TOPOLOGICAL FEATURE SELECTION FOR TIME SERIES DATA

PETER BUBENIK AND JOHNATHAN BUSH

ABSTRACT. We use tools from applied topology for feature selection on vector-valued time series data. We employ persistent homology and sliding window embeddings to quantify the coordinated dynamics of time series. We describe an algorithm for gradient descent to assigns scores, or weights, to the variables of the time series based on their contribution to the dynamics as quantified by persistent homology; the result is a convex combination of a subset of the variables. In this setting, we prove persistence vineyards are piecewise linear and we give a simple formula for the derivatives of the vines. We demonstrate our method of topological feature selection with synthetic data and *C. elegans* neuronal data.

1. INTRODUCTION

Many physical and biological systems exhibit periodic or quasiperiodic dynamics. The dynamics exhibited by the components of complex systems are often coordinated: neurons firing simultaneously, feedback loops of the endocrine system, animal swarming behaviors, etc. These dynamics (e.g., the activities of neurons) may be represented by vector-valued time series. In the spirit of dimensionality reduction, we consider the problem of assigning scores to the variables of these vector-valued time series based on their contribution to the global, coordinated dynamics. We refer to this problem as *topological feature selection for time series*.

Broadly, feature selection is a process aiming to identify the most relevant and informative features or variables from a given dataset. A primary goal of feature selection is to reduce dimensionality by eliminating irrelevant or redundant attributes. In our setting, we use persistent homology as a measure of a feature's relevancy and perform gradient descent to identify relevant features.

In Section 2, we describe related work on the persistent homology of sliding window embeddings, the nervous system of *C. elegans*, and the optimization of persistence-based functions. In Sections 3, 4, and 5, we summarize persistent homology, sliding window embeddings, and our notion of feature selection, respectively. In Section 6, we describe a gradient descent algorithm to identify optimal weighted subsets of time series variables with respect to their contribution to coordinated dynamics. We prove that the vines in the corresponding persistence vineyards are piecewise linear and prove that derivatives may be computed explicitly. Furthermore, we show that these derivatives are maximized along certain families of affine paths through the geometric simplex. Our focus is not on the problem of optimizing general functions of persistent homology, which has been addressed in various contexts and generality [21, 2, 16, 4, 15]; rather, we propose a simple pipeline for feature selection and emphasize the low-level interpretability of this pipeline's output.

As a motivating example, we apply our method for feature selection to neuronal activity in the model organism *C. elegans*. We show that the coordinated dynamics recorded across a large portion of the *C. elegans* brain are well-represented by a small subset of neurons.

Date: October 27, 2023.

2. Related work

Foundational results regarding the persistent homology of sliding window embeddings were established by Perea and Harer [20]. Persistent homology has been used to analyze time series data in many different settings, including gene expression [9, 19], nematode locomotion and behavior [23], periodic video [24], climate analysis [1], wheeze detection [10], and cryptocurrency variability [12], for example. These works commonly employ sliding window embeddings of varying subsets of data to draw conclusions about the statistics of the resulting collection of persistence diagrams. In this paper, we focus instead on sliding window embeddings across the entirety of a fixed data set.

Our primary motivating example is neuronal data collected from *Caenorhabditis elegans*, a transparent roundworm with a simple nervous system comprising just 302 neurons. Despite its simplicity, the *C. elegans* brain is complex enough to exhibit sensory modalities including a response to heat and smell. In particular, neurons in *C. elegans* are known to share information by engaging in coordinated dynamic activity that evolves cyclically near a low-dimensional manifold [14]. Notably, the authors of [14] use principle component analysis to reconstruct this activity, whereas we observe low-dimensional cyclic activity through a sliding window embedding. We remark that persistent homology was used to study the connectivity of the *C. elegans* nervous system in [13].

