Journal of Computational Mathematics and Data Science 4 (2022) 100050

Contents lists available at ScienceDirect

Journal of Computational Mathematics and Data Science

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

Variably Scaled Persistence Kernels (VSPKs) for persistent homology applications

Stefano De Marchi [a], Federico Lot [b], Francesco Marchetti [a,*], Davide Poggiali [c]

[a] Dipartimento di Matematica "Tullio Levi-Civita", Università di Padova, 35121 Padua, Italy
[b] Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, 35032 Marburg, Germany
[c] FAR Networks srl, 20063 Cernusco S.N., Milan, Italy

ARTICLE INFO

Keywords:
Kernel-based learning
Variably scaled persistence kernel
Persistence diagrams
Persistent homology

ABSTRACT

In recent years, various kernels have been proposed in the context of *persistent homology* to deal with *persistence diagrams* in supervised learning approaches. In this paper, we consider the idea of variably scaled kernels, for approximating functions and data, and we interpret it in the framework of persistent homology. We call them *Variably Scaled Persistence Kernels (VSPKs)*. These new kernels are then tested in different classification experiments. The obtained results show that they can improve the performance and the efficiency of existing standard kernels.

1. Introduction

Let $\Omega \subset \mathbb{R}^v$ and let $\mathcal{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\} \subset \Omega$ be a set of input data, $v, n \in \mathbb{N}$. In the classification setting, each element $\boldsymbol{x}_i$ of the dataset, $i = 1, \dots, n$, is uniquely associated to a *label* (or *class*) y_i, which is an element of a certain set of labels $\mathcal{T} = \{c_1, \dots, c_t\}$, $t \in \mathbb{N}$. In principle, the labels can be real numbers, but they can also represent categorical concepts to be learned. Then, we can define the dataset $\mathcal{Z} = \{(\boldsymbol{x}_i, y_i) \mid \boldsymbol{x}_i \in \mathcal{X},\ y_i \in \mathcal{T}\}$. The supervised learning task consists in finding a decision function $s : \Omega \longrightarrow \mathcal{T}$ such that:

(i) it models the input–output relation in $\mathcal{Z}$;
(ii) letting $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_m \in \Omega \setminus \mathcal{X}$ be unseen instances, $m \in \mathbb{N}$, with associated labels $\iota_i \in \mathcal{T}$, s models the input–output relation $\{(\boldsymbol{\xi}_i, \iota_i)\}_{i=1,\dots,m} \subset (\Omega \setminus \mathcal{X}) \times \mathcal{T}$, $m \geq 1$.

The *generalisation* capability required in (ii) is fundamental in the learning problem, since it is trivial to find a decision function that satisfies (i); for an introduction concerning (statistical) learning theory, refer to e.g. [1, §1.3] and [2].

Kernel methods are well-established tools in supervised machine learning, as well as in a variety of research and applied fields [3,4]. The flexibility provided by kernel-based schemes allows the handling of different kinds of possible structured data, e.g., graphs and words, which are encoded in some dot product space where even complex patterns may be distinguished [5–8].

In the following, our data consist of *persistence diagrams*, which represent an output of the so-called *persistent homology*, in the framework of Topological Data Analysis (TDA). The usage of topological methods and analysis techniques to extract significant features and patterns from data is receiving more and more interest; for a complete overview concerning TDA, see e.g. [9–11]. In the last years, several kernels specifically devoted to deal with the peculiar structure of persistence diagrams have been proposed [12–15], therefore the construction of suitable kernels is a very active research line in this context.

Here, we introduce what we call the Variably Scaled Persistence Kernels (VSPKs), which are inspired by the variably scaled kernels introduced in the context of approximation theory in [16] and employed as a feature augmentation strategy in [17,18]

* Corresponding author.
E-mail addresses: demarchi@math.unipd.it (S. De Marchi), lotf@mathematik.uni-marburg.de (F. Lot), francesco.marchetti@math.unipd.it (F. Marchetti), dr.davide.poggiali@gmail.com (D. Poggiali).

https://doi.org/10.1016/j.jcmds.2022.100050
Received 20 February 2022; Received in revised form 13 May 2022; Accepted 16 May 2022

for kernel-based learning. Indeed, our aim is to build a bridge between classical variably scaled kernels and kernels defined in the context of persistent homology. After defining VSPKs, we design some scaling functions and we test the performance of the resulting kernels by experimenting with three different datasets. The obtained results show that the variably scaled setting may yield to better classification outcomes and efficiency with respect to the standard setting, and thus can be considered for further investigations and applications involving persistence diagrams.

The paper is organised as follows. In Sections 2 and 3, we introduce kernel-based learning with Support Vector Machines (SVMs) and we recall some notions concerning persistent homology, also presenting some kernels of recent introduction that are dedicated to persistence diagrams. VSPKs are proposed and analysed in Section 4, and related numerical experiments are exposed in Section 5. Finally, in Section 6 we present some conclusions and final remarks.

2. Positive definite and variably scaled kernels

Let $\kappa : \Omega \times \Omega \longrightarrow \mathbb{R}$ be a kernel. Given a set of data $\mathcal{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\} \subset \Omega$, the $n \times n$ matrix K with elements $K_{ij} = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $i, j = 1, \dots, n$, is the *Gram matrix* of the kernel κ with respect to $\mathcal{X}$. If κ is positive definite (strictly positive definite) on $\Omega \times \Omega$, i.e., K is positive semi-definite (definite) for all possible datasets in Ω, then it is possible to decompose the kernel according to Mercer theorem [19], and to interpret such decomposition as an inner product in a *Reproducing Kernel Hilbert Space* (RKHS) $\mathcal{F}$. Indeed, there exists a (non-unique) *feature map* $\boldsymbol{\Phi} : \Omega \longrightarrow \mathcal{F}$ such that

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{\Phi}(\boldsymbol{x}), \boldsymbol{\Phi}(\boldsymbol{y}) \rangle_{\mathcal{F}} \qquad \boldsymbol{x}, \boldsymbol{y} \in \Omega,$$

being $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ the bilinear form related to the RKHS (e.g. $\boldsymbol{\Phi}(\boldsymbol{x}) = \kappa(\cdot, \boldsymbol{x})$). Moreover, the kernel κ induces a distance d_κ on Ω

$$d_\kappa(\boldsymbol{x}, \boldsymbol{y}) = \kappa(\boldsymbol{x}, \boldsymbol{x}) + \kappa(\boldsymbol{y}, \boldsymbol{y}) - 2\kappa(\boldsymbol{x}, \boldsymbol{y}). \tag{1}$$

Variably Scaled Kernels (VSKs) have been introduced in [16] in the context of kernel-based approximation, with the aim of overcoming instability issues. Then, they have been extended to work in a more general setting in [17], as presented in the following form. Let $\Lambda \subseteq \mathbb{R}^\nu$, $\nu > 0 \in \mathbb{N}$ and let $\kappa : \tilde{\Omega} \times \tilde{\Omega} \longrightarrow \mathbb{R}$ be a continuous (strictly) positive definite kernel, where $\tilde{\Omega} = \Omega \times \Lambda \subseteq \mathbb{R}^{\upsilon+\nu}$. Given a *scaling* function $\psi : \Omega \longrightarrow \Lambda$, and by defining $\Psi(\boldsymbol{x}) = (\boldsymbol{x}, \psi(\boldsymbol{x}))$, a VSK $\kappa^\Psi : \Omega \times \Omega \longrightarrow \mathbb{R}$ is defined as

$$\kappa^\Psi(\boldsymbol{x}, \boldsymbol{y}) = \kappa((\boldsymbol{x}, \psi(\boldsymbol{x})), (\boldsymbol{y}, \psi(\boldsymbol{y}))) \tag{2}$$

for $\boldsymbol{x}, \boldsymbol{y} \in \Omega$. The function Ψ can be interpreted as a *feature augmentation* map, which adds ν coordinates (features) to the original sample. In this view, the VSK setting has been analysed in [17] as a *stacking* technique, which is capable of enhancing the prediction performances of classical kernel-based classifiers such as, e.g., SVMs.

Letting $\boldsymbol{x} = (x_1, \dots, x_d)^\intercal \in \Omega$, we recall that a binary (i.e. $\mathcal{T} = \{-1, +1\}$) SVMs classifier is characterised by the decision function

$$s(\boldsymbol{x}) = \operatorname{sign}(h(\boldsymbol{x})) = \operatorname{sign}(\langle \boldsymbol{\Phi}(\boldsymbol{x}), \boldsymbol{w} \rangle_{\mathcal{F}} + b),$$

where

$$\boldsymbol{w} = \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{\Phi}(\boldsymbol{x}_i) \in \mathcal{F}.$$

The coefficients $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ are the solution of the following *soft margin* problem [20, §18, p. 346–347]

$$\begin{cases} \min_{\boldsymbol{\alpha} \in \mathbb{R}^n} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) - \sum_{i=1}^n \alpha_i, \\ \text{s.t. } \sum_{i=1}^n \alpha_i y_i = 0, \\ 0 \le \alpha_i \le \zeta, \quad i = 1, \dots, n, \end{cases}$$

where $[0, \zeta]^n$ is the *bounding box*, with $\zeta \in [0, +\infty)$. Usually, a binary SVMs classifier is extended to the multiclass setting by considering a *one-vs-rest* approach.

