus we can think of $\widetilde{f}_{\mathcal{P}}$ as a map $\widetilde{f}_{\mathcal{P}} : P(\Omega) \longrightarrow \mathbb{R}^n$.

To ensure the theory accounts for both types of partitions, let

$$G = \begin{cases} \Omega, & \text{if } \mathcal{P} \text{ is a partition of } \Omega \\ \bigcup_{(x,y)\in\Omega} E(x, y), & \text{if } \mathcal{P} \text{ is a partition of } \Omega' \end{cases}. \tag{48}$$

Note that $\Omega \subseteq G \subseteq \Omega'$, and the persistence block $\widetilde{f}(D)$ is a linear combination of characteristic functions $\chi_{E(x,y)}$ of the square $E(x,y)$ centered at $(x,y) \in \Omega$. Thus, whether $S_1, S_2, \ldots, S_n$ form a partition $\mathcal{P}$ of Ω or Ω', we have

$$\begin{aligned} I_{f,S_i}(D) &= \int_{S_i} \widetilde{f}(D) w dA \\ &= \int_{S_i \cap G} \widetilde{f}(D) w dA. \end{aligned} \tag{49}$$

The VPB $\widetilde{f}_{\mathcal{P}}(D)$ is a vector in $\mathbb{R}^n$, and so we may use the standard norms on $\mathbb{R}^n$ to measure the distance between two VPBs. In particular, for any two persistence diagrams $D, D_0 \in P_N(\Omega)$ and for $1 \leq p < \infty$, we have

$$\|\widetilde{f}_{\mathcal{P}}(D) - \widetilde{f}_{\mathcal{P}}(D_0)\|_p = \left(\sum_{i=1}^{n} |I_{f,S_i}(D) - I_{f,S_i}(D_0)|^p \right)^{1/p}, \tag{50}$$

and

$$\|\widetilde{f}_{\mathcal{P}}(D) - \widetilde{f}_{\mathcal{P}}(D_0)\|_\infty = \sup_{1 \leq i \leq n} |I_{f,S_i}(D) - I_{f,S_i}(D_0)|. \tag{51}$$

The induced persistence block map $\widetilde{f} : P_N(\Omega) \longrightarrow L^2(wdA)$ is continuous whenever the underlying function $f : [0,\infty) \times [0,\infty) \longrightarrow \mathbb{R}$ is bounded and continuous, by Theorem 3. In order to explore a similar continuity result for the VPB relative to the Hilbert space norm $\|\cdot\|_2$ given in (50) above, we establish an upper estimate for $\|\widetilde{f}_{\mathcal{P}}(D) - \widetilde{f}_{\mathcal{P}}(D_0)\|_2$, which involves the measure μ defined by

$$\mu(E) = \int_E d\mu = \int_E w dA.$$

Lemma 7 *Let $\Omega = [0,\alpha] \times [0,\beta]$ and $\Omega' = [-\beta, \alpha+\beta] \times [0, 2\beta]$ with $\alpha, \beta > 0$, let $\mathcal{P}$ be a partition of Ω or Ω' into n grid cells $S_1, S_2, \ldots, S_n$, and let G be the set defined in (48). For persistence diagrams $D = \{(x_j, y_j) : 1 \leq j \leq m\}$ and $D_0 = \{(a_j, b_j) : 1 \leq j \leq m\}$, we have*

$$\|\widetilde{f}_{\mathcal{P}}(D) - \widetilde{f}_{\mathcal{P}}(D_0)\|_2^2 \leq \sum_{i=1}^{n} \mu(S_i \cap G) \left[\sum_{j=1}^{m} \left(\int_{S_i} |f(x_j, y_j)\chi_{E(x_j,y_j)} - f(a_j, b_j)\chi_{E(a_j,b_j)}|^2 w dA \right)^{1/2} \right]^2.$$

Proof Observe that for integers i with $1 \leq i \leq n$, we have

$$
\begin{aligned}
& |I_{f,S_i}(D) - I_{f,S_i}(D_0)|^2 \\
= & \left| \int_{S_i \cap G} \widetilde{f}(D) w dA - \int_{S_i \cap G} \widetilde{f}(D_0) w dA \right|^2, \text{ by (49)} \\
\leq & \left(\int_{S_i \cap G} |\widetilde{f}(D) - \widetilde{f}(D_0)| w dA \right)^2 \\
\leq & \left(\int_{S_i \cap G} 1^2 w dA \right) \left(\int_{S_i \cap G} |\widetilde{f}(D) - \widetilde{f}(D_0)|^2 w dA \right), \text{ by the Cauchy-Schwarz Inequality} \\
\leq & \, \mu(S_i \cap G) \int_{S_i} |\widetilde{f}(D) - \widetilde{f}(D_0)|^2 w dA, \text{ because } S_i \cap G \subseteq S_i \text{ and the definition of } \mu \\
\leq & \, \mu(S_i \cap G) \left(\sum_{j=1}^{m} \| (f(x_j, y_j) \chi_{E(x_j,y_j)} - f(a_j, b_j) \chi_{E(a_j,b_j)}) \chi_{S_i} \|_2 \right)^2, \text{(use Triangle Inequality)} \\
= & \, \mu(S_i \cap G) \left(\sum_{j=1}^{m} \left(\int_{S_i} |(f(x_j, y_j) \chi_{E(x_j,y_j)} - f(a_j, b_j) \chi_{E(a_j,b_j)}|^2 w dA \right)^{1/2} \right)^2.
\end{aligned}
$$

Summing over all the grid cells S_i in the partition $\mathcal{P}$ leads to the desired inequality. ■

Similar to the identification map ρ, under some basic assumptions, we now show the vectorized persistence block map $\widetilde{f}_{\mathcal{P}} : P(\Omega) \longrightarrow \mathbb{R}^n$ is stable relative to the 1/2-Wasserstein distance.

