Persistence Paths and Signature Features in Topological Data Analysis
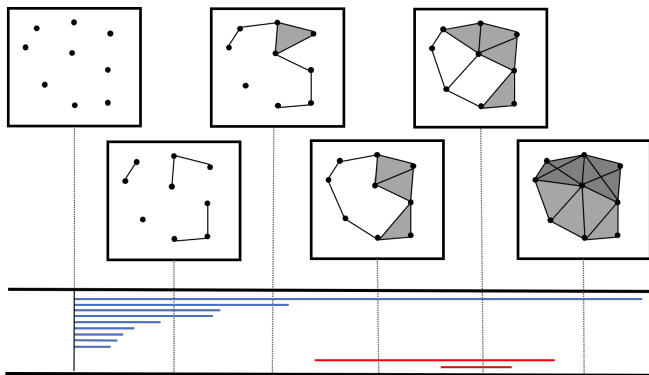
Ilya Chevyrev, Vidit Nanda, and Harald Oberhauser

Abstract—We introduce a new feature map for barcodes as they arise in persistent homology computation. The main idea is to first realize each barcode as a path in a convenient vector space, and to then compute its path signature which takes values in the tensor algebra of that vector space. The composition of these two operations — barcode to path, path to tensor series — results in a feature map that has several desirable properties for statistical learning, such as universality and characticticness, and achieves state-of-the-art results on common classification benchmarks.

Index Terms—Topological data analysis, barcodes, signature features, kernel learning

✦

1 INTRODUCTION

ALGEBRAIC topology provides a promising framework for extracting nonlinear features from finite metric spaces via the theory of *persistent homology* [1], [2], [3]. Persistent homology has solved a host of data-driven problems in disparate fields of science and engineering — examples include signal processing [4], proteomics [5], cosmology [6], sensor networks [7], molecular chemistry [8] and computer vision [9]. The typical output of persistent homology computation is called a *barcode*, and it constitutes a finite topological invariant of the coarse geometry which governs the shape of a given point cloud.

For the purposes of this introduction, it suffices to think of a barcode as a (multi)set of intervals $[b_\bullet, d_\bullet)$, each identifying those values of a scale parameter $\epsilon \geq 0$ at which some topological feature — such as a connected component, a tunnel, or a cavity — is present when the input metric space is thickened by ϵ. A central advantage of persistent homology is its remarkable *stability* theorem [10, Ch. 5.6]. This result asserts that the map **Met** → **Bar** which assigns barcodes to finite metric spaces is 1-Lipschitz when its source and target are equipped with certain natural metrics.

Persistence paths and signature features

Notwithstanding their usefulness for certain tasks, barcodes are notoriously unsuitable for standard statistical inference because **Bar** itself is a nonlinear metric space, and most scalable learning algorithms rely on linear methods. In this work, we construct a feature map of the form

$$\Phi_\bullet : \mathbf{Bar} \to \mathbf{T},$$

where $\mathbf{T} = \mathbf{T}(V)$ denotes the *tensor algebra* of a linear space V. The feature map $\Phi_\bullet$ is defined as composite, $\Phi_\bullet = \mathrm{S} \circ \iota_\bullet$, of a *persistence path embedding* $\iota_\bullet$ and the *path signature* S,

$$\mathbf{Bar} \xrightarrow{\iota_\bullet} \mathbf{BV}(V) \xrightarrow{\mathrm{S}} \mathbf{T}(V),$$

where the intermediate space $\mathbf{BV}(V)$ contains all continuous maps $[0,1] \to V$ of bounded variation.

Persistence path embedding $\iota_\bullet$ The maps $\iota_\bullet : \mathbf{Bar} \to \mathbf{BV}(V)$ disambiguate our $\Phi_\bullet$'s. There are many such embeddings, and they differ significantly in terms of stability, computability, and discriminative power.

Signature features S The map S represents a path as its $\mathbf{T}$-valued *signature*. This map is injective (modulo natural equivalence classes of paths), provides a hierarchical description of a path, and has a rich algebraic structure that captures natural operations on paths, such as concatenation and time reversal.

The concept of a persistence path embedding $\iota_\bullet$ reflects the interpretation of persistent homology as a dynamic description of the topological features which appear and disappear as a metric space is thickened across various scales. There is precedent for such constructions; e.g. Bubenik's landscapes [11] can be reformulated to give an important example of such an $\iota_\bullet$, which we denote with ι_{iL}. Despite their intuitive appeal, these approaches rely ultimately on a choice of feature map for paths, on which the resulting statistical learning guarantees depend.[1] We show here that

- *Ilya Chevyrev and Harald Oberhauser are with the Mathematical Institute, University of Oxford.*

- *Vidit Nanda is with the Mathematical Institute, University of Oxford, and the School of Mathematics, Institute for Advanced Study.*

1. For example, [11] chooses a functional on the Banach space of paths, but how to choose such functionals in a non-parametric fashion and evaluate them efficiently remains unclear (unless something special is known a priori about the probability distribution of the observed barcodes).

the composition with the signature map resolves such issues. For example, one of our results is that the feature map $\Phi_{iL} = S \circ \iota_{iL}$ is

universal: non-linear functions of the data are approximated by linear functionals in feature space: for every (sufficiently regular) function $f : \mathbf{Bar} \to \mathbb{R}$ there exists ℓ in the dual of $\mathbf{T}$, such that $f(B) \approx \langle \ell, \Phi_{iL}(B) \rangle$ uniformly over barcodes B.

characteristic: the expected value of the feature map characterizes the law of the random variable: the map which sends a probability measure μ on $\mathbf{Bar}$ to its expectation $\mu \mapsto \mathbb{E}_{B \sim \mu}[\Phi_{iL}(B)]$ in $\mathbf{T}$ is injective.

stable: the map $\Phi_{iL} : \mathbf{Bar} \to \mathbf{T}$ has explicit continuity properties, as recorded in Theorems 2 and 5 below.

Perhaps the biggest advantage of our approach is that it is not limited to ι_{iL}. Besides ι_{iL}, we will also discuss the following unstable path embeddings:

the naive embedding ι_N sorts all intervals decreasing in length (with intervals of equal length ordered by increasing birth times), enumerates them $\{1, \ldots, n\}$, and forms an n-dimensional path by running in the i-th coordinate with unit speed if the i-th bar is active (otherwise remaining constant).

Euler embedding ι_χ reduces a barcode to a single Euler characteristic curve (see [12, Sec. 3.2]). The resulting feature map Φ_χ is not stable, but is extremely fast to compute.

Betti embedding ι_β records only the Betti numbers as a function of the scale, and ignores information (contained in the barcode) which connects homology across different scale values.

envelope embedding ι_E constructed by sorting the intervals $[b_\bullet, d_\bullet)$ of a barcode in descending order by length, and then assembling the ordered sequence of $b_\bullet$'s and $d_\bullet$'s into two separate paths. This appears to be a completely new embedding.

Analogous statements for universality and characteristicness hold for the other $\Phi_\bullet$'s. Each of these embeddings $\iota_\bullet$ leads to different properties in terms of stability, computability, and discriminative power, for the associated feature maps $\Phi_\bullet$. For example, Φ_E has neither the stability of Φ_{iL} nor the computability of Φ_χ, but it gives state-of-the-art performance on supervised classification tasks. The emergence of a single feature map which is optimal along all three axes (stability, computability, discriminative power) appears unlikely, since these requirements tend to contravene each other. For example, stability requires the feature map to depend mostly on the longer intervals in a barcode, while in various problems (such as [5]), the signal of interest also resides in intervals of intermediate and short length.

Complexity

The dimension of V varies significantly between the different persistence path embeddings. If a barcode contains $\approx 10^3$ intervals, ι_N would map it to a path in a $\approx 10^3$-dimensional space, whereas ι_E always yields a path in 2 dimensions. Each of the above feature maps $\Phi_\bullet$ gives a *kernel* for barcodes $k_\bullet(B, B') := \langle \Phi_\bullet(B), \Phi_\bullet(B') \rangle$, and following [13], this kernel can be very efficiently computed

regardless of $\dim(V)$ as long as V carries an inner product. However, for low-dimensional embeddings, $\Phi_\bullet$ can be computed directly and performs very well (e.g. Φ_E for the envelope embedding ι_E).