The strength of kernel-based classifiers as SVMs mainly relies on the flexibility that concerns the construction of the kernel. Indeed, any type of structured data may potentially be *translated* into a certain Hilbert space (e.g., the classical $\mathbb{R}^n$) by a properly defined feature map. This is the case of *persistence diagrams*, which are useful tools in TDA characterised by a peculiar structure, as we outline in the next section.

3. Persistent homology and kernels

3.1. Basics on persistent homology

In the following, we recall some basic ideas about some tools of persistent homology. For a more detailed treatment, especially concerning the algebraic aspects of the construction, we refer e.g. to [10,21].

Data, manifolds and simplicial complexes. Let our data $\mathcal{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\} \subset \Omega$ be interpreted as a set of vertices sampled from some manifold M, and suppose that we wish to highlight some of its intrinsic homological properties. Intuitively, such properties provide important geometrical information on $\mathcal{X}$ by describing the number of holes and cavities that can be related to the data distribution. In order to do this, a concrete possibility consists in studying the geometrical features of a manifold built upon $\mathcal{X}$ and that can be considered as an approximation of M. Precisely, letting $\varepsilon > 0$, we study

$$M^{\mathcal{X},\varepsilon} = \bigcup_{i=1}^{n} B(\boldsymbol{x}_i, \varepsilon),$$

where $B(\boldsymbol{x}_i, \varepsilon)$ is the ball of radius ε and centre $\boldsymbol{x}_i$. The next step is to associate to $M^{\mathcal{X},\varepsilon}$ an abstract simplicial complex, which is formally defined as a family of sets that is closed with respect to the *subset operator*. In a sense, it can be considered as an enriched graph structure, where not only points and edges are included, but also triangles (i.e. formal combinations of three points in two-dimensional), tetrahedra (i.e. formal combinations of four points in three-dimensional), and so on. Moreover, the concept of *face* plays a central role. For example, an edge is characterised by two zero-dimensional faces (two extremal points), a triangle by three one-dimensional faces (three side edges), *et cetera*. Simplicial complexes can be defined by using different construction rules, which determine to what extent higher dimensional structures are built from low dimensional ones. Here, we associate to $M^{\mathcal{X},\varepsilon}$ a *Vietoris–Rips simplicial complex* $K^{\mathcal{X},\varepsilon}$, where two distinct vertices $\boldsymbol{x}_i, \boldsymbol{x}_j$ are connected by an edge if and only if $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \leq \varepsilon$, and we consider r-dimensional elements that are determined by $r+1$ connected $(r-1)$-dimensional faces as long as $r \leq v$. In Fig. 1 (top left/right and bottom left), we show a two-dimensional example ($v = 2$) where different numbers of edges and triangles are generated depending on the chosen ε.

The homology group. Formal linear combinations of r-dimensional faces in $K^{\mathcal{X},\varepsilon}$ form the *r-chains* algebraic group $C_r^{\mathcal{X},\varepsilon}$. Moreover, letting $[\boldsymbol{x}_{i_0}, \dots, \boldsymbol{x}_{i_r}]$ be the face constructed upon distinct vertices $\boldsymbol{x}_{i_0}, \dots, \boldsymbol{x}_{i_r}$ in $\mathcal{X}$, $i_0, \dots, i_r \in \{1, \dots, n\}$, we define the linear *boundary operator* $\partial_r : C_r^{\mathcal{X},\varepsilon} \longrightarrow C_{r-1}^{\mathcal{X},\varepsilon}$ as

$$\partial_r[\boldsymbol{x}_{i_0}, \dots, \boldsymbol{x}_{i_r}] = \sum_{j=0}^{r} (-1)^r [\boldsymbol{x}_{i_0}, \dots, \boldsymbol{x}_{i_{j-1}}, \boldsymbol{x}_{i_{j+1}}, \dots, \boldsymbol{x}_{i_r}].$$

The *r-cycles* and *r-boundaries* groups are defined as $Z_r^{\mathcal{X},\varepsilon} = \ker \partial_r$ and $B_r^{\mathcal{X},\varepsilon} = \mathrm{im}\partial_{r+1}$, respectively. Furthermore, it can be proved that the rank of the *r-homology* group $H_r^{\mathcal{X},\varepsilon} = Z_r^{\mathcal{X},\varepsilon} / B_r^{\mathcal{X},\varepsilon}$ expresses the concept of *r-dimensional holes* in $K^{\mathcal{X},\varepsilon}$, thus relatable to M and $\mathcal{X}$, which is the sought geometrical characterisation.

Persistent homology. Unfortunately, finding an optimal $\varepsilon^\star$ that represents the intrinsic geometric properties of M is a tough and unstable process. This is evident in the example in Fig. 1 (top left/right and bottom left), where different values of ε determine a completely different simplicial structure. To solve this issue, one can drop the idea of selecting a good parameter and analyse the whole *filtration* $\{M^{\mathcal{X},\varepsilon} \mid \varepsilon > 0\}$. In particular, letting $\varepsilon_1 < \cdots < \varepsilon_u$ be increasing real numbers, the nested sequence $K^{\mathcal{X},\varepsilon_1} \subseteq \cdots \subseteq K^{\mathcal{X},\varepsilon_u}$ is obtained. Then, for $r \geq 0$ and $i \in \{1, \dots, u\}$, we can associate to this filtration the *ℓ-persistent homology* group

$$H_{r,\ell}^{\mathcal{X},\varepsilon_i} = Z_r^{\mathcal{X},\varepsilon_i} / (Z_r^{\mathcal{X},\varepsilon_i} \cap B_r^{\mathcal{X},\varepsilon_{i+\ell}}).$$

The group $H_{r,\ell}^{\mathcal{X},\varepsilon_i}$ contains the homology classes that persist in the *time interval* $[i, i+\ell]$, i.e., they are *born* with $K^{\mathcal{X},\varepsilon_j}$ for some $j < i$ and they are *alive* with $K^{\mathcal{X},\varepsilon_{i+\ell}}$. We point out that such homology classes might persist *indefinitely* (we will denote this case as an ∞ level), or they might *die* with a certain ε_j, $i < j \leq u$. In other words, studying persistent homology groups provides information on holes and cavities in data along with their persistence: the more an element is persistent, the more is likely to be descriptive of the data structure.

Persistence diagrams. Consequently, each element of the persistent homology groups obtained by considering the whole filtration can be represented by a birth–death pair $(b, d) \in \mathbb{R}^2_+$, $b = \varepsilon_h$, $d = \varepsilon_k$ for some $h \in \{1, \dots, m\}$, $k \in \{1, \dots, m\} \cup \{\infty\}$, $h < k$. We say that $d - b$ is the *persistence* of (b, d), and that a birth–death pair is r-dimensional if it is related to r-dimensional homology groups. Moreover, letting $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_u)$, we define a *persistence diagram* $D_r(\mathcal{X}, \boldsymbol{\varepsilon})$ related to the filtration $K^{\mathcal{X},\varepsilon_1} \subseteq \cdots \subseteq K^{\mathcal{X},\varepsilon_m}$ as

$$D_r(\mathcal{X}, \boldsymbol{\varepsilon}) = \{(b, d) \mid (b, d) \in P_r(\mathcal{X}, \boldsymbol{\varepsilon})\} \cup \mathcal{B}, \quad r \geq 0, \tag{3}$$

where $P_r(\mathcal{X}, \boldsymbol{\varepsilon})$ denotes the set of r-dimensional birth–death pairs obtained with the filtration and $\mathcal{B} = \{(z, z) \mid z \geq 0\}$. We remark that $D_r(\mathcal{X}, \boldsymbol{\varepsilon})$ is a multiset, since a couple (b, d) might appear more than once, i.e., might have multiplicity greater than one. Furthermore, the bisector $\mathcal{B}$ is composed by an infinite number of elements characterised by infinite multiplicity, and it is added in order to achieve some *uniformity* among different persistence diagrams and facilitate the formulation of proper metrics, as we present below. In Fig. 1 (bottom right), we display the result of the discussed analysis with the example dataset. We observe that such analysis captures some intrinsic geometrical properties of $\mathcal{X}$. In particular, the unique most persistent zero-dimensional pair, which is significantly far from the others, suggests that a unique connected component underlies $\mathcal{X}$. Moreover, the persistent one-dimensional pair indicates the presence of a one-dimensional hole in the structure of the dataset, which is highlighted by the simplicial complex depicted in the bottom left figure.