Theorem 8 *Let $\Omega = [0, \alpha] \times [0, \beta]$ and $\Omega' = [-\beta, \alpha + \beta] \times [0, 2\beta]$ with $\alpha, \beta > 0$, let $\mathcal{P}$ be a partition of Ω or Ω' into n grid cells $S_1, S_2, \ldots, S_n$, and let G be the set defined in (48). Suppose $\lambda : \Omega \longrightarrow [0, \infty)$ is a length function satisfying (3), (4) and $w : \Omega' \longrightarrow [0, \infty)$ is a weight function satisfying (6). If the function $f : \Omega \longrightarrow \mathbb{R}$ and the length function $\lambda : \Omega \longrightarrow [0, \infty)$ are Lipschitz, then the vectorized persistence block map $\widetilde{f}_{\mathcal{P}} : P(\Omega) \longrightarrow \mathbb{R}^n$ is stable relative to the 1/2-Wasserstein distance. In particular, if the scalars $M, C > 0$ satisfy*

$$|f(x, y) - f(a, b)| \leq M \|(x, y) - (a, b)\|_\infty \text{ for all } (x, y), (a, b) \in \Omega; \tag{52}$$

$$|\lambda(x, y) - \lambda(a, b)| \leq M \|(x, y) - (a, b)\|_\infty \text{ for all } (x, y), (a, b) \in \Omega; \tag{53}$$

$$\|f\|_\infty \leq M, \ \|\lambda\|_\infty \leq M, \text{ and } \|w\|_\infty = C, \tag{54}$$

then we have

$$\|\widetilde{f}_{\mathcal{P}}(D) - \widetilde{f}_{\mathcal{P}}(D_0)\|_2 \leq 2\sqrt{C} M \sqrt{\mu(G) M (2M + 1)} \, W_{1/2}(D, D_0), \tag{55}$$

for any $D, D_0 \in P(\Omega)$.

Proof For any two persistence diagrams $D, D_0 \in P(\Omega)$, let N be the maximum cardinality of D and D_0. As in the proof of previous results, we may assume D, D_0 each has m points with $1 \leq m \leq 2N$, and the points in D, D_0 are arranged such that

$$D = \{(x_j, y_j) : 1 \leq j \leq m\} \text{ and } D_0 = \{(a_j, b_j) : 1 \leq j \leq m\}$$

and

$$W_{1/2}(D, D_0) = \sum_{j=1}^{m} \|(x_j, y_j) - (a_j, b_j)\|_\infty^{1/2}.$$

As in the proof of Theorem 5, we have two cases based on the area of $E(x_j, y_j) \cap E(a_j, b_j)$. For the case when $\text{Area}(E(x_j, y_j) \cap E(a_j, b_j)) > 0$, Lemma 2 and assumptions (52), (53) and (54) give us

$$\begin{aligned}
&\int_{S_i} |f(x_j, y_j)\chi_{E(x_j,y_j)} - f(a_j, b_j)\chi_{E(a_j,b_j)}|^2 w dA \\
&\leq \int_{\Omega'} |f(x_j, y_j)\chi_{E(x_j,y_j)} - f(a_j, b_j)\chi_{E(a_j,b_j)}|^2 w dA, \text{ because } S_i \subseteq \Omega' \\
&\leq 4CM^3 \Big(|f(x_j, y_j) - f(a_j, b_j)| + |\lambda(x_j, y_j) - \lambda(a_j, b_j)| + \|(x_j, y_j) - (a_j, b_j)\|_\infty \Big) \\
&\leq 4CM^3 \Big(M\|(x_j, y_j) - (a_j, b_j)\|_\infty + M\|(x_j, y_j) - (a_j, b_j)\|_\infty + \|(x_j, y_j) - (a_j, b_j)\|_\infty \Big) \\
&= 4CM^3(2M + 1)\|(x_j, y_j) - (a_j, b_j)\|_\infty.
\end{aligned}$$

For the case when $\text{Area}(E(x_j, y_j) \cap E(a_j, b_j)) = 0$, Lemma 2 gives us

$$\begin{aligned}
&\int_{S_i} |f(x_j, y_j)\chi_{E(x_j,y_j)} - f(a_j, b_j)\chi_{E(a_j,b_j)}|^2 w dA \\
&\leq \int_{\Omega'} |f(x_j, y_j)\chi_{E(x_j,y_j)} - f(a_j, b_j)\chi_{E(a_j,b_j)}|^2 w dA \\
&\leq 2CM^3 \ \|(x_j, y_j) - (a_j, b_j)\|_\infty \\
&\leq 4CM^3(2M + 1)\|(x_j, y_j) - (a_j, b_j)\|_\infty.
\end{aligned}$$

To finish the estimation of $\|\widetilde{f}_{\mathcal{P}}(D) - \widetilde{f}_{\mathcal{P}}(D_0)\|_2$ using Lemma 7, observe that

$$\begin{aligned}
\|\widetilde{f}_{\mathcal{P}}(D) - \widetilde{f}_{\mathcal{P}}(D_0)\|_2^2 &\leq \sum_{i=1}^{n} \mu(S_i \cap G) \left[\sum_{j=1}^{m} \left(\int_{S_i} |f(x_j, y_j)\chi_{E(x_j,y_j)} - f(a_j, b_j)\chi_{E(a_j,b_j)}|^2 w dA \right)^{1/2} \right]^2 \\
&\leq \sum_{i=1}^{n} \mu(S_i \cap G) \left[\sum_{j=1}^{m} \left(4CM^3(2M + 1)\|(x_j, y_j) - (a_j, b_j)\|_\infty \right)^{1/2} \right]^2 \\
&= 4CM^3(2M + 1) \sum_{i=1}^{n} \mu(S_i \cap G) \left[\sum_{j=1}^{m} \|(x_j, y_j) - (a_j, b_j)\|_\infty^{1/2} \right]^2 \\
&= 4CM^3(2M + 1) \ W_{1/2}(D, D_0)^2 \sum_{i=1}^{n} \mu(S_i \cap G) \\
&= 4CM^3\mu(G)(2M + 1) \ W_{1/2}(D, D_0)^2,
\end{aligned}$$