Benchmarks and related work

Statistical learning from barcodes has received a lot of attention, see the background section in [14] for a recent survey. The most common theme is to construct a kernel [15] or polynomial coordinates [16], [17] for barcodes. We believe one strength of our approach is the access to both, the kernel and its feature map (at least for the Betti, Euler, and envelope embeddings; in practice the naive embedding is only accessible via kernelization due the high-dimensionality of the persistence paths); the former gives access to well-developed tools from the kernel and Gaussian processes learning literature, while the latter allows us to use *any* learning method such as random forests or neural networks. A second advantage is that different choices of persistence path embeddings facilitates emphasis on different topological properties, and the feature space is a well-studied object (so in a supervised learning task, the optimal $\iota_\bullet$ can be determined by cross-validation).

2 BACKGROUND

As mentioned in the introduction, we construct maps of the form

$$\mathbf{Bar} \to \mathbf{BV}(V) \to \mathbf{T}(V),$$

from the space of persistence barcodes $\mathbf{Bar}$ to the tensor algebra $\mathbf{T}(V)$ of V via the space of bounded variation paths $\mathbf{BV}$. In this section we define these three spaces and recall important properties; see [3], [10] resp. [18] for more details about $\mathbf{Bar}$ resp. $\mathbf{BV}$.

2.1 Persistence, barcodes and stability

The *Vietoris-Rips* filtration [2, Sec. 3.1] associates a one-parameter nested family of finite simplicial complexes $\{K(t)\}_{t \geq 0}$ to each finite metric space (X, Δ) via the following rule. A subset $\{x_0, \ldots, x_k\}$ of X spans a k-dimensional simplex in $K(t)$ if and only if all the pairwise distances satisfy $\Delta(x_i, x_j) \leq t$. Thus, one has the inclusion $K(t) \subseteq K(s)$ whenever $t \leq s$. Computing the homology of this family [19, Ch. 2] with coefficients in a field $\mathbb{F}$ produces, in each dimension, a corresponding family of $\mathbb{F}$-vector spaces $\mathbf{U}_\bullet$ as follows:

$$\mathbf{U}_\bullet(t) = H_\bullet(K(t); \mathbb{F}),$$

and inclusions of simplicial complexes induce linear maps $\mathbf{U}_\bullet(t) \to \mathbf{U}_\bullet(s)$ for $t \leq s$. This data consisting of vector spaces and linear maps indexed by real numbers is called a *persistence module*.

The following result from [20] uses the fact that the polynomial ring $\mathbb{F}[x]$ in one variable x acts on sufficiently tame persistence modules. Since this ring is a principal ideal domain, $\mathbb{F}[x]$-modules have a particularly simple representation theory.

Theorem 1 (Structure). *Under mild assumptions (always satisfied by Vietoris-Rips homology of finite metric spaces), each*

persistence module **U** *is completely characterized up to isomorphism by a finite collection of (not necessarily distinct) intervals* $B = B(\mathbf{U}) = \{[b_\bullet, d_\bullet)\}$, *called its* barcode.

Thus, a barcode is simply a multi-set containing some subintervals of $[0, \infty]$. When the persistence module in question comes from the Vietoris-Rips construction as described above, its barcode provides a complete summary of all the intermediate homology of $K(t)$ across $t \in [0, \infty)$. In particular, the *i*-th *Betti number* of $K(t)$, written

$$\beta_i^t = \dim H_i(K(t); \mathbb{F}), \tag{1}$$

equals the number of intervals in the *i*-dimensional barcode $B_i = \text{Bar}(\mathbf{U}_i)$ that contain *t*. Similarly, for $t \leq s$, the rank of the induced map on homology

$$\beta_i^{t,s} = \text{rank}\big(H_i(K(t); \mathbb{F}) \to H_i(K(s); \mathbb{F})\big) \tag{2}$$

equals the number of intervals in B_i that contain $[t, s]$.

There are several efficient algorithms which take as input finite metric spaces and produce as outputs the barcodes of their Vietoris-Rips filtrations [21], [22]. Concurrently, the theory has also developed at a rapid pace, and the following result [10], [23] is an exemplar of its progress. (Note that **Met** is the collection of all finite metric spaces while **Bar** is the collection of all barcodes containing finitely many intervals.)

Theorem 2 (Stability). *The map* $\text{PH}_i : \mathbf{Met} \to \mathbf{Bar}$ *which assigns to each finite metric space its i-dimensional Vietoris-Rips persistent homology barcode is 1-Lipschitz for every $i \geq 0$. Here* **Met** *is endowed with the Gromov-Hausdorff distance [24] while* **Bar** *is endowed with the bottleneck distance (as defined in [23, Sec. 3.1]).*

Roughly, two barcodes lie within bottleneck distance ϵ of each other if it is possible to deform one to the other by moving the endpoints of all its intervals by at most ϵ (and vice-versa). Thus, the longer intervals are more stable to perturbation of the originating metric space (in particular, intervals of length smaller than 2ϵ might be created or destroyed during such a deformation).

2.2 Paths of bounded variation

Let V be a normed real vector space. Given a continuous path $x : [0, T] \to V$ (for some $T \geq 0$) and a finite partition of $[0, T]$

$$\overline{p} = (0 = p_0 \leq p_1 \leq \cdots \leq p_{\ell-1} \leq p_\ell = T),$$

the *1-variation of x along $\overline{p}$* is given by

$$\text{var}_1(x; \overline{p}) = \sum_{i=0}^{\ell-1} \|x(p_{i+1}) - x(p_i)\|_V.$$

Definition 2.1. The *total 1-variation* of a continuous path $x : [0, T] \to V$ is defined as

$$\|x\|_{1\text{-var}} = \|x(0)\| + \sup_{\overline{p}} \{\text{var}_1(x; \overline{p})\},$$

where the supremum is taken over all finite partitions of $[0, T]$. The normed real vector space $\mathbf{BV}(V)$ consists of all continuous paths x that satisfy $\|x\|_{1\text{-var}} < \infty$, with addition and scalar multiplication being defined pointwise. The induced metric on $\mathbf{BV}(V)$ is given as usual by $\|x - y\|_{1\text{-var}}$.

Functions of bounded variation lie strictly between Lipschitz-continuous functions and almost everywhere differentiable functions, and in particular every Lipschitz-continuous function $[0, T] \to V$ lies within $\mathbf{BV}(V)$.

2.3 The tensor algebra

Given a real vector space V and an integer $m \geq 0$, let $V^{\otimes m} = V \otimes V \otimes \cdots \otimes V$ denote the *m*-fold tensor product of V with itself. By convention, $V^{\otimes 0} = \mathbb{R}$. The tensor algebra $\mathbf{T}(V)$ of V is the direct product

$$\mathbf{T}(V) = \prod_{m \geq 0} V^{\otimes m}.$$

Thus, each element of $\mathbf{T}(V)$ is a sequence $(v_0, v_1, \ldots)$ where $v_m \in V^{\otimes m}$. We equip $\mathbf{T}(V)$ with the structure of a (graded) algebra under the tensor product operation, for which $V^{\otimes k} \otimes V^{\otimes \ell}$ takes values in $V^{\otimes(k+\ell)}$. Finally, let us emphasize that $\mathbf{T}(V)$ is a linear space which makes it a suitable feature space.

Example 2.2. For $V = \mathbb{R}^d$, $\mathbf{T}(V)$ is the feature space for one of the arguably most important feature maps for data in $\mathbb{R}^d$: the polynomial feature map

$$x \mapsto (1, x, x^{\otimes 2}, \ldots). \tag{3}$$

As we show in Section 3, the signature map for paths in V can be seen as a generalisation of the polynomial feature map (3).

3 FROM BARCODES TO PATHS

In this section we introduce several persistence path embeddings

$$\iota_\bullet : \mathbf{Bar} \to \mathbf{BV}(V).$$

To avoid technicalities, we make the following assumption (which is always met if B arises from the persistent homology of a finite metric space):[2]

Assumption 3.1. *Assume that every barcode B encountered in this section is* tame *in two senses: first, it has only finitely many intervals, and second, each interval is contained within $[0, T_B]$ for some sufficiently large T_B.*

3.1 The (integrated) landscape embedding

We first present an embedding with desirable stability properties. The persistence landscape of a barcode B is a single function $\Lambda : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$, but it is often convenient to denote each $\Lambda(k, -) : \mathbb{R} \to \mathbb{R}$ by λ_k. The best introduction to landscapes is visual:

2. The assumption can be sufficiently mollified, but since the underlying motivation for this work is computational and finitary, we do not lose any structure of interest by restricting to tame barcodes.