Persistence diagrams show some stability properties with respect to perturbations of the involved dataset [22]. To better clarify this aspect, let us recall some useful metrics. Let $\mathcal{X}, \mathcal{Y} \subset \Omega$ be two non-empty datasets. The *Hausdorff distance* is defined as

$$d_H(\mathcal{X}, \mathcal{Y}) = \max\left\{ \sup_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} \|x - y\|_\infty, \sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \|x - y\|_\infty \right\}.$$

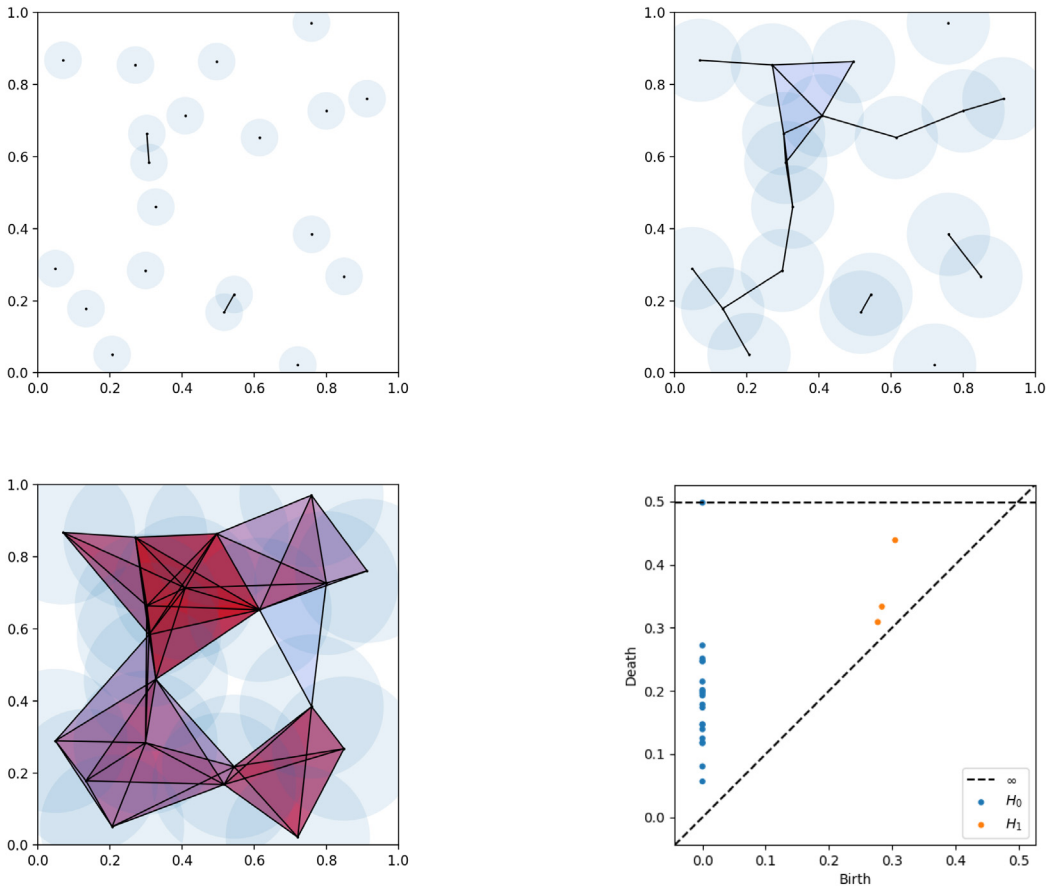

Fig. 1. Given a set of points $\mathcal{X}$, the construction of the Vietoris–Rips complex along the filtration is depicted on top (left: $\varepsilon = 0.1$; right: $\varepsilon = 0.22$) and bottom left ($\varepsilon = 0.4$) figures. The persistence diagrams $D_0(\mathcal{X}, \varepsilon)$ and $D_1(\mathcal{X}, \varepsilon)$ are overlapped in the bottom right figure.

Moreover, letting $D_r(\mathcal{X}, \varepsilon)$ and $D_r(\mathcal{Y}, \varepsilon)$ be persistence diagrams for some $r \geq 0$ and filtration radii vector $\boldsymbol{\varepsilon}$, we recall the *p-Wasserstein* distance

$$d_{W,p}(D_r(\mathcal{X}, \varepsilon), D_r(\mathcal{Y}, \varepsilon)) = \left(\inf_{\gamma \in \Gamma} \sum_{x \in D_r(\mathcal{X}, \varepsilon)} \|x - \gamma(x)\|_\infty^p \right)^{\frac{1}{p}},$$

where $\Gamma = \{\gamma : D_r(\mathcal{X}, \varepsilon) \longrightarrow D_r(\mathcal{Y}, \varepsilon) \mid \gamma \text{ is a bijection}\}$. In particular, letting $p \to \infty$, we obtain the *bottleneck distance*

$$d_{W,\infty}(D_r(\mathcal{X}, \varepsilon), D_r(\mathcal{Y}, \varepsilon)) = d_B(D_r(\mathcal{X}, \varepsilon), D_r(\mathcal{Y}, \varepsilon)) = \inf_{\gamma \in \Gamma} \sup_{x \in D_r(\mathcal{X}, \varepsilon)} \|x - \gamma(x)\|_\infty.$$

We have the stability result [13]

$$d_B(D_r(\mathcal{X}, \varepsilon), D_r(\mathcal{Y}, \varepsilon)) \leq d_H(\mathcal{X}, \mathcal{Y}),$$

i.e., the bottleneck distance between persistence diagrams is controlled as long as the underlying datasets are close in the Hausdorff metric.

3.2. Kernels for persistence diagrams

In order to better measure similarities between persistence diagrams, various positive definite kernels that are suitable for dealing with the peculiar structure of persistence diagrams have been introduced and studied in the recent literature. In our applications, we will consider the following ones. Moreover, sometimes we will denote as $D_1 = D_r(\mathcal{X}, \varepsilon)$ and $D_2 = D_r(\mathcal{Y}, \varepsilon)$ to simplify the notation.

- The **Persistence Scale Space (PSS) kernel** [15]

$$\kappa_\sigma(D_1, D_2) = \frac{1}{8\pi\sigma} \sum_{\substack{y \in D_1 \\ z \in D_2}} \exp\left(-\frac{\|y - z\|^2}{8\sigma}\right) - \exp\left(-\frac{\|y - \overline{z}\|^2}{8\sigma}\right), \tag{4}$$

where $\sigma > 0$ and $\overline{z} = (d, b)$ if $z = (b, d)$. It is 1-Wasserstein stable, i.e.:

$$\|\Phi_\sigma(D_1) - \Phi_\sigma(D_2)\|_{L_2(\Omega)} \leq \frac{1}{2\sqrt{\pi\sigma}} d_{W,1}(D_1, D_2),$$

where Φ_σ is the feature map related to κ_σ that is associated to the solution of the heat equation (see [15, Definition 1]).

- The **Persistence Weighted Gaussian (PWG) kernel** [13]

$$\kappa_G(D_1, D_2; \kappa, \omega) = \exp\left(-\frac{1}{2\tau^2}\|E_\kappa(\mu^\omega_{D_1}) - E_\kappa(\mu^\omega_{D_2})\|^2_{\mathcal{F}}\right) \quad \tau > 0, \tag{5}$$

which is built upon a *standard* Gaussian kernel κ and a weight function ω, where

$$E_\kappa(\mu^\omega_{D_1}) = \sum_{x \in D_1} \omega(x)\, \kappa(\cdot, x).$$

It is both 1-Wasserstein and bottleneck stable, if we choose as weight function

$$\omega_{\text{arc}}(x) = \arctan(C(d-b)^\delta) \quad x = (b, d),\ C > 0,\ \delta \in \mathbb{Z}_{>0}.$$

Indeed, there exist $\delta \in \mathbb{Z}_{>0}$ and $L > 0$ such that

$$\|E_\kappa(\mu^{\omega_{\text{arc}}}_{D_1}) - E_\kappa(\mu^{\omega_{\text{arc}}}_{D_2})\| \leq L d_B(D_1, D_2). \tag{6}$$

- The **Sliced Wasserstein (SW) kernel** [12]

$$\kappa_{SW}(D_1, D_2) = \exp\left(-\frac{SW(D_1, D_2)}{2\chi^2}\right). \tag{7}$$

which is based on the so called Sliced Wasserstein distance, which is equivalent to the 1-Wasserstein distance, i.e.,

$$\frac{1}{2M} d_{W,1}(D_1, D_2) \leq d_{SW}(D_1, D_2) \leq 2\sqrt{2}\, d_{W,1}(D_1, D_2)$$

for some positive constant M.

Remark 1. We observe that the PWG and SW kernels are in the form

$$\kappa(D_1, D_2) = \exp\left(-\beta d(D_1, D_2)\right)$$

for some $\beta > 0$, where $d(\cdot, \cdot)$ is the distance induced by the underlying metric in the case of the SW (for more details, we refer to the cited seminal papers), while for the PWG kernel the distance is induced by the kernel (see (1)).