The optimization of persistence-based functions is a topic of active research; see [21, 2, 16, 4, 15], for example. In particular, the authors of [4] describe a general framework for defining and computing gradients for persistence-based functions and give a sufficient condition for the convergence of stochastic subgradient descent methods. They also provide code containing machine learning layers to implement these methods of optimization. In this paper, we de-emphasize computational and algorithmic aspects of optimization and focus on the low-level interpretability of optimal persistence diagrams arising from time series data and, more generally, linear combinations of certain symmetric matrices. As a point of clarification, we show in Section 6 that the gradient may be computed efficiently and explicitly; however, we also describe a modified algorithm for stochastic gradient descent that may be more appropriate for certain applications.

3. Persistent homology

In this section, we give an overview of persistent homology, which is the primary tool of our topological analysis. Throughout, we fix a base field $\mathbb{F}$ and take the real numbers $(\mathbb{R}, \leq)$ as an indexing poset whenever possible.

3.1. **Preliminaries.** A *simplicial complex* is a collection K of finite subsets, called *faces*, of a set V such that if $\sigma \in K$ and $\tau \subset \sigma$, then $\tau \in K$. The elements of K are called *simplices* and the elements of V are called *vertices*. Given a simplicial complex K, a *filtration of K*, also called a *filtered simplicial complex*, is a collection of simplicial complexes $\{K_t \mid t \in \mathbb{R}\}$ such that $K_r \subset K_s \subset K$ whenever $r \leq s$. Given a filtration of a finite simplicial complex K, we call the function $f \colon K \to \mathbb{R}$ defined by $\sigma \mapsto \inf\{t \in \mathbb{R} \mid \sigma \in K_t\}$ the *filtration function*, and note that $K_t = f^{-1}((-\infty, t])$.

A *persistence module* over $(\mathbb{R}, \leq)$ is a collection of $\mathbb{F}$-vector spaces $\{V_t \mid t \in \mathbb{R}\}$ and linear maps $\{f_r^s \colon V_r \to V_s \mid r \leq s\}$ such that $f_s^t \circ f_r^s = f_r^t$ whenever $r \leq s \leq t$.

A filtered simplicial complex naturally gives rise to a persistence module. For a fixed degree d, taking reduced simplicial homology with coefficients in $\mathbb{F}$ of a filtered simplicial complex

$\mathcal{K} = \{K_t \mid t \in \mathbb{R}\}$ yields a collection of $\mathbb{F}$-vector spaces $\{\tilde{H}_d(K_t) \mid t \in \mathbb{R}\}$ and linear maps $\{\tilde{H}_d(\iota_r^s) \colon \tilde{H}_d(K_r) \to \tilde{H}_d(K_s) \mid r \leq s\}$, where $\iota_r^s \colon K_r \hookrightarrow K_s$ denotes the inclusion. We let $\mathrm{PH}_d(\mathcal{K})$ denote this persistence module.

3.2. Vietoris–Rips persistent homology.
In applied topology, one associates to a finite metric space X a filtered simplicial complex such that the induced persistence module captures topological information about X at varying scales or resolutions. A computationally efficient construction along these lines is the *Vietoris–Rips filtration*, defined for a finite metric space (X, d_X) and $t \in \mathbb{R}$ by

$$\mathrm{VR}(X; t) := \left\{ \sigma \subset X \ \middle| \ \max_{x,y \in \sigma} d_X(x,y) \leq t \right\}.$$

Note that $\mathrm{VR}(X; t) = \varnothing$ for all $t < 0$. Then, $\mathrm{VR}(X) := \{\mathrm{VR}(X; t) \mid t \in \mathbb{R}\}$ is a filtered simplicial complex and we refer to the data of the induced persistence module $\mathrm{PH}_d(\mathrm{VR}(X))$ as the *Vietoris–Rips persistent homology of the metric space* (X, d_X).