To the left is a barcode containing only three intervals, where each $[b,d]$ is shown as a point in the plane with coordinates (b,d). The construction of three associated landscape functions $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$, which are shown to the right, proceeds by first projecting these points onto the diagonal, and then extracting successive maximal envelopes of the resulting arrangement of line segments. The higher λ_k for this illustrated barcode are all identically zero.

Definition 3.2. [11, Def. 3] The **landscape** $\Lambda = \Lambda^B$ of the barcode B is the (continuous) function $\mathbb{N} \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ given by

$$\Lambda^B(k,t) = \lambda_k^B(t) = \sup\{s \geq 0 \mid \beta^{t-s,t+s} \geq k\}.$$

Here, $\beta^{t-s,t+s}$ equals the number of intervals in B which contain $[t-s,t+s]$ (see (2) for the case of barcodes arising from persistent homology). Moreover, we adopt the usual convention that $\Lambda^B(k,t) = 0$ whenever the supremum is being taken over the empty set.

For tame barcodes B, one can safely exclude ∞ from the codomain of Λ^B. Moreover, each λ_k^B becomes bounded and compactly supported (in addition to continuous), so we may view the assignment of landscapes to barcodes as a function

$$\Lambda^\bullet : \mathbf{Bar} \to L^p(\mathbb{N} \times \mathbb{R})$$

for *every* $p \in [1,\infty]$.

Definition 3.3. The **landscape embedding** $\iota_\mathrm{L} : \mathbf{Bar} \to \mathbf{BV}(\ell^\infty)$ assigns to each landscape B a path $\iota_\mathrm{L}(B)$ in $\mathbf{BV}(\ell^\infty)$ whose k-th component is the k-th landscape function for each $k \in \mathbb{N}$:

$$[\iota_\mathrm{L}(B)]_k(t) = \lambda_k(t).$$

Similarly, the **integrated landscape embedding** $\iota_\mathrm{iL} : \mathbf{Bar} \to \mathbf{BV}(\ell^\infty)$ is defined as

$$[\iota_\mathrm{iL}(B)]_k(t) = \int_{-\infty}^t \lambda_k(s)\mathrm{d}s.$$

The choice of $\mathbf{BV}(\ell^\infty)$ for the target space above is somewhat arbitrary: we may as well have mapped to $\mathbf{BV}(\ell^p)$ for any $p \in [1,\infty]$ or to $\mathbf{BV}(\mathbb{R}^n)$ via truncation. We now show that ι_iL inherits stability (in the sense of Theorem 2) from barcodes via their landscapes. The two spaces defined below will appear in the proof.

Definition 3.4. For $p,q \in [1,\infty]$, define

1) the Banach space $L^{p,q}(\mathbb{N} \times \mathbb{R})$ consisting of all functions $y : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ for which the following (p,q)-norm is finite: $\|y\|_{L^{p,q}} = [\int_\mathbb{R} (\sum_{k=1}^\infty |y(k,t)|^p)^{q/p}\mathrm{d}t]^{1/q}$, and

2) the Sobolev path space $W^{1,q}(\mathbb{R},\ell^p)$ consisting of all functions $x : \mathbb{R} \to \ell^p$ for which there exists some $\dot{x} \in L^q(\mathbb{R},\ell^p)$ such that $x(t) = \int_{-\infty}^t \dot{x}(s)\mathrm{d}s$. The norm of x in this case is defined by $\|x\|_{W^{1,q}} = \|\dot{x}\|_{L^q}$.

We remark that one usually defines the Sobolev norm on $W^{1,q}$ as $\|x\|_{L^q} + \|\dot{x}\|_{L^q}$, while our definition drops the term $\|x\|_{L^q}$. For paths x defined on a compact interval $[0,T]$ with $x(0) = 0$, we note that these norms are equivalent (but on unbounded domains, this is no longer the case). Our choice of norm is motivated by the upcoming Lemma 3.5.

At special values of p and q, the two spaces defined above become more familiar. For instance, $L^{p,p}(\mathbb{N} \times \mathbb{R})$ is

the L^p space obtained by equipping $\mathbb{N} \times \mathbb{R}$ with the product of the counting and Lebesgue measures, as considered in [11, Sec. 2.4]. Similarly, $W^{1,\infty}(\mathbb{R},\ell^p)$ is the space of 1-Hölder paths in ℓ^p, while $W^{1,1}(\mathbb{R},\ell^p)$ is the subspace of absolutely continuous paths in $\mathbf{BV}(\ell^p)$. See [25, Sec. 1.4] for further details. In any case, landscapes in the image of $\Lambda^\bullet$ lie in $L^{p,q}(\mathbb{N} \times \mathbb{R})$ for every possible $p,q \in [1,\infty]$.

Lemma 3.5. *For all* $p,q \in [1,\infty]$, *the map* $\mathbb{J} : \Lambda(\star,\bullet) \mapsto \int_{-\infty}^\bullet \Lambda(\star,s)\mathrm{d}s$ *is an isometry*

$$L^{p,q}(\mathbb{N} \times \mathbb{R}) \xrightarrow{\simeq} W^{1,q}(\mathbb{R},\ell^p).$$

Proof. The map $\varphi : L^{p,q}(\mathbb{N} \times \mathbb{R}) \to L^q(\mathbb{R},\ell^p)$, $\varphi(y)(t) = (y(1,t),(y(2,t),\ldots)$, is an isometry:

$$\|y\|_{L^{p,q}}^q = \int_\mathbb{R} \left(\sum_{k=1}^\infty |y(k,t)|^p\right)^{q/p} \mathrm{d}t = \int_\mathbb{R} \|\varphi(y)(t)\|_{\ell^p}^q \mathrm{d}t$$
$$= \|\varphi(y)\|_{L^q}^q.$$

By definition of $W^{1,q}(\mathbb{R},\ell^p)$, the map $\int : L^q(\mathbb{R},\ell^p) \to W^{1,q}(\mathbb{R},\ell^p)$, $x \mapsto \int_{-\infty}^\bullet x(s)\mathrm{d}s$, is also an isometry. The conclusion follows by observing that $\mathbb{J} = \int \circ \varphi$. □

We now obtain a desirable stability property for the integrated landscape embedding.

Theorem 3. *The map* $\mathbf{Met} \to \mathbf{BV}(\ell^\infty)$ *obtained by composing the i-dimensional Vietoris-Rips persistent homology of Theorem 2 with the integrated landscape embedding of Definition 3.3, i.e.,*

$$\mathbf{Met} \xrightarrow{\mathrm{PH}_i} \mathbf{Bar} \xrightarrow{\Lambda^\bullet} L^\infty(\mathbb{N} \times \mathbb{R}) \xrightarrow{\mathbb{J}} \mathbf{BV}(\ell^\infty),$$

is 1-Lipschitz for every $i \geq 0$. *Here* $\mathbf{Met}$ *has the Gromov-Hausdorff metric while* $\mathbf{BV}(\ell^\infty)$ *has the 1-Hölder norm*

$$\|x\|_{1\text{-Höl}} := \|x(0)\|_{\ell^\infty} + \sup_{s \neq t} \|x(t) - x(s)\|_{\ell^\infty} |t-s|^{-1} .$$
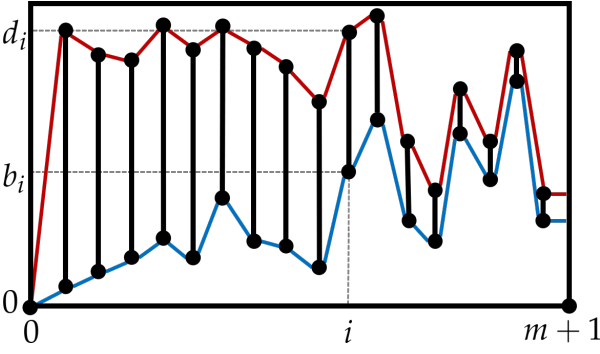
Proof. It holds that PH_i is 1-Lipschitz by Theorem 2 and $\Lambda^\bullet$ is 1-Lipschitz by [11, Thm. 13]. Taking $p = q = \infty$ in Lemma 3.5 and noting that $\Lambda^B(k,t) = 0$ for all $t \leq 0$, $k \in \mathbb{N}$, and barcodes B, it follows that $\mathbb{J}$ is also 1-Lipschitz. □