which concludes the proof. ■

Similar to the identification map ρ discussed in the previous subsection, the stability of the vectorized persistence block map $\widetilde{f}_{\mathcal{P}}$ relative to the 1/2-Wasserstein distance is independent of the number of points in the persistence diagrams $D, D_0 \in P(\Omega)$. If we restrict our attention to the persistence diagrams in $P_N(\Omega)$, the following corollary shows the vectorized persistence block map $\widetilde{f}_{\mathcal{P}}$ is stable relative to the p-Wasserstein distance for $\frac{1}{2} < p < \infty$. The proof of the corollary utilizes Hölder's inequality, following the exact same argument as in the proof of Corollary 6, and we need not repeat the argument here.

Corollary 9 *Assume the hypotheses of Theorem 8. For any given integer $N \geq 1$, the weight function $w : \Omega' \longrightarrow [0, \infty)$ can be scaled so that the VPBs $\widetilde{f}_{\mathcal{P}}(D)$ and $\widetilde{f}_{\mathcal{P}}(D_0)$ satisfy*

$$\|\widetilde{f}_{\mathcal{P}}(D) - \widetilde{f}_{\mathcal{P}}(D_0)\|_2 \leq \begin{cases} \sqrt{\mu(G)}\ W_p(D, D_0)^{1/2}, & \text{if } 1 \leq p < \infty \\ \sqrt{\mu(G)}\ W_p(D, D_0)^{1/2p}, & \text{if } \frac{1}{2} < p < 1 \end{cases},$$

for any $D, D_0 \in P_N(\Omega)$. In particular for $p = 2$, we have

$$\|\widetilde{f}_{\mathcal{P}}(D) - \widetilde{f}_{\mathcal{P}}(D_0)\|_2 \leq \sqrt{\mu(G)}\ W_2(D, D_0)^{1/2}.$$

The vectorization map $\widetilde{f}_{\mathcal{P}}(D)$ is stable with respect to the $p-$Wasserstein distance, when $1/2 \leq p < \infty$. The estimations provided in Theorem 8 and Corollary 9, do not even depend on the number n of grid cells in the partition $\mathcal{P}$ of Ω or Ω'. In practice, we can take $\mu(G)$ to be a fixed constant and we can have a maximum number N of points in a persistence diagram, and scale a weight function that serves our purpose in Corollary 9. However, the scaling is not necessary in Theorem 8.

4. Simulations

In this section we present our simulation results to illustrate the utility of vectorized persistence blocks (VPB) for classification, clustering and change point detection tasks. In all of the simulations, our results are compared with the corresponding results produced from persistence images (PI). Two of the simulations closely follow those of Sections 6.1 and 6.4 in Adams et al. (2017). The simulations involve both synthetic and real datasets. On 2D datasets, we extract topological information using the Vietoris-Rips filtration Ghrist (2008). For large 3D datasets, we resort to the α-shape complexes [see Edelsbrunner and Mücke (1994); Edelsbrunner and Harer (2010)] to avoid computational challenges that arise from using the Vietoris-Rips complexes in that situation.

To compute VPBs, we experiment with several choices of the weight function w and the side length function λ. For example, for λ we consider functions of the form

$$\lambda = F(y; \tau, n, m) = 2\tau y \left(\frac{y}{y_{max}}\right)^n \left(1 - \frac{y}{y_{max}}\right)^m,$$

where $\tau \in (0, 1]$, n, m are non-negative integers and y_{max} is the maximum of persistence values y for the PD in question. All our results are based on the side length $\lambda = F(y; \tau, 0, 0) = 2\tau y$, where $0 < \tau \leq 1$ and the weight function $w(x, y) = (x + y)/N^2$, where N is the maximum

number of points in all PDs under consideration for a given homological dimension. Note that when VPBs are standardized as inputs in the context of a learning task, the value of N will not affect the results and therefore can be set to one for convenience. Moreover, with the selected functions for weight and side length, VPBs can be computed analytically. PDs are computed using the `R` packages `TDA` Fasy et al. (2021) and `TDAstats` Wadhwa et al. (2018). To produce PIs we utilize the `R` package `kernelTDA` Padellini et al. (2020).

4.1 Clustering using k-medoids on a synthetic dataset

This simulation study considers the synthetic dataset used in Sections 6.1 of Adams et al. (2017) which consists of 25 point clouds of size 500 sampled uniformly from each of the following six shapes: a unit cube, a circle of radius $1/2$, a sphere of radius $1/2$, three clusters with random centers in the unit cube, three clusters each containing another set of three clusters, a torus with inner and outer diameters of $1/2$ and 1 respectively. Furthermore, Gaussian noise is added to each point cloud at two levels: $\eta = 0.05$ and $\eta = 0.1$. Topological signatures (in the form of VPBs and PIs) are computed from PDs for homological dimensions H_0 and H_1. In this study, the goal is to cluster the given 150 point clouds into 6 respective shape classes. For PIs, we use the parameter settings adopted by the authors of Adams et al. (2017): standard deviation $\sigma = 0.1$, a 20×20 grid and the linear weighting function. VPBs are computed with the candidate values of τ ranging in $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ on a grid of size 6×6. The final value of τ is selected based on the internal measure of the Davies-Bouldin Index Halkidi et al. (2002). More specifically, among the five different clustering performances (one for each τ), we choose that for which the Davies-Bouldin Index is minimum.