4. Variably scaled persistence kernels

In the following, our purpose is to interpret the idea underlying VSKs in the context of persistent homology. The main difference between standard kernels and kernels for persistence diagrams is the structure of the input data. Since persistent diagrams consists of a collection of topological features, i.e. birth–death couples in $\mathbb{R}^2_{\geq \mathcal{B}} = \{x = (b, d) \in \mathbb{R}^2_+ \mid d \geq b\}$, introducing a scaling function whose output lies outside $\mathbb{R}^2_{\geq \mathcal{B}}$ would be meaningless. Hence, letting $\mathfrak{D}_r(\varepsilon) = \{D_r(\mathcal{X}, \varepsilon) \,|\, \mathcal{X} \subset \Omega\}$, we propose the following definition.

Definition 1. Let $\kappa : \mathfrak{D}_r(\varepsilon) \times \mathfrak{D}_r(\varepsilon) \longrightarrow \mathbb{R}$ be a kernel for persistence diagrams and let $\Psi : \mathfrak{D}_r(\varepsilon) \to \mathfrak{D}_r(\varepsilon)$. A variably scaled persistence kernel κ^Ψ on $\mathfrak{D}_r(\varepsilon) \times \mathfrak{D}_r(\varepsilon)$ is defined as

$$\kappa^\Psi(D_r(\mathcal{X}, \varepsilon), D_r(\mathcal{X}, \varepsilon)) = \kappa(\Psi(D_r(\mathcal{X}, \varepsilon)), \Psi(D_r(\mathcal{Y}, \varepsilon)))$$

for $D_r(\mathcal{X}, \varepsilon), D_r(\mathcal{Y}, \varepsilon) \in \mathfrak{D}_r(\varepsilon)$.

As in other contexts, a proper function Ψ needs to be designed. We will construct it upon an auxiliary function $\psi : \mathfrak{D}_r(\varepsilon) \longrightarrow \mathbb{R}^2_{\geq \mathcal{B}}$, whose design is deepened in Section 4.3. In the next subsections, we propose two different structures for Ψ, which will be denoted as Ψ_a and Ψ_ρ.

4.1. The feature augmentation map: Ψ_a

We define

$$\Psi_a(D_r(\mathcal{X}, \varepsilon)) = D_r(\mathcal{X}, \varepsilon) \cup \psi(D_r(\mathcal{X}, \varepsilon)).$$

The map Ψ_a plays the role of a *feature augmenting* map, since an additional generator is added in the persistence diagram.

We have the following stability results for the VSPK framework in terms of the classical setting.

Proposition 1. *Let κ_σ be the PSS kernel defined in (4) and let D_1, D_2 be r-dimensional persistence diagrams for some $r \geq 0$. We have*

$$\|\Phi^{\Psi_a}_\sigma(D_1) - \Phi^{\Psi_a}_\sigma(D_2)\|_{L_2(\Omega)} \leq \frac{1}{2\sqrt{\pi\sigma}}\left(d_{W,1}(D_1, D_2) + \|\psi(D_1) - \psi(D_2)\|_\infty\right).$$

Proof. By definition, for any $\delta > d_{W,1}(D_1, D_2)$ there exists a bijection γ between D_1 and D_2 such that $\sum_{\boldsymbol{x}\in D_1} \|\boldsymbol{x} - \gamma(\boldsymbol{x})\|_\infty \leq \delta$. Then, by denoting $u_{\boldsymbol{y}}(\boldsymbol{x}) = \frac{1}{4\pi\sigma}\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{4\sigma}\right)$ and imposing the extension $\gamma(\psi(D_1)) = \psi(D_2)$, we obtain

$$\begin{aligned}\|\boldsymbol{\Phi}_\sigma^{\Psi_a}(D_1) - \boldsymbol{\Phi}_\sigma^{\Psi_a}(D_2)\|_{L_2(\Omega)} &= \|\boldsymbol{\Phi}_\sigma(D_1 \cup \psi(D_1)) - \boldsymbol{\Phi}_\sigma(D_2 \cup \psi(D_2))\| \\ &= \left\| \sum_{\boldsymbol{y}\in D_1\cup\psi(D_1)} (u_{\boldsymbol{y}} - u_{\overline{\boldsymbol{y}}}) - (u_{\gamma(\boldsymbol{y})} - u_{\overline{\gamma(\boldsymbol{y})}}) \right\|_{L_2(\Omega)} \\ &= \|\boldsymbol{\Phi}_\sigma(D_1) - \boldsymbol{\Phi}_\sigma(D_2) + \boldsymbol{\Phi}_\sigma(\psi(D_1)) - \boldsymbol{\Phi}_\sigma(\psi(D_2))\|_{L_2(\Omega)}.\end{aligned}$$

Therefore, by following the proof of [15, Theorem 2] and by observing that $d_{W,1}(\psi(D_1), \psi(D_2)) = \|\psi(D_1) - \psi(D_2)\|_\infty$, the triangular inequality leads to

$$\|\boldsymbol{\Phi}_\sigma^{\Psi_a}(D_1) - \boldsymbol{\Phi}_\sigma^{\Psi_a}(D_2)\|_{L_2(\Omega)} \leq \frac{1}{2\sqrt{\pi\sigma}}(\delta + \|\psi(D_1) - \psi(D_2)\|_\infty)$$

and the proof is concluded. □

Proposition 2. *Let $\kappa_G(\cdot,\cdot,\kappa,\omega_{\mathrm{arc}})$ be the PWG kernel defined in (5), where $\kappa(\boldsymbol{x},\boldsymbol{y}) = \exp(-\|\boldsymbol{x}-\boldsymbol{y}\|^2/(2\eta^2))$, and let D_1, D_2 be r-dimensional persistence diagrams for some $r \geq 0$. By denoting as $E_\kappa^{\Psi_a}(\mu_{D_1}^{\omega_{arc}}) = E_\kappa(\mu_{D_1\cup\psi(D_1)}^{\omega_{arc}})$, we have*

$$\|E_\kappa^{\Psi_a}(\mu_{D_1}^{\omega_{arc}}) - E_\kappa^{\Psi_a}(\mu_{D_2}^{\omega_{arc}})\| \leq L d_B(D_1, D_2) + L^\psi(D_1, D_2)$$

with $L^\psi(D_1, D_2) = 2\omega_{arc}(\psi(D_1))\left(1 - \exp\left(-\frac{\|\psi(D_1)-\psi(D_2)\|^2}{2\eta^2}\right)\right) + |\omega_{arc}(\psi(D_1)) - \omega_{arc}(\psi(D_2))|$ *and L as in (6).*

Proof. We can follow the proof of [23, Theorem D.1] and the assumptions therein. As in Proposition 1, we impose the extension $\gamma(\psi(D_1)) = \psi(D_2)$ for the bijection between D_1 and D_2, obtaining then

$$\begin{aligned}\|E_\kappa^{\Psi_a}(\mu_{D_1}^{\omega_{\mathrm{arc}}}) - E_\kappa^{\Psi_a}(\mu_{D_2}^{\omega_{\mathrm{arc}}})\| &\leq \sum_{\boldsymbol{x}\in D_1\cup\psi(D_1)} \omega_{\mathrm{arc}}(\boldsymbol{x})\|\kappa_G(\cdot,\boldsymbol{x}) - \kappa_G(\cdot,\gamma(\boldsymbol{x}))\|_{\mathscr{F}} \\ &+ \sum_{\boldsymbol{x}\in D_1\cup\psi(D_1)} |\omega_{\mathrm{arc}}(\boldsymbol{x}) - \omega_{\mathrm{arc}}(\gamma(\boldsymbol{x}))|\|\kappa_G(\cdot,\gamma(\boldsymbol{x}))\|_{\mathscr{F}} \\ &+ \sum_{\boldsymbol{y}\in(D_2\cup\psi(D_2))'} \omega_{\mathrm{arc}}(\boldsymbol{y})\|\kappa_G(\cdot,\gamma(\boldsymbol{y}))\|_{\mathscr{F}} \\ &\leq L d_B(D_1, D_2) + \omega_{\mathrm{arc}}(\psi(D_1))\|\kappa_G(\cdot,\psi(D_1)) - \kappa_G(\cdot,\psi(D_2))\|_{\mathscr{F}} \\ &+ |\omega_{\mathrm{arc}}(\psi(D_1)) - \omega_{\mathrm{arc}}(\psi(D_2))|\|\kappa_G(\cdot,\psi(D_2))\|_{\mathscr{F}}.\end{aligned}$$

We conclude by observing that

$$\|\kappa_G(\cdot,\psi(D_1)) - \kappa_G(\cdot,\psi(D_2))\|_{\mathscr{F}} = 2\left(1 - \exp\left(-\frac{\|\psi(D_1) - \psi(D_2)\|^2}{2\eta^2}\right)\right). \quad \square$$