3.2 The envelope embedding

Consider $B \in \mathbf{Bar}$. Order the intervals $\{[b_i,d_i]\}_{i=1}^m$ of B in descending order by their lengths $(d_i - b_i)$ (with intervals of equal length ordered by increasing birth times), and embed them into $\mathbb{R}^2$ as the (disjoint) union

$$U(B) = \bigcup_{i=1}^m \{i\} \times [b_i,d_i].$$

The *upper* envelope $u_B : \mathbb{R} \to \mathbb{R}$ of B is the piecewise linear curve obtained by linearly interpolating between the highest points (i,d_i) of $U(B)$ across $i \in \{0,\ldots,m\}$, with $d_0 = 0$ by convention. Similarly, the *lower* envelope ℓ_B is obtained by interpolating between the lowest points (i,b_i), again with $b_0 = 0$. Both curves are uniquely extended to have domain $\mathbb{R}$ by keeping them constant on the intervals $(-\infty,0]$ and $[m,\infty)$. We illustrate both envelopes in the accompanying figure.

Definition 3.6. The **envelope embedding** is a map $\iota_E :$ **Bar** $\to$ **BV**$(\mathbb{R}^2)$ defined as follows. To each barcode B, it associates the path $\iota_E(B) : \mathbb{R} \to \mathbb{R}^2$, given by

$$[\iota_E(B)](t) = (\ell_B(t), u_B(t)).$$

Here u_B and ℓ_B are the upper and lower envelopes of B as described above.

For values of t near zero, the upper and lower envelopes, $\ell_B(t), u_B(t)$ capture only the longest, most stable intervals of B. As $t \geq 0$ increases, more of the smaller intervals get included, and the output $\iota_E(B)$ becomes more volatile to small perturbations of B (in the bottleneck metric). This motivates us to truncate after a given time: pick an integer $N \geq 1$ and let $\iota_E^N :$ **Bar** $\to$ **BV**$(\mathbb{R}^2)$ be the *restricted envelope embedding* obtained by setting $\iota_E^N(B)(t)$ equal to

$$\begin{cases} \iota_E(B)(t) & \text{if } t \leq N, \\ \iota_E(B)(N) & \text{if } t > N, \end{cases}$$

If the feature map associated to this truncated envelope embedding performs well for small values of N and poorly for large ones, then one obtains evidence in favor of the hypothesis that the signal of interest genuinely resides in the larger, more stable intervals.

3.3 The Betti and Euler embeddings

In contrast to the previous two subsections, we now consider embeddings which depend on all homological dimensions.

Definition 3.7. Denote $B = \bigcup_{i \geq 0} B_i$ where B_i contains all intervals of homological dimension i. Choose an integer $n \geq 1$ and numbers $a_i^k \in \mathbb{R}$ for $i \geq 0$ and $1 \leq k \leq n$. Setting $a = \{a_i^k\}$, the **generalised Betti embedding** $\beta(B;a) \in$ **BV**$(\mathbb{R}^n)$ is defined as follows. Let $\{t_j\}_{j=1}^m$ be the (ordered set of) all endpoints of intervals in B (which lie in $[0,T]$) together with $t_1 = 0$ and $t_m = T$. For each $j \in \{1,\dots,m\}$, set

$$\beta(B;a)(t_j) = \left(\sum_{i \geq 0} a_i^1 \beta_i^{t_j}, \dots, \sum_{i \geq 0} a_i^n \beta_i^{t_j} \right),$$

where $\beta_i^{t_j}$ is the Betti number of B as in (1). We extend the definition to points $t \in (t_j, t_{j+1})$ in a piecewise linear fashion

$$\beta(B;a)(t) = \beta(B;a)(t_j) + \frac{t - t_j}{t_{j+1} - t_j}\Big(\beta(B;a)(t_{j+1}) - \beta(B;a)(t_j) \Big).$$

The **Betti embedding** $\iota_\beta(B) \in$ **BV**$(\mathbb{R}^n)$ is defined by setting $a_i^k = 1$ if $k - 1 = i$ and $a_i^k = 0$ otherwise:

$$\iota_\beta(B)(t_j) = (\beta_0^{t_j}, \dots, \beta_{n-1}^{t_j}).$$

The **Euler embedding** $\iota_\chi(B) \in$ **BV**$(\mathbb{R})$ is defined by setting $a_i^1 = (-1)^i$:

$$\iota_\chi(B)(t_j) = \sum_{i=0}^n (-1)^i \beta_i^{t_j}.$$

Remark 3.8. When B arises from the persistent homology of a finite metric space, it is computationally convenient to recall that $\iota_\chi B(t_j)$ is the Euler characteristic of the associated Vietoris-Rips simplicial complex $K(t_j)$, and is thus also given by an alternating count of simplices across dimension:

$$\iota_\chi(B)(t_j) = \sum_{i \geq 0} (-1)^i \cdot \#\{i\text{-simplices in } K(t_j)\}.$$

Hence, $\iota_\chi(B)$ can be computed without knowing the actual homology of K. In fact, one could further consider generalised simplex embeddings

$$\sigma^k(B;a)(t_j) = \sum_{i \geq 0} a_i^k \cdot \#\{i\text{-simplices in } K(t_j)\},$$

which can all be computed without knowing the homology of K (albeit do not in general capture homological invariants).

Remark 3.9. Since ι_β and particularly ι_χ are massive numerical reductions of B, it is apparent that metric spaces with very different persistence barcodes might have identical Betti and Euler embeddings. On the other hand, for barcodes arising from certain popular models of random metric spaces, the expected value of $|\iota_\chi(B)(t)|$ is a remarkably good predictor of the Betti number β_i^t for each $t \geq 0$ — see [26, Sec. 5.3] and the references therein for details.

3.4 Stability and injectivity

As mentioned in the introduction, the emergence of a single feature map which is optimal in terms of stability, discriminative power and computability is unlikely since these three properties tend to contravene each other. Indeed, our persistence path embeddings vary drastically in terms of stability, discriminative power, and computability. In terms of discriminative power, the (integrated) landscape and envelope embeddings are injective maps from the space of barcodes to spaces of bounded variation paths, but neither Betti or Euler are injective, see Remark 3.9. In terms of stability, the only embedding that is stable is the (integrated) landscape embedding; for the other embeddings simple counterexamples can be constructed.[3]

4 FROM PATHS TO TENSORS

We introduce the second component S for our feature map $\Phi_\bullet = S \circ \iota_\bullet$.

Definition 4.1. For a Banach space V, the signature map is defined as

$$S : \mathbf{BV}(V) \to \mathbf{T}(V) \quad x \mapsto (S_0(x), S_1(x),\dots)$$

---

3. Consider adding small bars $[b_n, d_n]$ where $[b_n, d_n] = [0, n^{-1})$ if n is even, $[b_n, d_n] = [c - n^{-1}, c)$ if n odd for a fixed sufficiently large $c > 0$.

where $S_0(x) = 1$ and

$$S_m(x) := \int_{0 < t_1 < \cdots < t_m < T} dx(t_1) \otimes dx(t_2) \otimes \cdots \otimes dx(t_m) \in V^{\otimes m}$$

is defined as a Riemann–Stieltjes integral over the $m-1$ simplex of length T.

Example 4.2 (The case $V = \mathbb{R}^n$). For a path $x \in \mathbf{BV}(\mathbb{R}^n)$, it holds that

$$S_m(x) = \sum_{i_1, \ldots, i_m} S^{i_1, \ldots, i_m}(x) e_{i_1} \otimes \cdots \otimes e_{i_m} \in (\mathbb{R}^n)^{\otimes m},$$

where the sum is taken over all multi-indexes $(i_1, \ldots, i_m) \in \{1, \ldots, n\}^m$, $(e_i)_{i=1}^n$ is a basis of $\mathbb{R}^n$, and

$$S^{i_1, \ldots, i_m}(x) = \int_{0 < t_1 < \cdots < t_m < T} dx_{i_1}(t_1) dx_{i_2}(t_2) \cdots dx_{i_m}(t_m).$$

Hence the term $S_m(x)$ can simply be interpreted as a collection of n^m real numbers.