Following the approach in Adams et al. (2017), clustering is performed employing the k-medoids method [see Rdusseeun and Kaufman (1987); Park and Jun (2009)], which is a robust version of the well-known the k-means algorithm Lloyd (1982). It uses actual data points as cluster centers (called medoids) instead of centroids as in k-means and can be applied with any dissimilarity measure. To implement k-medoids in `R`, we use the `cluster::pam()` function with its default parameter settings Maechler et al. (2021).

A 150×150 matrix of pairwise distances (also called a dissimilarity matrix) between topological signatures (VPBs or PIs) is computed and passed to k-medoids (specifying the number of clusters to be six) which returns a list of cluster assignments. The performance is evaluated based on the accuracy of cluster predictions against the true cluster labels. Comparison of the clustering accuracies is given in Table 1. The results show that k-medoids produces slightly better accuracies on VPBs. Cluster visualizations of VPBs reduced via principal component analysis (PCA) are given in Figure 3.

Furthermore, we explore the sensitivity of accuracies for VPBs with respect to changes in tuning parameters. From Figures 4 and 5, we can observe that the accuracies appear to be fairly stable with respect to τ and grid size when the noise level $\eta = 0.05$.

Concerning the computational time, we observe that the cost to produce VPBs (0.09 and 0.188 seconds for H_0 and H_1) is lower than that of PIs (0.615 and 0.714 seconds for H_0 and H_1) despite running k-medoids five times for different values of τ. Here, the cost is the combined time to compute the topological signatures and to run the k-medoids algorithm. However, the observed difference in computational cost is partly due to the grid sizes being unequal (i.e. 6×6 for VPBs and 20×20 for PIs).

Metric	Dim	Noise	PI	VPB
L_1	H_0	0.05	0.933	**0.947**
L_2	H_0	0.05	0.927	**0.940**
L_∞	H_0	0.05	0.940	**0.947**
L_1	H_0	0.10	0.953	**0.960**
L_2	H_0	0.10	0.953	**0.960**
L_∞	H_0	0.10	**0.960**	**0.960**
L_1	H_1	0.05	**1.000**	**1.000**
L_2	H_1	0.05	**1.000**	**1.000**
L_∞	H_1	0.05	**1.000**	**1.000**
L_1	H_1	0.10	0.960	**0.973**
L_2	H_1	0.10	0.960	**0.973**
L_∞	H_1	0.10	**0.960**	**0.960**

Table 1: k-medoids accuracy results.

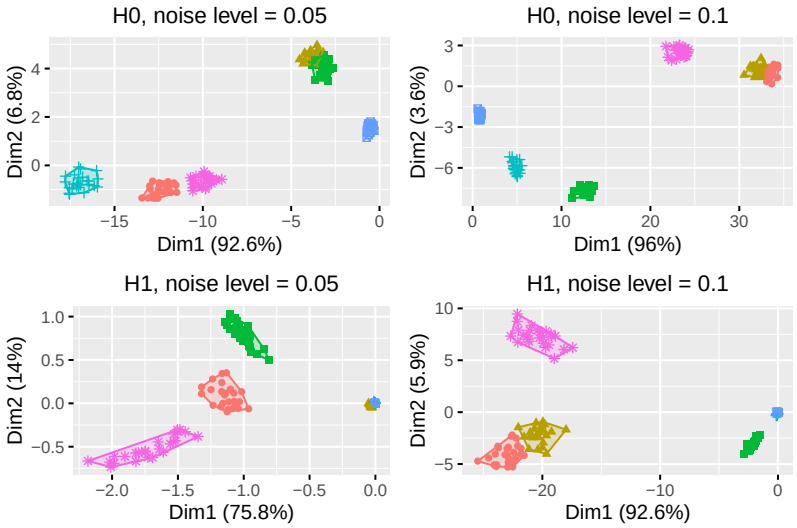

Figure 3: Visualization of VPB-based clusters using PCA.

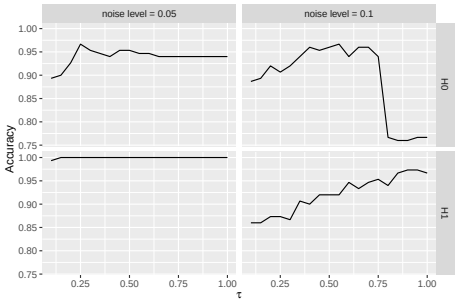

Figure 4: VPB accuracy as a function of τ. Grid size 6×6, metric L_2.

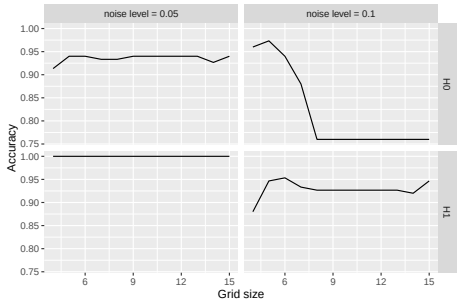

Figure 5: VPB accuracy as a function of grid size. $\tau = 0.75$, metric L_2.

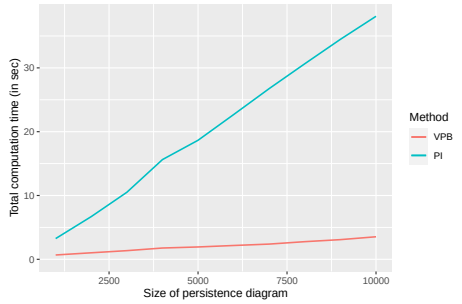

Figure 6: Comparison of computational cost.