Proposition 3. *Let κ_{SW} be the SW kernel defined in (7) and let D_1, D_2 be r-dimensional persistence diagrams for some $r \geq 0$. We have*

$$d_{SW}^{\Psi_a}(D_1, D_2) \leq d_{SW}(D_1, D_2) + d_{SW}(\psi(D_1), \psi(D_2)).$$

Proof. We follow the assumptions of [12, Theorem 3.3 & Theorem 3.4]. Letting $\theta \in \mathbb{R}^2$, with $\|\theta\|_2 = 1$, we introduce $\mu_D^\theta = \sum_{p\in D} \delta_{\pi_\theta(p)}$ and $\mu_{DB}^\theta = \sum_{p\in D} \delta_{\pi_\theta(p)\circ\pi_B}$, where $\pi_\theta : \mathbb{R}^2 \to L(\theta)$ is the orthogonal projection onto the line $L(\theta) = \{\lambda\theta \ : \ \lambda \in \mathbb{R}\}$, π_B is the orthogonal projection onto the diagonal and δ is the Dirac function. Furthermore, let

$$\mathcal{W}(\mu,\nu) = \inf_{P\in\Pi(\mu,\nu)} \iint_{\mathbb{R}\times\mathbb{R}} |x - y| P(dx, dy)$$

where $\Pi(\mu,\nu)$ is the set of measures on $\mathbb{R}^2$ with marginals μ and ν. Then,

$$d_{SW}^{\Psi_a}(D_1, D_2) = \frac{1}{2\pi}\int_{\mathbb{S}} \mathcal{W}(\mu_{\Psi_a(D_1)}^\theta + \mu_{\Psi_a(D_2)B}^\theta, \mu_{\Psi_a(D_2)}^\theta + \mu_{\Psi_a(D_1)B}^\theta) d\theta,$$

and, under the condition of [12, Remark 2.2], we have that

$$\mathcal{W}(\mu_{D_1}^\theta, \mu_{D_2}^\theta) = \|S_{D_1}^\theta - S_{D_2}^\theta\|_1,$$

where $S_{D_1}^\theta, S_{D_2}^\theta$ are the vectors of points $x \in D_1$ and $y \in D_2$, respectively, projected onto the line $L(\theta)$ and then sorted from the smallest value to the highest. Then, we have

$$\begin{aligned}&\mathcal{W}(\mu_{\Psi_a(D_1)}^\theta + \mu_{\Psi_a(D_2)B}^\theta, \mu_{\Psi_a(D_2)}^\theta + \mu_{\Psi_a(D_1)B}^\theta) = \\ &\qquad = \|S_{\Psi_a(D_1)\cup\Psi_a(D_2)B}^\theta - S_{\Psi_a(D_2)\cup\Psi_a(D_1)B}^\theta\|_1\end{aligned}$$

$$\leq \|S^{\theta}_{D_1\cup D_2\mathcal{B}} - S^{\theta}_{D_2\cup D_1\mathcal{B}}\|_1 + \|S^{\theta}_{\psi(D_1)\cup\psi(D_2)\mathcal{B}} - S^{\theta}_{\psi(D_2)\cup\psi(D_1)\mathcal{B}}\|_1$$
$$= \mathcal{W}(\mu^{\theta}_{D_1} + \mu^{\theta}_{D_2\mathcal{B}}, \mu^{\theta}_{D_2} + \mu^{\theta}_{D_1\mathcal{B}}) + \mathcal{W}(\mu^{\theta}_{\psi(D_1)} + \mu^{\theta}_{\psi(D_2)\mathcal{B}}, \mu^{\theta}_{\psi(D_2)} + \mu^{\theta}_{\psi(D_1)\mathcal{B}}),$$

from which our statement follows. □

Corollary 1. *In the hypotheses of Proposition 3, we have*

$$d^{\Psi_a}_{SW}(D_1, D_2) \leq 2\sqrt{2}(d_{W,1}(D_1, D_2) + \|\psi(D_1) - \psi(D_2)\|_\infty).$$

Proof. We can write

$$d^{\Psi_a}_{SW}(D_1, D_2) \leq d_{SW}(D_1, D_2) + d_{SW}(\psi(D_1), \psi(D_2))$$
$$\leq 2\sqrt{2}(d_{W,1}(D_1, D_2) + d_{W,1}(\psi(D_1), \psi(D_2)))$$
$$\leq 2\sqrt{2}(d_{W,1}(D_1, D_2) + \|\psi(D_1) - \psi(D_2)\|_\infty). \quad \square$$

4.2. The feature extraction map: Ψ_ρ

Let $\rho \in \mathbb{N}$ and $\widetilde{D}_r(\mathcal{X}, \varepsilon) = D_r(\mathcal{X}, \varepsilon) \setminus \mathcal{B}$. We define the set $\widetilde{D}_r(\mathcal{X}, \varepsilon, \rho)$, which consists of the ρ most persistent elements in $\widetilde{D}_r(\mathcal{X}, \varepsilon)$, and the function

$$\Psi_\rho(D_r(\mathcal{X}, \varepsilon)) = \widetilde{D}_r(\mathcal{X}, \varepsilon, \rho) \cup \psi\left(D_r(\mathcal{X}, \varepsilon) \setminus \widetilde{D}_r(\mathcal{X}, \varepsilon, \rho)\right) \cup \mathcal{B}.$$

We remark that $(D_r(\mathcal{X}, \varepsilon) \setminus \widetilde{D}_r(\mathcal{X}, \varepsilon, \rho)) \in \mathfrak{D}_r(\varepsilon)$, therefore $\Psi_\rho(D_r(\mathcal{X}, \varepsilon))$ is well defined.

The map Ψ_ρ performs a *feature extraction* procedure. Indeed, the resulting persistence diagram consists of the most ρ persistent elements, $\mathcal{B}$ while the remaining elements, which are possible large in numbers, are *compressed* into a single generator. Nevertheless, it can be related to Ψ_a by the formula

$$\Psi_\rho = \widetilde{D}_r(\mathcal{X}, \varepsilon, \rho) \cup \Psi_a(D_r(\mathcal{X}, \varepsilon) \setminus \widetilde{D}_r(\mathcal{X}, \varepsilon, \rho)) \setminus (D_r(\mathcal{X}, \varepsilon) \setminus \widetilde{D}_r(\mathcal{X}, \varepsilon, \rho)).$$

As we experiment in Section 5, Ψ_ρ turns out to be useful especially in cases where many generators are related to *noisy* structures in the data, leading to a consistent saving in terms of computational costs.

4.3. The auxiliary function ψ

We define the auxiliary function $\psi : \mathfrak{D}_r(\varepsilon) \to \mathbb{R}^2_{\geq\mathcal{B}}$ as

$$\psi(D_r(\mathcal{X}, \varepsilon)) = \frac{1}{W} \sum_{\boldsymbol{x}\in\widetilde{D}_r(\mathcal{X},\varepsilon)} w(\boldsymbol{x})\,\boldsymbol{x} \tag{8}$$

where $w : \mathbb{R}^2_+ \longrightarrow \mathbb{R}_+$ is a weight function and $W = \sum_{\boldsymbol{x}\in\widetilde{D}_r(\mathcal{X},\varepsilon)} w(\boldsymbol{x})$. We observe that $\widetilde{D}_r(\mathcal{X}, \varepsilon)$ contains a finite number of elements (generators), therefore the sum in (8) is always defined (see also Remark 2).

We propose the following weights.

1. Let $w_1(\boldsymbol{x}) = 1/|\widetilde{D}_r(\mathcal{X}, \varepsilon)|$, being $|\widetilde{D}_r(\mathcal{X}, \varepsilon)|$ the cardinality of the multiset, i.e., each element is counted with its multiplicity. We denote as *centre of uniform mass* the resulting auxiliary function
$$\psi_1(D_r(\mathcal{X}, \varepsilon)) = \frac{1}{|\widetilde{D}_r(\mathcal{X}, \varepsilon)|} \sum_{\boldsymbol{x}\in\widetilde{D}_r(\mathcal{X},\varepsilon)} \boldsymbol{x}.$$
2. Let $\boldsymbol{x} = (b, d) \in \widetilde{D}_r(\mathcal{X}, \varepsilon)$, where $b, d \in \mathbb{R}_+$ are the birth–death *time* of the element $\boldsymbol{x}$ (see (3)), and let $w_2(x) = d - b$ be the persistence of $\boldsymbol{x}$. We denote as *centre of persistence* the auxiliary function
$$\psi_2(D_r(\mathcal{X}, \varepsilon)) = \frac{1}{\sum_{\boldsymbol{x}=(b,d)\in\widetilde{D}_r(\mathcal{X},\varepsilon)}(d-b)} \sum_{\boldsymbol{x}=(b,d)\in\widetilde{D}_r(\mathcal{X},\varepsilon)} (d-b)\,\boldsymbol{x}.$$

Both the options aim at extracting a *representative* element from the generators of the persistence diagram. While the centre of uniform mass is the *barycentre* of the elements of the multiset, the centre of persistence assigns different weights according to the persistence of the elements. This is a natural choice to be analysed, since elements with low persistence are more likely to be related to *noise* structures resulting in the filtration, while elements of large persistence are linked to more representative geometrical features of the dataset (see the example in Fig. 1). For these reasons, in Section 5 we mainly use the function ψ_2 in our experiments, but we also compare ψ_1 and ψ_2 in Section 5.1.