The mapping S is essentially injective up to a natural equivalence class of paths, called tree-like equivalence.[4]

Theorem 4. *[27] Let $x, y \in \mathbf{BV}(V)$. Then $S(x) = S(y)$ iff x and y are tree-like equivalent.*

Tree-like equivalence can be a useful equivalence relation, e.g. it identifies paths that differ only by time-parametrization.[5] For our applications, it is instructive to think of $S_m(x)$ as the natural generalisation of the monomial of order m of a vector $v \in V$ to pathspace.

Example 4.3. Let $v \in \mathbb{R}^n$ and consider the path $x(t) = tv$ where $t \in [0, 1]$. The components of its signature are given by

$$S_m(x) = \int d(vt_1) \otimes \cdots \otimes d(vt_m) = \int v^{\otimes m} dt_1 \cdots dt_m = \left(\frac{v^{\otimes m}}{m!} \right),$$

Thus $S(x)$ indeed recovers the moment map $v \mapsto (1, v, \frac{v^{\otimes 2}}{2!}, \ldots)$.

A further useful similarity between the signature map and monomials is that the space of linear functions of the signature is closed under multiplication, which is commonly known as the shuffle identity.

Lemma 4.4 (Shuffle identity). *Suppose $\ell_i \in (V')^{\otimes m_i}$ for $i = 1, 2$ (where V' is the continuous dual space of V). Then there exists $\ell_3 \in (V')^{\otimes (m_1 + m_2)}$ such that for all $x \in \mathbf{BV}(V)$*

$$\langle S(x), \ell_1 \rangle \langle S(x), \ell_2 \rangle = \langle S(x), \ell_3 \rangle$$

Proof. See [18, Thm. 2.5] for an elementary proof when V is finite dimensional, and [28, Cor. 3.9] for the general case. □

Remark 4.5. The linear functional ℓ_3 is known as the shuffle product of ℓ_1 and ℓ_2.

---

4. x, y are tree-like equivalent iff there exists some $\mathbb{R}$-tree (i.e., a metric space in which any two points are connected by a unique arc which is isometric to a real interval) $\mathfrak{T}$ so that $x * \overleftarrow{y}$, the concatenation of x with the time-reversal of y, decomposes as

$$[0, T] \xrightarrow{\phi} \mathfrak{T} \xrightarrow{\psi} V$$

where ϕ and ψ are continuous maps with $\phi(0) = \phi(T)$. We call a set of continuous paths in $\mathbf{BV}(V)$ *reduced* if none of its distinct elements are tree-like equivalent.

5. x and $x' = (x(\varphi(t)))$ are tree-like equivalent for any time-change φ.

In light of Lemma 4.4, one can ask whether linear combinations of such "monomials" $S_m(x)$ are dense in a space of functions, and, whether the sequence of expected "moments" characterizes the law of the random path. For compact subsets of $\mathbf{BV}(V)$, the answer to both questions is yes, as we shall see in Theorem 5, and follows by a standard Stone–Weierstrass argument; the general, non-compact case is more subtle (cf. classical moment problem) but is known to be true under suitable integrability conditions [29].

5 STATISTICAL LEARNING

In this section we discuss the problem of statistical learning on the space of barcodes **Bar**. The space **Met** has traditionally been of more interest; however, since the persistent homology map PH : **Met** → **Bar** from Section 2.1 is well-understood, we focus here on **Bar**. Results for **Bar**-valued random variables pull back along PH to results for **Met**-valued random variables.

We are interested in two standard learning problems: given independent random samples $B_1, \ldots, B_k \sim B$ of a **Bar**-valued random variable B, our aim is to

(i) learn a function $f(B)$ of the data $f : \mathbf{Bar} \to \mathbb{R}$, and
(ii) characterize the law μ of the data $B \sim \mu$.

As mentioned in the introduction, the standard approach to both problems is to find a feature map $\Phi : \mathbf{Bar} \to \mathbf{T}(V)$ which is universal and characteristic (addressing points (i) and (ii) respectively). Let us establish these properties for our feature map

$$\Phi = S \circ \iota : \mathbf{Bar} \to \mathbf{T}(V),$$

where V is a Banach space, ι is any persistence path embedding (e.g., one of the maps from Section 3) and S is the signature map of Section 4. Due to the injectivity of S (up to tree-like equivalence), Φ preserves essentially all the information captured by ι. In particular, if ι maps some domain $\mathcal{D} \subset \mathbf{Bar}$ injectively into the space of tree-reduced paths (as is the case for any embedding once time is added as a coordinate), then Φ is also injective on $\mathcal{D}$. To make this precise, we use suitable quotient spaces.

Definition 5.1. For $\iota : \mathbf{Bar} \to \mathbf{BV}(V)$, define the equivalence relation $B \sim_\iota B'$ iff $\iota(B)$ and $\iota(B')$ are tree-like equivalent. Let **Bar**/ι denote the quotient of **Bar** under $\sim_\iota$, and equip **Bar**/ι with the initial topology with respect to the map

$$\mathbf{Bar}/\iota \to \mathbf{BV}(V)/\sim_t, \quad [B] \mapsto [\iota(B)],$$

where $\sim_t$ denotes tree-like equivalence and $\mathbf{BV}(V)/\sim_t$ is equipped with the quotient topology (recall that $\mathbf{BV}(V)$ bears the 1-variation topology).

Theorem 5. *Define*

$$\Phi : \mathbf{Bar}/\iota \to \mathbf{T}(V), \quad B \mapsto S \circ \iota(B).$$

On each compact subset $K \subset \mathbf{Bar}/\iota$, the map Φ has the following properties.

(1) (**Universal**) *Let $f : K \to \mathbb{R}$ be continuous. For each $\epsilon > 0$, there exists ℓ in $\bigoplus_{m \geq 0} (V')^{\otimes m}$ (the dual space of the tensor algebra) such that*

$$\sup_{B \in K} |f(B) - \langle \Phi(B), \ell \rangle| < \epsilon.$$

(2) (**Characteristic**) *Denoting by* $\mathcal{M}$ *the set of Borel probability measures on K, the map*

$$\mathcal{M} \to \mathbf{T}(V), \quad \mu \mapsto \mathbb{E}_{B \sim \mu}[\Phi(B)]$$

is injective.

(3) (**Kernelized**) *Suppose further that V is a Hilbert space. Then the map*

$$k : K \times K \to \mathbb{R}, \quad k(B, B') = \langle \Phi(B), \Phi(B') \rangle$$

defines a bounded, continuous kernel[6] which is universal for the space of continuous functions $C(K, \mathbb{R})$ and characteristic for Borel probability measures on K.

Proof. By Theorem 4, the continuity of the signature map in the 1-variation topology [18, Thm. 3.10], and the definition of $\mathbf{Bar}/\iota$, it follows that the map Φ is continuous and separates the points of $\mathbf{Bar}/\iota$. Combining these properties with Lemma 4.4 shows that the set of functions $\{B \mapsto \langle \Phi(B), \ell \rangle : \ell \in \bigoplus_{m \geq 0} (V')^{\otimes m}\}$ is a point-separating subalgebra of $C(K, \mathbb{R})$, hence Point (1) follows from the Stone–Weierstrass theorem.