We may also define Vietoris–Rips persistent homology of any positive symmetric matrix with zeros on the diagonal. Given such a matrix $M = (M_{ij})_{1 \leq i,j \leq k}$ and $t \in \mathbb{R}$, define

$$\mathrm{VR}(M; t) := \left\{ \sigma \subset \{1, \ldots, k\} \ \middle| \ \max_{i,j \in \sigma} M_{ij} \leq t \right\}.$$

Then, $\mathrm{VR}(M) := \{\mathrm{VR}(M; t) \mid t \in \mathbb{R}\}$ is a filtered simplicial complex and we refer to the data of the induced persistence module $\mathrm{PH}_d(\mathrm{VR}(M))$ as the *Vietoris–Rips persistent homology of the matrix* M. In the case that M is a distance matrix of a finite metric space X, we have equality of persistence modules $\mathrm{PH}_d(\mathrm{VR}(M)) = \mathrm{PH}_d(\mathrm{VR}(X))$, and any bijection $\{1, 2, \ldots, |X|\} \to X$ induces a canonical simplicial bijection $\mathrm{VR}(M; t) \to \mathrm{VR}(X; t)$ for all $t \in \mathbb{R}$.

Remark 3.1. These constructions generalize to infinite metric spaces and matrices as long as simplices are constrained to have finite cardinalities.

3.3. Persistence diagrams and barcodes.
Given an interval $\lambda \subset \mathbb{R}$, which may or may not contain its endpoints, we let λ_* denote the *interval persistence module* such that $\lambda_t = \mathbb{F}$ if $t \in \lambda$, $\lambda_t = \{0\}$ if $t \notin \lambda$, and all linear maps that are not forced to be zero are identity maps. The following is a version of the structure theorem for persistent homology [7], which states that the persistence module of the Vietoris–Rips filtration of a totally bounded space is isomorphic to a direct sum of interval modules. A metric space is said to be *totally bounded* if there exists a finite cover of X by ε-balls for all $\varepsilon > 0$. For example, a subset of euclidean space is totally bounded if and only if it is bounded.

Theorem 1. *If X is a totally bounded metric space, there is a family of intervals Λ such that the persistence module induced by the Vietoris–Rips filtration decomposes as*

$$\mathrm{PH}_d(\mathrm{VR}(X)) = \bigoplus_{\lambda \in \Lambda} \lambda_*.$$

In the setting of Theorem 1, the collection of intervals Λ is called the *degree d barcode* of the persistence module. Because a totally bounded space is bounded, and because we compute reduced homology, each interval in Λ has finite endpoints. The multiset of these endpoints, denoted $\mathrm{PD}_d(X)$, is called a *persistence diagram*. The elements of $\mathrm{PD}_d(X)$ belong to the *diagonal upper half-plane*

$\mathbb{R}^2_{\leq} := \{(x,y) \in \mathbb{R}^2 \mid x \leq y\}$. Given a point $(b,d) \in \mathrm{PD}_d(X)$, we refer to b as the *birth*, d as the *death*, and $d - b$ as the *lifetime* of the corresponding topological feature.

For a finite metric space X, the persistence diagram $\mathrm{PD}_d(X)$ summarizing its Vietoris–Rips persistent homology is guaranteed to contain only a finite number of points, and each point can appear with only finite multiplicity; for brevity, we call such a persistence diagram *finite*.

3.4. **Persistence pairs.** Given a filtered simplicial complex $\mathcal{K} = \{K_t \subset K \mid t \in \mathbb{R}\}$, let $f\colon K \to \mathbb{R}$ denote the corresponding filtration function. The persistence algorithm [26] computes *persistence pairs*, which are pairs of simplices $(\sigma_b, \sigma_d) \in K \times K$ such that

$$(f(\sigma_b), f(\sigma_d)) \in \mathrm{PD}_d(\mathcal{K}).$$

We call σ_b the *birth simplex* and call σ_d the *death simplex* of the point (p,q).

In general, the filtration function $f\colon K \to \mathbb{R}$ induces a total preorder on the set of simplices of $\mathcal{K}$, which can be made into a total order in a fairly arbitrary way (as long as this order respects face relations). The persistence algorithm depends on this choice of total order, but the corresponding points in the persistence diagram do not.