Remark 2. If we take $\boldsymbol{x} \in D_r(\mathcal{X}, \varepsilon)$ in (8), additional conditions on the weight function w are needed to guarantee the convergence of the sum. However, such infinite setting is not meaningful to be analysed, since elements in the bisector carry no topological information concerning the dataset. As a further observation, the centre of persistence might be formally computed by summing over $D_r(\mathcal{X}, \varepsilon)$, as in this case $w(\boldsymbol{x}) = 0$ for $\boldsymbol{x} \in \mathcal{B}$.

Table 1
Demographic details and baseline cognitive status measures of the study population.

	AD (mean)	AD (st.dev.)	Control (mean)	Control (st.dev.)
No. of subjects	225	–	248	–
Gender (F/M)	114/111	–	126/122	–
Hand preference (A/L/R)	5/23/197	–	6/26/216	–
Age at entry	74.41	7.60	65.21	9.62
Education (years)	14.77	3.08	16.04	2.51
MMSE	20.33	6.38	29.27	1.30

Remark 3. Since $D_r(\mathcal{X}, \epsilon)$ contains an infinite number of elements for all $\mathcal{X} \subset \Omega$ by definition, the VSPK κ_Ψ is still well defined on $\mathfrak{D}_r(\epsilon) \times \mathfrak{D}_r(\epsilon)$. Moreover, if κ is (strictly) positive definite, so it is κ_Ψ.

Remark 4. Referring to Remark 1, if κ is a PWG or SW kernel, then κ_Ψ can be directly expressed in terms of the distance $d_\Psi(\cdot,\cdot)$ induced in the variably scaled setting.

5. Experiments

Our aim is to show how the SVM classifier may benefit of the introduced VSPKs.

In the experiments the kernels are handled using Python 3.8 and the modulus scikit-learn [24] on a 2.6 GHz Dual-Core Intel Core i5. Persistence diagrams are constructed via the modulii *persim* [25], *ripser* [26] and *GUDHI* [27]. Free and open source PYTHON software is available at

https://github.com/reevost/vspk_code .

We validate the following hyperparameters. About the considered kernels, we follow the guidelines provided by the authors of the seminal papers.

- Concerning the SVM classifier, we validate $\zeta \in \{10^j \mid j = -3, \dots, 3\}$.
- Concerning the PSS kernel, we take $\sigma \in \{10^j \mid j = -3, \dots, 3\} \cup \{5 \cdot 10^j \mid j = -3, \dots, 2\}$.
- Concerning the PWG kernel, the parameters C and τ of the PWGK are chosen in $\{10^j \mid j = -2, \dots, 2\}$, while δ is set to 10 (see [23, Theorem 3.2]). Moreover, as underlying standard kernel we use the Gaussian.
- Concerning the SW kernel, χ is obtained following the procedure carried out in [12, §4].

To assess the performance of the classifiers, we consider the following scores.

- The accuracy score

$$\text{accuracy} = \frac{\text{true positives} + \text{true negatives}}{\text{true positives} + \text{true negatives} + \text{false positives} + \text{false negatives}}.$$

- The f_1-score [28]

$$\text{f}_1\text{-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}},$$

where

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}},$$

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}.$$

The f_1-score is indeed the harmonic mean between precision and recall, and it is then widely-used in literature since it represents a useful trade-off between these two important metrics.

5.1. Alzheimer's disease diagnosis

The Open Access Series of Imaging Studies (OASIS) is a project aimed at making neuroimaging data sets of the brain freely available for the scientific community. In particular, OASIS-3 is a compilation of MRI and PET imaging and related clinical data for 1098 participants who were collected across several ongoing studies in the Washington University Knight Alzheimer Disease Research Center over the course of 15 years. Imaging data is accompanied by dementia and APOE status and longitudinal clinical and cognitive outcomes [29,30].

We consider a subset of the full study group, in order to have a balanced set of data.

A summary of demographic and neuropsychological details of the subjects considered in our study is presented in Table 1.

For each subject, we build the persistence diagrams using the estimation of cortical thickness on 34 points in both right and left hemisphere of the brain, for a total of 64 values. For simplicity, in the study we consider the same coordinates of the above

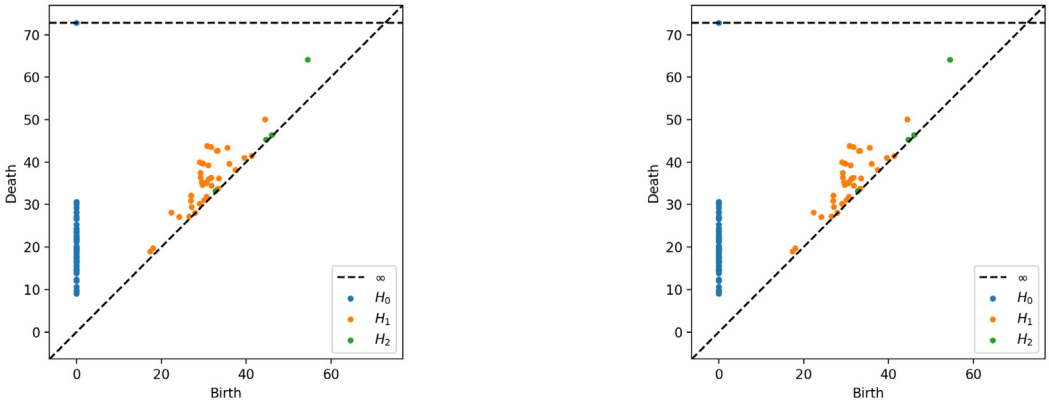

Fig. 2. Persistence diagrams of an AD subject with a MMSE of 30 (left) and the persistence diagram of a control subject with a MMSE of 7 (right).

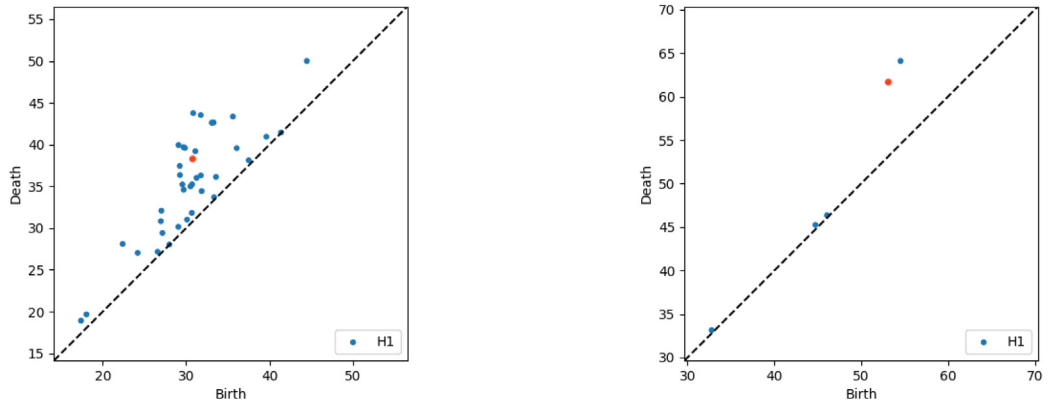

Fig. 3. 1-dimensional (left) and 2-dimensional (right) persistence diagram of an AD subject. The red dot is the added centre of persistence via Ψ_a and ψ_2.

Table 2
OASIS-3 dataset. Results of SVMs classification obtained considering H_1 and H_2 persistence diagrams and using the SW kernel.

	Ψ	ψ	Accuracy	f_1-score	Validation time (s)
SW (H_1)	–	–	0.741	0.716	220
VSP-SW (H_1)	Ψ_a	ψ_2	0.732	0.700	223
VSP-SW (H_1)	Ψ_ρ	ψ_2	0.731	0.693	184
SW (H_2)	–	–	0.741	0.712	183
VSP-SW (H_2)	Ψ_a	ψ_2	0.753	0.720	300

mentioned points for all subjects. The coordinates are computed with the *scipy* toolbox [31]. From this coordinates we build the persistence diagrams and we extract 1 and 2-dimensional topological features, i.e., we obtain the generators associated with H_1 and H_2 homological groups.

In Fig. 2 we show two examples of persistence diagrams, and in Fig. 3 we highlight the generator added as centre of persistence.

We evaluate the performance achieved by a SVMs classifier that makes use of the presented PSS, PWG and SW kernels, both in the classical and in the variably scaled settings.

In each test, we perform a random 70%/30% splitting of the dataset for training and testing, and we consider 5-fold cross validation on the training set for the tuning of the hyperparameters. The results displayed in Tables 2 3 4 have been averaged over 10 runs of tests. There, the reported validation time refers to the time required by the whole cross-validation process.