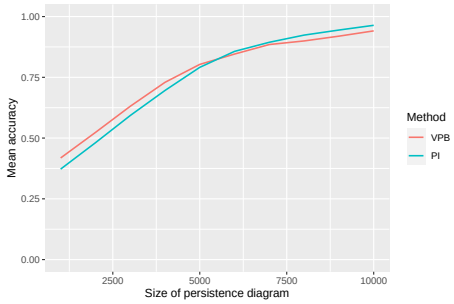

Figure 7: Comparison of accuracy.

Next, in order to have an objective comparison between VPBs and PIs, we examine their performance in terms of computational cost and accuracy, computed over the grids of equal size. We simulate 25 PDs from 7 different models and cluster them on the basis of the computed VPBs and PIs using k-medoids. We track the computational cost and clustering accuracy as functions of the size of PDs. More specifically, let $\{D_{ij}\}_{i=1,j=1}^{25,7}$ be PDs of size $n \in \{1000, 2000, \ldots, 10000\}$, where the birth values are randomly sampled from the uniform distribution on $(0, 1)$ and the persistence values from the beta distributions with the shape parameters $\alpha_1 = 1$, $\beta_1 = 1$; $\alpha_2 = 1$, $\beta_2 = 1.1$; $\alpha_3 = 1.1$, $\beta_3 = 1.1$; $\alpha_4 = 1.1$, $\beta_4 = 1$; $\alpha_5 = 1$, $\beta_5 = 0.9$; $\alpha_6 = 0.9$, $\beta_6 = 0.9$ and $\alpha_7 = 0.9$, $\beta_7 = 1$. We superimpose a grid of size of 10×10 to compute the VPBs and PIs. We use $\tau = 0.3$ for the VPBs and one half of the overall maximum persistence divided by the grid size (i.e., 10) as σ for the PIs. The plot of the total computational cost (in seconds, averaged over 100 independent trials) versus the size of PDs n is given in Figure 6, which reveals that, on average, VBPs are 8.4 times faster to compute than PIs. Figure 7 depicts the mean accuracy (computed over 100 independent trials) versus the persistence diagram size. According to the plot, the mean accuracies of the two methods are quite comparable. Since the differences among the 7 parameter settings for the beta distributions are very small, the resulting PDs, when plotted, are hard to visually distinguish from one another. Consequently, the mean accuracies are initially low but gradually improve as the size of PDs n increases.

4.2 Clustering using k-medoids on a real dataset

In our second simulation study, we perform clustering and shape retrieval tasks on a 3D dataset. The dataset consists of 100 point clouds derived from 3D scans of ten non-rigid toys, each having approximately 4000 points Limberger et al. (2017). All toys are scanned in ten different poses by moving various body parts around their joints (see Figures 8 and 9). The dataset contains only vertex information and no triangle mesh is provided. The PDs are computed via α-shape filtrations built on top of the point clouds. On average, the resulting PDs contain about 4000, 7500 and 1200 topological features of homological dimensions zero, one and two respectively.

As in Section 4.1, we perform clustering using the k-medoids algorithm. For PIs, we select a 20×20 grid and the default value of σ which is $1/2$ of the maximum persistence across all PDs divided by the grid size. VPBs are computed on a 12×12 grid. All the other

Figure 8: Toys used for SHREC17 dataset Limberger et al. (2017).

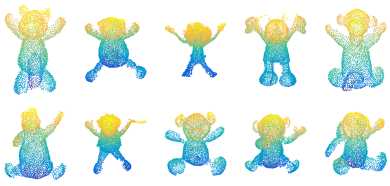

Figure 9: Sample point clouds from SHREC17 dataset. The points are colored based on the y and z coordinates Limberger et al. (2017).

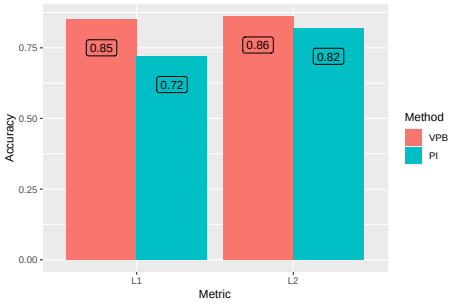

Figure 10: Clustering accuracy on SHREC17 dataset for H_1.

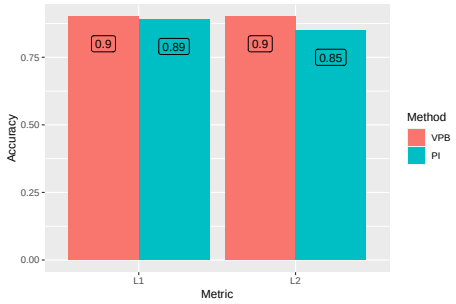

Figure 11: Clustering accuracy on SHREC17 dataset for H_2.

configurations carry over from the previous section. In particular, similar to the previous experiment, we perform clustering for each homological dimension. We find that VPBs and PIs are most discriminative for dimensions H_1 and H_2 and therefore we report our results only for these two dimensions. Figures 10 and 11 show the accuracies achieved by the two methods.

The 100×100 dissimilarity matrix used as an input for k-medoids, can also be viewed as a result of a 3D shape retrieval problem. In shape retrieval, as in clustering, no class information is allowed to be used to compute this matrix. We evaluate the retrieval performance using five standard statistics (Nearest Neighbor (NN), First-Tier (FT), Second-Tier (ST), E-Measure (E) and Discounted Cumulative Gain (DCG) Shilane et al. (2004)) computed from the dissimilarity matrix.

The selected dataset for our experiment was initially used for the 2017 Shape Retrieval Contest (SHREC) track: Point-Cloud Shape Retrieval of Non-Rigid Toys. This track invited researchers to contribute to the development of new discriminative and efficient descriptors of 3D non-rigid shapes. A total of eight groups participated in this content. Each group was allowed to submit up to 6 dissimilarity matrices. We compare our results with not only the performance of PIs but also with the best results of all participants.