Point (2) follows by duality: the dual of $C(K, \mathbb{R})$ are the Radon measures on K (these include the Borel probability measures) and universality implies that the map $\ell \mapsto (B \mapsto \langle \Phi(B), \ell \rangle)$ is dense in this dual; hence, every Radon measure is characterized by

$$\left\{ \int_K \langle \Phi(B), \ell \rangle d\mu(B) : \ell \in \bigoplus_{m=0}^{\infty} (V')^{\otimes m} \right\}$$

$$= \left\{ \left\langle \int_K \Phi(B) d\mu(B), \ell \right\rangle : \ell \in \bigoplus_{m=0}^{\infty} (V')^{\otimes m} \right\}.$$

In Point (3) the boundedness follows since from continuity of Φ and compactness of K. Finally, every inner product kernel $k(B, B') = \langle \Phi(B), \Phi(B') \rangle$ is universal [resp. characteristic] if and only if the feature map Φ is universal [resp. characteristic]; this follows from a general argument about reproducing kernels, see for example [31, Prop. E.3] for details. □

The computational bottleneck for Φ is typically the calculation of the signature, since an element of $\mathbf{T}(V)$ truncated at level M needs $O(l \dim(V)^M)$ real numbers if $\iota(B)$ is assumed piecewise linear with at most l time points. This gets prohibitively large for moderate dimensions. By contrast, the kernel on $\mathbf{Bar}$, $k(B, B') = \langle \Phi(B), \Phi(B') \rangle$, is defined via the canonical inner product on $\mathbf{T}(V)$. Using [13], the level-M approximation to $k(B, B')$ can be computed in $O(l^2 c^2 M)$ time[7] and $O(l^2)$ memory, where c is the cost of evaluating one inner product in V.

---

6. A kernel on a set $\mathcal{X}$ is a positive definite map $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$. The completion $\mathcal{H}$ of $\mathcal{H}_0 := \{k(x, \cdot) : x \in \mathcal{X}\} \subset \mathbb{R}^{\mathcal{X}}$ with respect to the inner product $\langle k(x, \cdot), k(y, \cdot) \rangle := k(x, y)$ forms a so-called reproducing kernel Hilbert space. A kernel is called universal for a topological vector space $\mathcal{F} \subset \mathbb{R}^{\mathcal{X}}$ if $\mathcal{H}_0$ embeds continuously into a dense subspace of $\mathcal{F}$ and called universal if the transpose map $\mathcal{F}' \to \mathcal{H}$ is injective, see [30] for details.

7. The low-rank approximation algorithm in [13] reduces this cost further to $O(lcM)$ time and $O(M)$ memory.

5.1 Hyperparameters

Each of the feature maps $\Phi_\bullet$ naturally generalises to a parametrised feature map $\Phi_\bullet^\pi$ where π denotes a set of parameters (which will be typically chosen by the learning algorithm). The set of parameters $\pi = \{M, \tau, \Delta, \phi\}$ are

- $M \in \mathbb{N}$ is the *truncation level* of S, meaning that we only consider the first M components $(S_0(x), S_1(x), \ldots, S_M(x))$ of every signature $S(x)$.
- $\tau \in \{0, 1\}$ is the *time-augmentation* parameter. If it is non-zero, then we replace each path $x(t) \in \mathbf{BV}(V)$ by the path $(t, x(t)) \in \mathbf{BV}(\mathbb{R} \times V)$ before computing its signature.
- $\Delta \in V^l$, for some $l \geq 0$, is a *lag vector* containing non-negative real numbers. We replace $x(t) \in \mathbf{BV}(V)$ by

$$\left(x(t), x(\max(t - \Delta_1, 0)), \ldots, x(\max(t - \Delta_l, 0)) \right)$$

  in $\mathbf{BV}(V^{l+1})$ before computing its signature.
- a *non-linearity* $\varphi : V \to W$. Since $\iota_\bullet$ composed with a sufficiently regular map $\varphi : V \to W$ to another vector space W, lifts barcodes to $\mathbf{BV}(W)$. If φ is injective and non-linear, one expects to obtain more efficient learning from signatures of paths in $\mathbf{BV}(W)$ since more non-linearities are provided by $S \circ \varphi \circ \iota_\bullet(B) \in \mathbf{T}(W)$ than by $S \circ \iota_\bullet(B) \in \mathbf{T}(V)$.[8]

For the choice $M = \infty$, $\tau = 0$, $\Delta = 0$ (with $l = 0$) and $\varphi = \mathrm{id} : V \to V$, the corresponding Φ^π recovers Φ, and $k^\pi(B, B') = \langle \Phi^\pi(B), \Phi^\pi(B') \rangle$ recovers k. With slight abuse of notation, we write Φ for Φ^π and k for k^π for the remainder of the article.

6 EXPERIMENTS

We evaluate our feature map on three supervised classification tasks: orbits, textures, and shapes. These are common benchmarks, taken from recent papers[9] and are described in Figures 1, 2, and 3. For kernel methods $k_\bullet$ we used a support vector classifier and for feature maps $\Phi_\bullet$ we used a random forest classifier. For kernel methods we used the same Nyström approximation to deal with the quadratic growth of the Gram matrix.

6.1 Computational complexity

For a persistence path $\iota_\bullet(B) \in \mathbf{BV}(\mathbb{R}^n)$, truncation of the feature map $\Phi_\bullet(B)$ at tensors at level less than or equal to M leads $O(n^M)$ coordinates (again, akin to classic polynomial features in machine learning). This combinatorial explosion limits the choice of $\iota \in \{\iota_E, \iota_\chi, \iota_\beta\}$ for Φ_ι in practice: the integrated landscape embedding ι_{iL} produces paths in $\mathbf{BV}(\mathbb{R}^n)$ for large n, which rules out Φ_{iL}. However, the kernelization

---

8. Generically W will be very high- or infinite-dimensional which prevents the direct calculation of $S \circ \varphi \circ \iota_\bullet$. But if φ is the feature map of a kernel on V, (i.e. (W, κ) is a reproducing kernel Hilbert space over V with kernel $\kappa(u, v) := \langle \varphi(u), \varphi(v) \rangle$), then the signature kernelization [13] still allows to compute the kernel $k^\pi(B, B') = \langle \Phi^\pi(B), \Phi^\pi(B') \rangle$. Typically we choose $V = \mathbb{R}^n$ and use a classic kernel on $\mathbb{R}^n$ for κ such as the RBF kernel.

9. It can be hard to replicate reported results in the literature, since the preprocessing is often not fully specified; e.g. on OUTEX various downsampling methods combined with CBLP are possible before barcodes are computed. We used the same data set of barcodes and train-test split for all experiments to allow for a fair comparison.

$k_\bullet$ needs constant memory and computation time $O(Mn)$; in practice this allows us to evaluate up to $n \approx 10^2$ on standard laptops [13]; thus $k_\bullet$ can be efficiently computed for all $\iota_\bullet$'s as long as the persistence paths are not too long (long persistence paths are in principle possible by using the low-rank algorithm from [13] which we did not implement). On the other hand $\Phi_E(B)$, $\Phi_\chi(B)$, and $\Phi_\beta(B)$ are fast to compute directly since their ι's produce low-dimensional paths.

6.2 Implementation

We implement our feature map $\Phi_\bullet$ and the kernel $k_\bullet$ in Python's sci-kit learn package [32]. As part of this, we use for the signature kernel computation $k_\bullet$ the legacy code from [13, Alg. 3], but we did not implement the low-rank algorithm (which limits $k_\bullet$ to paths with ≤ 300 time ticks on laptops, which is the reason why we do not report results for k_χ and k_β on the Orbits dataset); for $\Phi_\bullet|_M = S|_M \circ \iota_\bullet$ we make use of Terry Lyons' ESIG package to compute the truncated signature $S|_M$.

To allow for a fair comparison, we also implemented the sliced Wasserstein kernel k_{SW} [15] and the persistence image feature map [33] combined with a random forest. This allows for a reasonable benchmarking since [15], [33] compares performance of k_{SW} and Φ_{PI} against a number of other methods — see [15, Sec. 4.2 and 4.3] and [33, Sec. 6].

6.3 Hyperparameter tuning

We use a grid search and 5-fold cross-validation for parameter tuning. The hyperparameter grid for $\Phi_\bullet^\pi$ consists of parameters π that are: the truncation level M in the tensor algebra and time-augmentation parameter τ; additionally the parameters of the classifiers were used, and if the kernelized version is used, then additionally the parameters of the kernel on V (we used throughout the Gauss kernel so this is just $\sigma > 0$). For the envelope embedding E a further grid search was performed over the restriction parameter N described in Section 3.2. Adding lags described in Section 5.1 is possible but was not tested here.

6.4 Performance

Table 1 reports the mean accuracy of repeating each experiment 20 times together with the standard deviation. As benchmark we report performance of the sliced Wasserstein kernel k_{SW} and the persistence image features Φ_{PI}. For k_{SW}, we used the approximation given in [15, Alg. 1] with six directions. For Φ_{PI}, we used the `persim` package[10] with the Gauss kernel, linear weight function, and a grid search over the number of pixels and variance $\sigma > 0$.