For the variably scaled setting, we consider Ψ_a for both H_1 and H_2 diagrams, while Ψ_ρ, with $\rho = 10$, is employed with H_1 only, since H_2 diagrams are limited in the number of generators, and therefore compressing features is not meaningful.

Furthermore, we use ψ_2 as auxiliary function. Indeed, as highlighted in Table 5, ψ_2 definitely outperforms ψ_1 in our setting. Moreover, we observe that the performances achieved by the auxiliary function alone, i.e., taking the centres of mass or persistence in place of the persistence diagram, are definitely not competitive with respect to the classical and variably scaled settings.

We observe that VSPKs are competitive with respect to the classical setting, improving the performance in some cases. Moreover, the usage of Ψ_ρ leads to a consistent saving in validation time.

Table 3
OASIS-3 dataset. Results of SVMs classification obtained considering H_1 and H_2 persistence diagrams and using the PWG kernel.

	Ψ	ψ	Accuracy	f_1-score	Validation time (s)
PWGK (H_1)	–	–	0.749	0.723	18165
VSP-PWG (H_1)	Ψ_a	ψ_2	0.750	0.726	18082
VSP-PWG (H_1)	Ψ_ρ	ψ_2	0.759	0.735	8731
PWG (H_2)	–	–	0.716	0.699	4422
VSP-PWG (H_2)	Ψ_a	ψ_2	0.709	0.683	4713

Table 4
OASIS-3 dataset. Results of SVMs classification obtained considering H_1 and H_2 persistence diagrams and using the PSS kernel.

	Ψ	ψ	Accuracy	f_1-score	Validation time (s)
PSS (H_1)	–	–	0.743	0.721	9238
VSP-PSS (H_1)	Ψ_a	ψ_2	0.752	0.728	9045
VSP-PSS (H_1)	Ψ_ρ	ψ_2	0.750	0.723	3330
PSS (H_2)	–	–	0.781	0.762	2825
VSP-PSS (H_2)	Ψ_a	ψ_2	0.775	0.755	3120

Table 5
OASIS-3 dataset. Results of SVMs classification obtained by using the centre of mass and persistence alone in place of the persistence diagrams.

	Accuracy	f_1-score
ψ_1	0.56	0.55
ψ_2	0.65	0.72

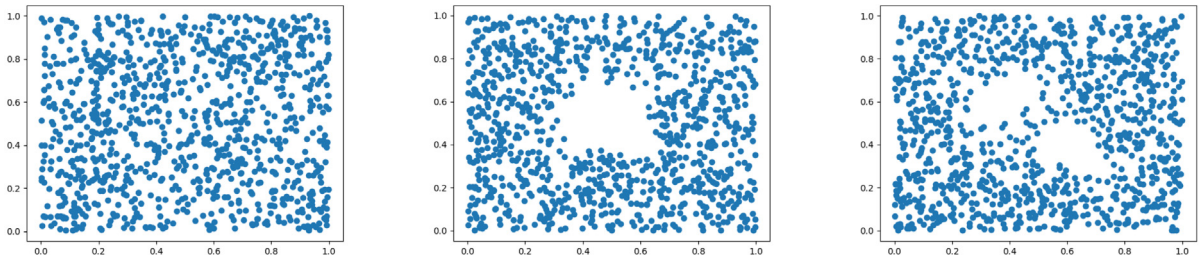

Fig. 4. Fixed $(x_0, y_0) \in [0,1]^2$, the orbits resulting from the linked twisted map taking $r = 2.5, 4.1, 4.3$, from left to right, respectively.

5.2. Orbit recognition

As second experiment, we follow the idea proposed in [32] and we analyse the linked twisted map, which models fluid flows. The corresponding orbits are computed via the discrete system

$$\begin{cases} x_{n+1} = x_n + r y_n(1 - y_n) & \mod 1 \\ y_{n+1} = y_n + r x_{n+1}(1 - x_{n+1}) & \mod 1 \end{cases}$$

where $(x_0, y_0) \in [0,1] \times [0,1]$ is the initial position and $r > 0$ is a real parameter that influences the orbit. The topological structure of the orbit changes with the initial position and r, as displayed in Figs. 4 and 5, where we depict the first 1000 *iterations* $\{(x_n, y_n) : n = 0, \dots, 1000\}$.

In the following tests, accordingly to [32], we choose a set of five parameters $r = 2.5, 3.5, 4, 4.1, 4.3$ as set of classification labels. For each label, we compute the first 1000 points of 50 orbits, with random starting point. Therefore, the dataset consists of 250 elements. Then, we compute the persistence diagram related to each orbit.

Here, since each persistence diagram has a huge number of generators ($\approx 10^5$), we restrict to the most $\rho = 10$ persistent elements in each diagram. Indeed, in our experiments, we observed that considering a larger number of generators does not significantly affect the classification performance, while leading to a higher computational cost. More precisely, here Ψ_a is computed with respect to such 10 elements, while in the case of Ψ_ρ the discarded less persistent generators are compressed in a unique element via ψ_2. Moreover, since we are dealing with 2-dimensional orbits, we compute the H_1 homology group only.

As in Section 5.1, we consider a 5-fold cross validation on the training set, and the results displayed in Tables 6–8 are averaged over 10 runs with 70%/30% training-test split of the data.

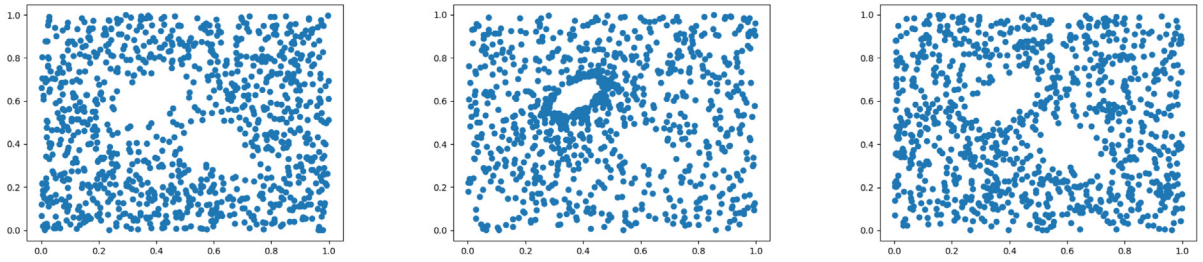

Fig. 5. Fixed $r = 4.3$, the orbits resulting from the linked twisted map taking different starting points $(x_0, y_0) \in [0, 1]^2$.

Table 6
Orbit Recognition. Results of SVMs classification on H_1 persistence diagrams using the SW kernel.

	Ψ	ψ	Accuracy	f_1-score
SW (H_1)	–	–	0.832	0.831
VSP-SW (H_1)	Ψ_a	ψ_2	0.814	0.812
VSP-SW (H_1)	Ψ_ρ	ψ_2	0.833	0.832

Table 7
Orbit Recognition. Results of SVMs classification on H_1 persistence diagrams using the PWG kernel.

	Ψ	ψ	Accuracy	f_1-score
PWG (H_1)	–	–	0.858	0.866
VSP-PWG (H_1)	Ψ_a	ψ_2	0.858	0.866
VSP-PWG (H_1)	Ψ_ρ	ψ_2	0.846	0.853

Table 8
Orbit Recognition. Results of SVMs classification on H_1 persistence diagrams using the PSS kernel.

	Ψ	ψ	Accuracy	f_1-score
PSS (H_1)	–	–	0.806	0.803
VSP-PSS (H_1)	Ψ_a	ψ_2	0.824	0.821
VSP-PSS (H_1)	Ψ_ρ	ψ_2	0.826	0.823

Table 9
Details about the considered datasets.

Category	Number of samples	Number/type of labels
Ant	72546	5 (head, torso, tail, legs, antennas)
Airplane	59546	4 (head, torso, tail, wings)
Bird	50902	4 (head, torso, tail, wings)
Octopus	66248	2 (head, legs)

5.3. 3D shape segmentation

Here, we follow an experiment proposed also in [12]. We consider some categories of the mesh segmentation benchmark introduced in [33], which contains different 3D shapes of several categories. Precisely, we consider four different categories/datasets: *Ant*, *Airplane*, *Bird* and *Octopus*. In the datasets, each sample is a *face*, which is represented as a triplet of 3D points. Separately for each category, the classification task consists in associating each sample to its corresponding segmentation label, which describes the relative position of the sample with respect to the shape. We point out that four different classification tasks are indeed considered, and each dataset is related to a different number of labels, as detailed in Table 9. A H_1 persistence diagram is then computed for each face by using the geodesic distance on the 3D shape; for more details concerning the datasets, we refer to [34], while in Fig. 6 we display some examples of shape segmentation.

As in the previous subsections, we consider a 5-fold cross validation on the training set, and the results displayed in Tables 10 and 11 are averaged over 10 runs with 20%/80% training-test split of the data.