We use L_1 metric to compute the dissimilarity matrices for VPBs and PIs in homological dimension H_2. For VPBs, $\tau = 0.1$ gives a slightly better result than that of $\tau = 0.3$ for which the Davies-Bouldin Index is minimum. From Table 2, we see that PIs slightly outperform

Method	NN	FT	ST	E	DCG
BoW-RoPS-DMF-3	1.0000	0.9778	0.9978	0.4390	0.9979
BPHAPT	0.9800	0.9111	0.9544	0.4273	0.9743
PI	0.9700	0.8989	0.9767	0.4376	0.9740
VPB	0.9800	0.8400	0.9567	0.4351	0.9573
MFLO-FV-IWKS	0.8900	0.7911	0.8589	0.4024	0.9038
CDSPF	0.9200	0.6744	0.8156	0.4005	0.8851
SnapNet	0.8800	0.6633	0.8011	0.3985	0.8663
AlphaVol1	0.7900	0.5878	0.7578	0.3980	0.8145
SQFD(WKS)	0.5400	0.3111	0.4467	0.2507	0.6032
m3DSH-2	0.4400	0.1867	0.2856	0.1932	0.4997

Table 2: Retrieval results on SHREC17 Shilane et al. (2004). The methods are ranked according to overall performance.

VPBs. BPHAPT and our method show comparable results - we have the same NN score and surpass in terms of ST and EE scores.

Furthermore, VPBs exhibit substantial computational advantages over both BoW-RoPS-DMF-3 and BPHAPT. For example, for BoW-RoPS-DMF-3 which utilizes the Bag-of-Words scheme, it takes around 30 minutes for the codebook module to run, whereas our method needs only about 48 seconds to compute all topological summaries (46 seconds for PDs and 2 seconds for VPBs). In terms of data preprocessing, BPHAPT has an extra step of constructing triangular meshes using the Poisson reconstruction method, while our approach does not preprocess the data and only uses the vertex information. Finally, our method has simpler complexity with regard to the number of tuning parameters involved - we only need to choose the parameter τ and the grid size.

4.3 Classification of parameter values of a discrete dynamical system

In this section, we explore the utility of VPBs in the context of classification and change point detection problems. These studies are also motivated by Adams et al. (2017) and involve data generated by the linked twist map defined by the dynamical system:

$$x_{n+1} = x_n + r \cdot y_n(1 - y_n) \bmod 1$$
$$y_{n+1} = y_n + r \cdot x_{n+1}(1 - x_{n+1}) \bmod 1,$$

where $r > 0$ Hertzsch et al. (2007). As in Adams et al. (2017), we consider values of r ranging in $\{2.0, 3.5, 4.0, 4.1, 4.3\}$, and take $\sigma = 0.005$ and a 20×20 grid for the computation of PIs.

For classification, we generate 100 truncated orbits $\{(x_n, y_n) : n = 0, 1, \ldots, 1000\}$ for each value of r, where the initial points (x_0, y_0) are selected at random from the unit square (see Figures 12, 14 and 15; a typical scatter plot for $r = 3.5$, not shown in the figures due to space limitations, does not exhibit any pattern and appears to be uniformly distributed over the unit square). The figures show that the scatter plots for $r = 4.0$ and $r = 4.1$ both have one hole in the center but they slightly differ in size and that scatter plots for the same

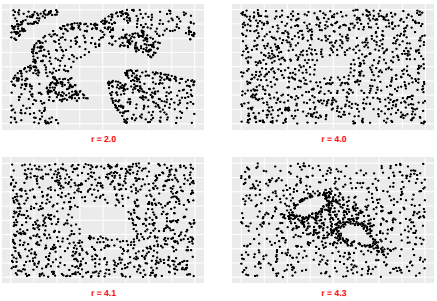

Figure 12: Scatter plots of sample truncated orbits corresponding to $r = 2.0, 4.0, 4.1, 4.3$ respectively.

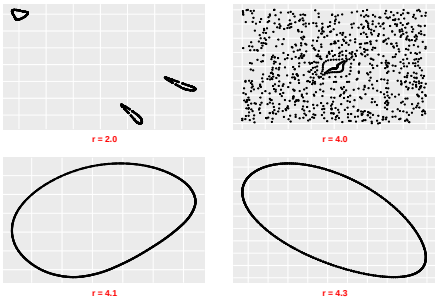

Figure 13: Scatter plots of atypical orbits corresponding to $r = 2.0, 4.0, 4.1, 4.3$ respectively.

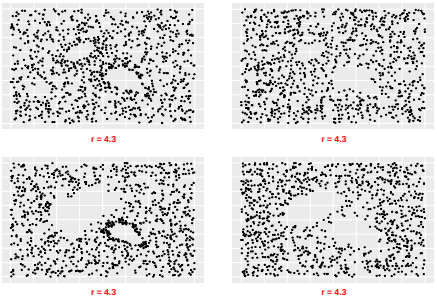

Figure 14: Scatter plots of sample truncated orbits corresponding to $r = 4.3$ and different (x_0, y).

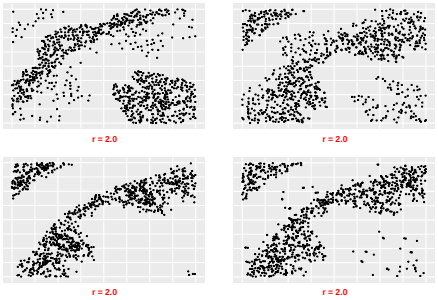

Figure 15: Scatter plots of sample truncated orbits corresponding to $r = 2.0$ and different (x_0, y).

value of r can show noticeable within-class variation in shape making the classification task more challenging.