Table 1 shows that our approach performs very competitively, achieving state-of-the art in two common benchmarks. To the best of our knowledge, the Betti embedding ι_β beats the state-of-the-art for shapes in feature form and achieves close to state-of-the-art for textures in kernelized form. This is encouraging since both k_{SW} and Φ_{PI} provide very competitive benchmarks.

Remark 6.1. Our implementation of k_{SW} achieves better accuracy, particularly for Orbits, than reported in the original

10. https://github.com/sauln/persim

TABLE 1: Mean accuracy ($\pm$ standard deviation).

Method	Textures	Orbits	Shapes
k_{SW}	96.8 ± 1.0	94.6 ± 1.3	95.8 ± 1.6
Φ_{PI}	93.7 ± 1.0	$\mathbf{99.86 \pm 0.21}$	90.3 ± 2.3
k_E	90.4 ± 1.5	96.6 ± 0.9	92.7 ± 1.5
k_χ	94.9 ± 0.6	NA	92.4 ± 3.0
k_β	$\mathbf{97.8 \pm 0.2}$	NA	93.0 ± 3.0
Φ_E	88.1 ± 0.8	98.1 ± 1.0	95.0 ± 0.9
Φ_χ	92.9 ± 0.7	98.8 ± 0.6	98.0 ± 1.1
Φ_β	96.6 ± 0.6	97.7 ± 0.8	$\mathbf{98.1 \pm 0.7}$

papers [15], where it is 96.1 ± 0.4 for Textures and 83.7 ± 0.5 for Orbits. We believe this is due to a different Gram matrix approximation (we use Nyström for all kernel methods). Similarly Φ_{PI} drastically outperformed the results for the same orbits experiment in [33]; we believe this is due to our choice of classifier, namely random forest (vs. discriminant subspace ensemble).

The best results for k_{SW}, Φ_{PI}, and E were achieved by looking at the 0th homology for texture and orbits, and 1st homology for shapes (we did not combine homologies of different dimensions for these methods).

The values M for the optimal truncation level in the tensor algebra (as chosen by cross-validated gridsearch) is given in Table 2; we ran all tests for $1 \leq M \leq 8$. The features $\Phi_\bullet$ performed best at higher levels, while $k_\bullet$ performed best at lower levels. We suspect that the reason is that the Gaussian kernel non-linearity that lifts barcodes to paths in infinite-dimensional spaces, allows to capture the needed information already on level 2 or 3. We also ran the same experiments for the landscape[11] and the naive embedding, ι_L and ι_N, but the results were not competitive on either dataset.

TABLE 2: Best truncation level M.

M	Textures	Orbits	Shapes
k_E	2	2	2
k_χ	3	NA	2
k_β	3	NA	2
Φ_E	5	8	4
Φ_χ	8	7	7
Φ_β	6	8	5

7 CONCLUSION

We have presented a two-step methodology to build a feature map for barcodes: the first step identifies a barcode as a persistence path; the second step computes the signature of this path. The motivation for the first step is that it captures the dynamic, time-indexed nature of topological persistence. In other words, just as important as the appearance and disappearance of topological features across various scales is the *order* in which these events occur. The motivation for the second step is that the signature provides an extremely well-studied feature map for paths. The four central advantages of this approach are:

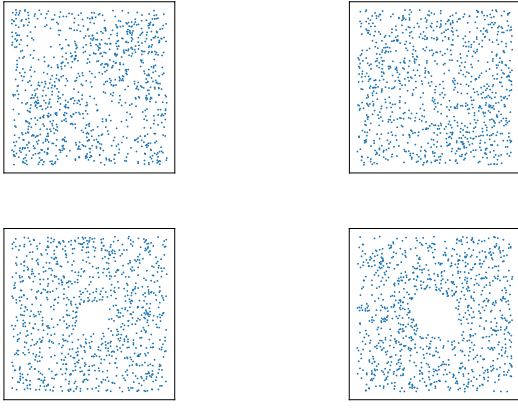11. We employed [34, Alg. 1] to compute persistence landscapes

Fig. 3: **Shapes.** The synthetic shapes dataset [33], [36] consists of point cloud data with six labels: random, circle, sphere, clusters, clusters-of-clusters and torus. Within each label, there are 50 point clouds containing 500 points each. The Gaussian noise model used in each case has standard deviation 0.1. We use a 50% to 50% test-train split and the persistence diagrams provided by [36] https://github.com/bziiuj/pcodebooks. Note that our learning method is different to that of [33], [36] where *K*-medoids clustering is used.

Fig. 1: **Orbits.** The dataset consists of 500 orbits. Each orbit is a set of 1001 points in $\mathbb{R}^2$, $\{(x_n, y_n) : n = 0, \ldots, 1000\} \subset \mathbb{R}^2$, generated by one out of five discrete dynamical systems with a random initial value (x_0, y_0). The five dynamical systems are given by taking the parameter $r \in \{2.5, 3.5, 4, 4.1, 4.3\}$, which thus acts as label, and update rule $x_{n+1} = x_n + ry_n(1 - y_n) \mod 1$, and $y_{n+1} = y_n + rx_n(1 - x_n) \mod 1$ where (x_0, y_0) is chosen uniformly at random in $(0,1)^2$. For each of the labels r we generated 100 orbits and used 50% to 50% as train-test split for the resulting 500 orbits. Above shows one orbit for each value of $r = 2.5, 3.5, 4, 4.1$ (left to right and top to bottom).
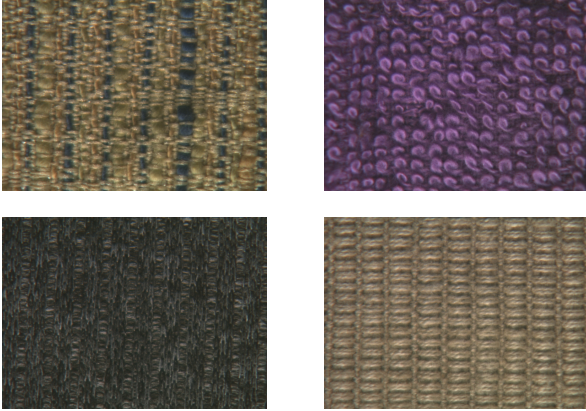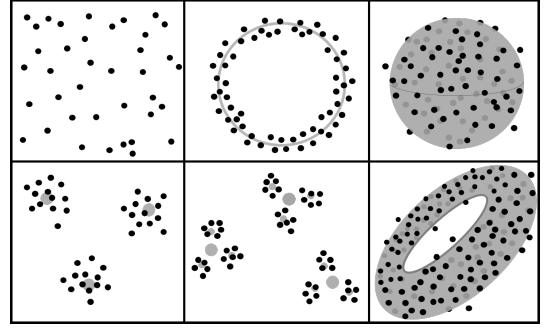

Fig. 2: **Textures.** The *Outex_TC_00000* dataset of surface textures [35]. The data set consists of 240 images for training, 240 for testing, as prescribed by the test suite. Each sample carries one out of 24 labels. Above shows texture of four different labels.

Theory. Barcodes are represented as elements in the tensor algebra given as iterated integrals. The tensor algebra has a rich and well-understood algebraic structure that makes it easy to establish both universality and characteristicness. Moreover, this allows tools from stochastic analysis and non-commutative algebra to interpret barcodes produced by topological data analysis.

Performance. We benchmarked our feature maps on standard datasets against two recent, very competitive methods (one kernelized, one unkernelized) and achieved state-of-the art performance on two out of the three benchmarks.

Flexibility. Persistence path embeddings differ in stability, computability, and discriminative power. This flexibility to emphasize one aspect over the other can prove decisive for a given dataset. For example, the Betti and Euler embeddings provide a large dimensionality reduction, with the Euler embedding yielding further computational advantages since the whole barcode never has to be computed. In this case, the price is a reduced discriminative power, which was not relevant to the benchmark datasets considered here.

(Un)kernelized learning. We provide both kernelized and unkernelized versions of our feature maps. The former allows us to use the well-developed modular kernel learning approach that provides statistical learning guarantees. The latter allows to use far more general classifiers (e.g., those which use regularization beyond Tikhonov to avoid overfitting), and hence potentially deal with large scale problems that are beyond the current reach of kernel-based methods.