We remark that the elements of the persistence diagrams in this experiment are limited in numbers and free of noisy generators, because of the particular construction of the diagrams obtained via the geodesic metric. Therefore, the usage of Ψ_ρ is not significant.

6. Conclusions and future work

In this paper, we proposed VSPKs for dealing with persistence diagrams in the context of persistent homology. The proposed framework, which is directly inspired by the variably scaled setting explored in kernel-based approximation and learning, may

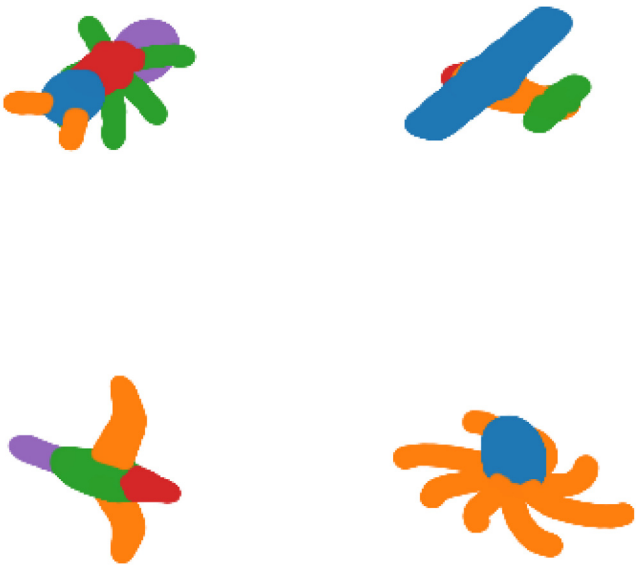

Fig. 6. Examples of 3D shapes and corresponding segmentation labels. From top left, clock-wise, Ant, Airplane, Bird and Octopus.

Table 10
Accuracy achieved by the SVMs classifiers in the carried out tests. In the variably scaled setting, we used Ψ_a and ψ_2 functions.

	PSS (H_1)	VSP-PSS (H_1)	SW (H_1)	VSP-SW (H_1)
Ant	0.925	0.920	0.925	0.934
Airplane	0.822	0.820	0.822	0.843
Bird	0.881	0.858	0.881	0.882
Octopus	0.915	0.925	0.915	0.942

Table 11
f1-score achieved by the SVMs classifiers in the carried out tests. In the variably scaled setting, we used Ψ_a and ψ_2 functions.

	PSS (H_1)	VSP-PSS (H_1)	SW (H_1)	VSP-SW (H_1)
Ant	0.910	0.905	0.910	0.913
Airplane	0.811	0.813	0.811	0.814
Bird	0.826	0.779	0.826	0.807
Octopus	0.912	0.922	0.912	0.940

enhance the performance and the efficiency of existing kernels for persistence diagrams, as suggested by the obtained results. Future work consists of investigating more on the design of the scaling function, which plays a key role in the construction of the kernel. In this view, the analysis of VSPKs in the context of algebraic topology may provide useful insights.

Declaration of competing interest

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.

Acknowledgements

This research has been accomplished within the Rete ITaliana di Approssimazione (RITA) and the thematic group on Approximation Theory and Applications of the Italian Mathematical Union. We also received the support of GNCS-INδAM. Data were provided in part by OASIS-3: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P50 AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, R01 EB009352.

References

[1] Schölkopf B, Smola AJ. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press; 2002.
[2] Vapnik V. An overview of statistical learning theory. IEEE Trans Neural Netw 1999;10(5):988–99.

[3] Schaback R, Wendland H. Kernel techniques: from machine learning to meshless methods. Acta Numer 2006;15:543–639, 2269747.
[4] Shawe-Taylor J, Cristianini N. Kernel Methods for Pattern Analysis. Cambridge University Press; 2004.
[5] Borgwardt KM, Ong CS, Schönauer S, Vishwanathan SVN, Smola AJ, Kriegel H-P. Protein function prediction via graph kernels. Bioinformatics 2005;21(suppl-1):i47–56.
[6] Collins M, Duffy N. Convolution kernels for natural language. In: Proceedings of the 14th international conference on neural information processing systems: natural and synthetic. NIPS'01, Cambridge, MA, USA: MIT Press; 2001, p. 625–32.
[7] Lampert CH. Kernel methods in computer vision. Found Trends® Comput Graph Vis 2009;4(3):193–285.
[8] Vishwanathan SVN, Schraudolph NN, Kondor R, Borgwardt KM. Graph kernels. J Mach Learn Res 2010;11:1201–42, 2645450.
[9] Carlsson G. Topology and data. Bull Am Math Soc (N.S.) 2009;46(2):255–308, 2476414.
[10] Edelsbrunner H, Letscher D, Zomorodian A. Topological persistence and simplification. In: 41st Annual symposium on foundations of computer science (Redondo Beach, CA, 2000). Los Alamitos, CA: IEEE Comput. Soc. Press; 2000, p. 454–63, 1931842.
[11] Fomenko A. Visual Geometry and Topology. Berlin: Springer-Verlag; 1994, p. xvi+324, Translated from the Russian by Marianna V. Tsaplina 1261418.
[12] Carrière M, Cuturi M, Oudot S. Sliced wasserstein kernel for persistence diagrams. In: Proceedings of the 34th International Conference on Machine Learning - Volume 70. ICML'17, JMLR.org; 2017, p. 664–73.
[13] Kusano G, Fukumizu K, Hiraoka Y. Kernel method for persistence diagrams via kernel embedding and weight factor. J Mach Learn Res 2017;18. Paper No. 189, 41 3827077.
[14] Le T, Yamada M. Persistence Fisher kernel: A Riemannian manifold kernel for persistence diagrams. In: NIPS'18, Red Hook, NY, USA: Curran Associates Inc.; 2018, p. 10028–39.
[15] Reininghaus J, Huber S, Bauer U, Kwitt R. A stable multi-scale kernel for topological machine learning. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition. 2015, p. 4741–8.
[16] Bozzini M, Lenarduzzi L, Rossini M, Schaback R. Interpolation with variably scaled kernels. IMA J Numer Anal 2015;35(1):199–219, 3335202.
[17] Campi C, Marchetti F, Perracchione E. Learning via variably scaled kernels. Adv Comput Math 2021;47(4). Paper No. 51, 23 4280526.
[18] Marchetti F, De Martino F, Shamseddin M, De Marchi S, Brisken C. Variably scaled kernels improve classification of hormonally-treated patient-derived xenografts. In: 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems. 2020, p. 1–6.
[19] Mercer J, Forsyth AR. XVI. Functions of positive and negative type, and their connection the theory of integral equations. Philos Trans R Soc Lond Ser A 1909;209(441–458):415–46.
[20] Fasshauer G, McCourt M. Kernel-based approximation methods using MATLAB. WORLD SCIENTIFIC; 2015.
[21] Guillemard M, Iske A. Interactions between kernels, frames, and persistent homology. In: Recent applications of harmonic analysis to function spaces, differential equations, and data science. Appl. Numer. Harmon. Anal., Birkhäuser/Springer, Cham; 2017, p. 861–88, 3701278.
[22] Cohen-Steiner D, Edelsbrunner H, Harer J. Stability of persistence diagrams. Discrete Comput Geom 2007;37(1):103–20, 2279866.
[23] Kusano G, Fukumizu K, Hiraoka Y. Persistence weighted Gaussian kernel for topological data analysis. ICML'16, JMLR.org; 2016, p. 2004–13.
[24] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30, 2854348.
[25] Saul N, Tralie C. Scikit-TDA: Topological data analysis for python. 2019.
[26] Tralie C, Saul N, Bar-On R. Ripser.py: A lean persistent homology library for python. J Open Source Softw 2018;3(29):925.
[27] The GUDHI Project. GUDHI User and Reference Manual. 3.5.0. GUDHI Editorial Board; 2022.
[28] van Rijsbergen CJ. Information Retrieval. Butterworth; 1979.
[29] LaMontagne PJ, Benzinger TL, Morris JC, Keefe S, Hornbeck R, Xiong C, et al. OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. 2019, MedRxiv.
[30] Pachauri D, Hinrichs C, Chung MK, Johnson SC, Singh V. Topology-based kernels with application to inference problems in Alzheimer's disease. IEEE Trans Med Imaging 2011;30(10):1760–70.
[31] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental algorithms for scientific computing in python. Nature Methods 2020;17:261–72.
[32] Adams H, Emerson T, Kirby M, Neville R, Peterson C, Shipman P, et al. Persistence images: a stable vector representation of persistent homology. J Mach Learn Res 2017;18. Paper No. 8, 35 3625712.
[33] Chen X, Golovinskiy A, Funkhouser T. A benchmark for 3D mesh segmentation. ACM Trans Graph (Proc. SIGGRAPH) 2009;28(3).
[34] Carrière M, Oudot SY, Ovsjanikov M. Stable topological signatures for points on 3D shapes. Comput Graph Forum 2015;34(5):1–12.