Vietoris-Rips filtrations are used to compute PDs. To curb the computation cost to produce the PDs, we terminate the filtrations at scale value of 0.45. In other words, simplices with at least one pairwise distance greater than 0.45 do not enter any complex in the filtration. At this scale value, the holes in the point clouds (for $r = 4.0, 4.1, 4.3$) are already filled by 2-simplices (triangles). Going beyond the scale value of 0.45 does not essentially provide any additional useful topological information.

Interestingly, when generating the orbits for our simulation studies, we notice that for certain values of r, some initial points (x_0, y_0) lead to atypical-looking orbits (see Figure 13). Clearly, these orbits would mislead any classification method, so we devise a simple heuristic to omit such orbits from our analysis. We compute VPBs on a 7×7 grid and vary τ over $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. The optimal value of τ is selected by 10-fold cross-validation. We use standard machine learning methods such as Random Forest, Support Vector Machines (SVM), GLMNet and Model Averaged Neural Network to classify the orbits in the test set according to the associated parameter value of r.

We run the classifiers on the H_0 and H_1 features as well as when these features are concatenated. H_1-based test accuracies for 50/50 and 70/30 splits of the data, averaged over

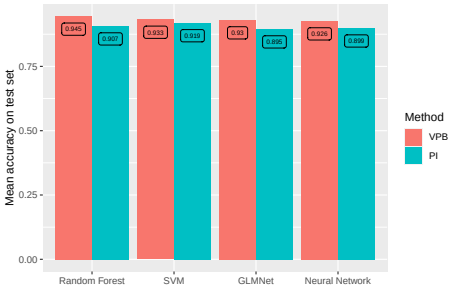

Figure 16: 50/50 split of data. VPBs and PIs are constructed from H_1 features.

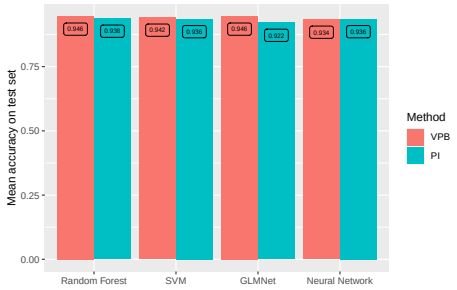

Figure 17: 70/30 split of data. VPBs and PIs are constructed from H_1 features.

	Change point #			
	1^{st}	2^{nd}	3^{rd}	4^{th}
VPB	0.0	43.6	7.4	33.5
PI	0.0	10.5	3.6	19.2

Table 3: Mean absolute errors for H_0.

	Change point #			
	1^{st}	2^{nd}	3^{rd}	4^{th}
VPB	0.0	0.5	1.7	1.0
PI	0.0	17.7	1.6	6.6

Table 4: Mean absolute errors for H_1.

10 independent trials, are given in Figures 16 and 17. When the H_0 and H_1 are concatenated, the accuracies come out to be only about 1% higher than those for H_1, while the H_0 features yield the lowest results (around 62% accuracy). From the figures we can see that, VPBs, when compared to PIs, on average produce slightly higher test accuracies across almost all the machine learning methods employed.

Next, we present a study on change point detection. Let $r_1 = 2$, $r_2 = 3.5$, $r_3 = 4.0$, $r_4 = 4.1$, $r_5 = 4.3$. In this study, we generate 50 truncated orbits for each value of r using the linked twist map and assume they are ordered in time so that the change points occur at times $t = 51, 101, 151, 201$: $X_{t,r} = \{(x_n(r), y_n(r)) : n = 0, 1, \ldots, 1000\}$, where $r = r_{\lceil t/50 \rceil}$, $t = 1, \ldots, 250$ and initial conditions (x_0, y_0) are selected at random. To estimate the true change points, we use a nonparametric multiple change point detection method called E-Divisive which determines statistically significant change points based on a permutation test Matteson and James (2014). Our choice is motivated by the fact that the distribution of the topological features remains unknown. We input $VPB_{t,r}$ and $PI_{t,r}$ associated with $X_{t,r}$ to the `e.divisive()` function from `ecp` package which outputs a list of estimated change points James and Matteson (2013). The change point results are given in Tables 3 and 4, which contain the mean absolute errors computed over 10 independent trials. In case the expected change point is not detected, we set the absolute error equal to 50 - the gap between any two consecutive true change points.

Finally, we examine the optimal values of the turning parameter τ that should have been selected based on the ground truth. We find that if there is little to no noise in the data, then $\tau = 0.3$ was optimal for homological dimensions H_1 and H_2 (when relevant) across all our simulations on clustering and classification. Moreover, we observe that all the classifiers identify $\tau = 0.3$ as optimal based on 10-fold cross-validation as well.

5. Conclusion and future work

The induced persistence block map presents a way to transform a PD into a vector in the Hilbert space $L^2(wdA)$ where the area measure wdA is induced by a continuous nonnegative weight function w. We establish the necessary and sufficient condition for the induced persistence block map to be continuous with respect to the p-Wasserstein distance. As a special case, when the persistence block map is induced by the constant function 1, we prove additional results on stability. A VPB is a finite dimensional vector summary of the image of the induced persistence block map for a given PD. Being vectors of $\mathbb{R}^n$, VPBs can easily be used as inputs within many machine learning frameworks. VPBs are faster to compute compared to other vector-based summaries such as persistence images and persistence landscapes. The parameters involved in the computation of VPBs are scale-free which makes their selection or tuning easier in practice. Through several simulation studies covering various learning task such as clustering, classification and change point detection, we have shown that VPBs produce improved performance results over persistence images in most cases.