REFERENCES

[1] R. Ghrist, "Barcodes: the persistent topology of data," *Bulletin of the American Mathematical Society*, vol. 45, no. 1, pp. 61–75, 2008.

[2] V. Nanda and R. Sazdanović, "Simplicial models and topological inference for biological systems," in *Discrete and Topological Models in Molecular Biology*, N. Jonoska and M. Saito, Eds. Springer, 2014, ch. 6, pp. 109–141.

[3] S. Oudot, *Persistence theory: from quiver representations to data analysis*, ser. Mathematical Surveys and Monographs. American Mathematical Society, 2015, vol. 209.

[4] J. A. Perea and J. Harer, "Sliding windows and persistence: An application of topological methods to signal analysis," *Foundations of Computational Mathematics*, vol. 15, no. 3, pp. 799–838, Jun 2015. [Online]. Available: https://doi.org/10.1007/s10208-014-9206-z

[5] M. Gameiro, Y. Hiraoka, S. Izumi, M. Kramar, K. Mischaikow, and V. Nanda, "A topological measurement of protein compressibility," *Japan Journal of Industrial and Applied Mathematics*, vol. 32, no. 1, pp. 1–17, 2015.

[6] T. Sousbie, "The persistent cosmic web and its filamentary structure – I. Theory and implementation," *Monthly Notices of the Royal Astronomical Society*, vol. 414, no. 1, pp. 350–383, 2011.

[7] V. de Silva and R. Ghrist, "Coverage in sensor networks via persistent homology," *Algebraic and Geometric Topology*, vol. 7, no. 1, pp. 339–358, 2007.

[8] K. Xia, X. Feng, Y. Tong, and G. W. Wei, "Persistent homology for the quantitative prediction of fullerene stability," *Journal of Computational Chemistry*, vol. 36, pp. 408–422, 2015.

[9] J. Lamar-Leon, R. Alonso-Baryolo, E. Garcia-Reyes, and R. Gonzalez-Diaz, "Persistent homology-based gait recognition robust to upper body variations," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, Dec 2016, pp. 1083–1088.

[10] F. Chazal, V. de Silva, M. Glisse, and S. Oudot, *Structure and stability of persistence modules*, ser. SpringerBriefs in Mathematics. Springer, 2016.

[11] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *Journal of Machine Learning Research*, vol. 16, p. 77–102, 2015.

[12] K. Turner, S. Mukherjee, and D. M. Boyer, "Persistent homology transform for modeling shapes and surfaces," *Information and Inference: A Journal of the IMA*, vol. 3, no. 4, pp. 310–344, 2014.

[13] F. J. Király and H. Oberhauser, "Kernels for sequentially ordered data," *ArXiv e-prints, 1601.08169*, 2016, accepted in Journal of Machine Learning Research subject to minor revision.

[14] H. Adams, S. Chepushtanova, T. Emerson, E. Hanson, M. Kirby, F. Motta, R. Neville, C. Peterson, P. Shipman, and L. Ziegelmeier, "Persistence Images: A Stable Vector Representation of Persistent Homology," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 218–252, 2017.

[15] M. Carrière, M. Cuturi, and S. Oudot, "Sliced Wasserstein kernel for persistence diagrams," in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. International Convention Centre, Sydney, Australia: PMLR, 06–11 Aug 2017, pp. 664–673. [Online]. Available: http://proceedings.mlr.press/v70/carriere17a.html

[16] A. Adcock, E. Carlsson, and G. Carlsson, "The ring of algebraic functions on persistence bar codes," *Homology Homotopy Appl.*, vol. 18, no. 1, pp. 381–402, 2016. [Online]. Available: https://doi.org/10.4310/HHA.2016.v18.n1.a21

[17] B. Di Fabio and M. Ferri, "Comparing persistence diagrams through complex vectors," in *Image analysis and processing—ICIAP 2015. Part I*, ser. Lecture Notes in Comput. Sci. Springer, Cham, 2015, vol. 9279, pp. 294–305. [Online]. Available: https://doi.org/10.1007/978-3-319-23231-7_27

[18] T. J. Lyons, M. Caruana, and T. Lévy, "Differential equations driven by rough paths," Berlin, pp. xviii+109, 2007, lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard.

[19] A. Hatcher, *Algebraic topology*. Cambridge University Press, 2002.

[20] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete and Computational Geometry*, vol. 33, pp. 249–274, 2005.

[21] K. Mischaikow and V. Nanda, "Morse theory for filtrations and efficient computation of persistent homology," *Discrete and Computational Geometry*, vol. 50, no. 2, pp. 330–353, 2013.

[22] R. Ghrist and G. Henselman, "Matroid filtrations and computational persistent homology," *arXiv:1606.00199v2[math.at]*, 2016.

[23] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, "Stability of persistence diagrams," *Discrete and Computational Geometry*, vol. 37, no. 1, pp. 103–120, 2007.

[24] M. Gromov, "Groups of polynomial growth and expanding maps," *Inst. Hautes Études Sci. Publ. Math.*, vol. 53, p. 53–73, 1981.

[25] P. K. Friz and N. B. Victoir, *Multidimensional stochastic processes as rough paths: theory and applications*, ser. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press, 2010.

[26] M. Kahle, "Topology of random simplicial complexes: a survey," *AMS Contemporary Math*, vol. 620, pp. 201–222, 2014.

[27] H. Boedihardjo, X. Geng, T. Lyons, and D. Yang, "The signature of a rough path: uniqueness," *Adv. Math.*, vol. 293, pp. 720–737, 2016. [Online]. Available: https://doi.org/10.1016/j.aim.2016.02.011

[28] T. Cass, B. K. Driver, N. Lim, and C. Litterer, "On the integration of weakly geometric rough paths," *J. Math. Soc. Japan*, vol. 68, no. 4, pp. 1505–1524, 2016. [Online]. Available: https://doi.org/10.2969/jmsj/06841505

[29] I. Chevyrev and T. Lyons, "Characteristic functions of measures on geometric rough paths," *Ann. Probab.*, vol. 44, no. 6, pp. 4049–4082, 2016. [Online]. Available: https://doi.org/10.1214/15-AOP1068

[30] C.-J. Simon-Gabriel and B. Schölkopf, "Kernel Distribution Embeddings: Universal Kernels, Characteristic Kernels and Kernel Metrics on Distributions," *ArXiv e-prints*, Apr. 2016.

[31] I. Chevyrev and H. Oberhauser, "Signature moments to characterize laws of stochastic processes," *ArXiv e-prints*, Oct. 2018.

[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[33] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, "Persistence images: A stable vector representation of persistent homology," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 218–252, 2017.

[34] P. Bubenik and P. Dł otko, "A persistence landscapes toolbox for topological statistics," *J. Symbolic Comput.*, vol. 78, pp. 91–114, 2017. [Online]. Available: https://doi.org/10.1016/j.jsc.2016.03.009

[35] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen, "Outex-new framework for empirical evaluation of texture analysis algorithms," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 1. IEEE, 2002, pp. 701–706.

[36] B. Zielinski, M. Juda, and M. Zeppelzauer, "Persistence Codebooks for Topological Data Analysis," *ArXiv e-prints*, Feb. 2018.

Ilya Chevyrev obtained his DPhil in mathematics in 2016 from the University of Oxford. He then worked as postdoctoral researcher at the Technical University of Berlin before returning to Oxford at the end of 2016 for a Junior Research Fellowship with St John's College. His research focuses on stochastic analysis, particularly rough paths theory and stochastic PDEs.

Vidit Nanda is an applied and computational algebraic topologist. He obtained his PhD in mathematics from Rutgers University in 2012, and has held post-doctoral positions at the University of Pennsylvania, the Alan Turing Institute, and the Institute for Advanced Study. In 2018 he joined the Mathematical Institute at the University of Oxford as an Associate Professor.

Harald Oberhauser obtained his PhD in 2010 in the Statslab at the University of Cambridge and held post-doctoral positions at the Technical University of Berlin, and the Oxford-Man Institute. In 2014 he joined Mathematics Department of University College London, and in 2015 he joined the Mathematical Institute in Oxford as Associate Professor. His research focuses on stochastic analysis and its applications.