In our simulation studies, VPBs are derived from persistence blocks induced by the identity function. In the future, we plan to consider a wider class of such functions and investigate the corresponding stability properties.

References

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. *The Journal of Machine Learning Research*, 18(1):218–252, 2017.

Nieves Atienza, Rocío González-Díaz, and Manuel Soriano-Trigueros. On the stability of persistent entropy and new summary functions for topological data analysis. *Pattern Recognition*, 107:107509, 2020.

Eric Berry, Yen-Chi Chen, Jessi Cisewski-Kehe, and Brittany Terese Fasy. Functional summaries of persistence diagrams. *Journal of Applied and Computational Topology*, 4(2): 211–262, 2020.

Peter Bubenik. Statistical topological data analysis using persistence landscapes. *The Journal of Machine Learning Research*, 16(1):77–102, 2015.

Peter Bubenik and Tane Vergili. Topological spaces of persistence modules and their properties. *Journal of Applied and Computational Topology*, 2(3):233–269, 2018.

Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2): 255–308, 2009.

Frédéric Chazal, Vin De Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, 2014.

Yen-Chi Chen, Daren Wang, Alessandro Rinaldo, and Larry Wasserman. Statistical analysis of persistence intensity functions. *arXiv preprint arXiv:1510.02502*, 2015.

Yu-Min Chung and Austin Lawson. Persistence curves: A canonical framework for summarizing persistence diagrams. *arXiv preprint arXiv:1904.07768*, 2019.

David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & computational geometry*, 37(1):103–120, 2007.

David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have l p-stable persistence. *Foundations of computational mathematics*, 10(2): 127–139, 2010.

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

Herbert Edelsbrunner and Ernst P Mücke. Three-dimensional alpha shapes. *ACM Transactions on Graphics (TOG)*, 13(1):43–72, 1994.

Herbert Edelsbrunner, John Harer, et al. Persistent homology-a survey. *Contemporary mathematics*, 453:257–282, 2008.

Brittany T Fasy, Jisu Kim, Fabrizio Lecci, Clement Maria, David L Millman, Vincent Rouvreau, and Maintainer Jisu Kim. Package 'tda', 2021.

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

Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis. Clustering validity checking methods. *ACM Sigmod Record*, 31(3):19–27, 2002.

Jan-Martin Hertzsch, Rob Sturman, and Stephen Wiggins. Dna microarrays: design principles for maximizing ergodic, chaotic mixing. *Small*, 3(2):202–218, 2007.

Umar Islambekov, Monisha Yuvaraj, and Yulia R Gel. Harnessing the power of topological data analysis to detect change points. *Environmetrics*, 31(1):e2612, 2020.

Nicholas A James and David S Matteson. ecp: An r package for nonparametric multiple change point analysis of multivariate data. *arXiv preprint arXiv:1309.3295*, 2013.

Tomasz Kaczynski, Konstantin Mischaikow, and Marian Mrozek. *Computational homology*, volume 157. Springer Science & Business Media, 2006.

Genki Kusano, Yasuaki Hiraoka, and Kenji Fukumizu. Persistence weighted gaussian kernel for topological data analysis. In *International Conference on Machine Learning*, pages 2004–2013. PMLR, 2016.

Chunyuan Li, Maks Ovsjanikov, and Frederic Chazal. Persistence-based structural recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1995–2002, 2014.

Frederico A Limberger, Richard C Wilson, M Aono, N Audebert, A Boulch, B Bustos, A Giachetti, A Godil, B Le Saux, B Li, et al. Shrec'17 track: Point-cloud shape retrieval of non-rigid toys. In *10th Eurographics workshop on 3D Object retrieval*, pages 1–11, 2017.

Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

Martin Maechler, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. *cluster: Cluster Analysis Basics and Extensions*, 2021. URL `https://CRAN.R-project.org/package=cluster`. R package version 2.1.2 — For new features, see the 'Changelog' file (in the package source).

David S Matteson and Nicholas A James. A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109 (505):334–345, 2014.

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

Tullia Padellini, Francesco Palini, Pierpaolo Brutti, Chih-Chung Chang, Chih-Chen Lin, Michael Kerber, Dmitriy Morozov, Arnur Nigmetov, and Maintainer Tullia Padellini. Statistical learning with kernel for persistence diagrams, 2020.

Hae-Sang Park and Chi-Hyuck Jun. A simple and fast algorithm for k-medoids clustering. *Expert systems with applications*, 36(2):3336–3341, 2009.

LKPJ Rdusseeun and P Kaufman. Clustering by means of medoids. In *Proceedings of the Statistical Data Analysis Based on the L1 Norm Conference, Neuchatel, Switzerland*, pages 405–416, 1987.

Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4741–4748, 2015.

Eitan Richardson and Michael Werman. Efficient classification using the euler characteristic. *Pattern Recognition Letters*, 49:99–106, 2014.

Bastian Rieck, Filip Sadlo, and Heike Leitte. Topological machine learning with persistence indicator functions. In *Topological Methods in Data Analysis and Visualization*, pages 87–101. Springer, 2017.

Joseph J Rotman. *An introduction to algebraic topology*, volume 119. Springer Science & Business Media, 2013.

Walter Rudin. *Functional Analysis, Second Edition*. International Series in Pure and Applied Mathematics. McGraw-Hill, 1991. ISBN 0070542368.

Philip Shilane, Patrick Min, Michael Kazhdan, and Thomas Funkhouser. The princeton shape benchmark. In *Proceedings Shape Modeling Applications, 2004.*, pages 167–178. IEEE, 2004.

Raoul R Wadhwa, Drew FK Williamson, Andrew Dhawan, and Jacob G Scott. Tdastats: R pipeline for computing persistent homology in topological data analysis. *Journal of open source software*, 3(28):860, 2018.

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