4. Time series and sliding window embeddings

We define a *time series* to be any (possibly discontinuous) function $\mathbb{R} \to \mathbb{R}^n$.

Definition 4.1. Given a time series $f\colon \mathbb{R} \to \mathbb{R}^n$, a positive integer L, and a nonnegative real number τ, the *sliding window embedding of f* is the function

$$\mathrm{SW}[f]\colon \mathbb{R} \to \mathbb{R}^{L \cdot n}$$
$$t \mapsto [f(t) \;\; f(t+\tau) \;\; f(t+2\tau) \;\; \cdots \;\; f(t+(L-1)\tau)]$$

where $[\cdot]$ denotes concatenation of vectors. We call the numbers L and τ *window parameters*. Given a finite subset $D \subset \mathbb{R}$, we define

$$\mathrm{SW}[f|_D] := \mathrm{Im}(\mathrm{SW}[f]|_D)$$

and call the resulting set a *sliding window point cloud*.

Window parameters will be clear through context and are suppressed from the notation. The construction of a sliding window embedding is motivated by Takens' theorem [22], which gives conditions for which a dynamical system can be reconstructed from observations of the state space.

Definition 4.2. A *continuous time dynamical system* is a pair (X, Φ), where X is a topological space and $\Phi\colon \mathbb{R} \times X \to X$ is a continuous map such that $\Phi(0,p) = p$ and $\Phi(s, \Phi(t,p)) = \Phi(s+t,p)$ for all $p \in X$ and for all $s, t \in \mathbb{R}$. Here, X is called the *state space* of the system. For a fixed $p \in X$, the set $\{\Phi(t,p) \mid t \in \mathbb{R}\}$ is called the *orbit* through p.

One may have access only to a continuous observation function $F\colon X \to \mathbb{R}^n$, rather than the dynamical system itself. In that case, the evaluation of F along the orbit through an initial state $p \in X$ yields a time series $F \circ \Phi(-, p)\colon \mathbb{R} \to \mathbb{R}^n$. Informally, Takens' theorem implies that a sliding window embedding of this time series generically will recover the topology of the orbit $\{\Phi(t,p) \mid t \in \mathbb{R}\}$. The assumptions of Takens' theorem are impossible to check in practice, but we view this as a motivating heuristic for our constructions.

In practice, data may be collected as a finite list of observations of an unknown underlying time series $f\colon \mathbb{R} \to \mathbb{R}^n$. In these cases, the corresponding sliding window point cloud is a finite sampling of the image of SW$[f]$. By Takens' theorem, we expect this point cloud to approximate an orbit of the underlying unknown dynamical system, and we may compute persistent homology to quantitatively summarize topological features of the orbit.

4.1. **Embeddings of periodic time series.** For an integer N, a real number ϕ, and window parameters τ and L, the sliding window embedding of the periodic function $\zeta(t) = \sin(Nt + \phi)$ may be computed exactly as

$$\mathrm{SW}[\zeta](t) = \sin(Nt + \phi)u + \cos(Nt + \phi)v, \tag{1}$$

where $u, v \in \mathbb{R}^L$ are such that $u = \mathrm{SW}[\cos(Nt)]|_{t=0}$ and $v = \mathrm{SW}[\sin(Nt)]|_{t=0}$ [20]. It follows that the image of SW$[\zeta]$ is a planar ellipse whenever $L \geq 2$ and u and v are linearly independent, with semi-major and semi-minor axes expressible as functions of N, τ, and L.

While the sliding window embeddings of more complicated functions are understood only in certain cases [18, 11, 25], Equation 1 provides a heuristic that is verified experimentally: namely, that sliding window embeddings of time series exhibiting periodic or near-periodic behaviour will tend to manifest as topological loops (or more generally, tori). This structure is then detectable and quantifiable through persistent homology.

5. Feature selection

Given a time series $f\colon \mathbb{R} \to \mathbb{R}^n$ with components $f(t) = (f_1(t), \ldots, f_n(t))$, we call the sliding window embedding SW$[f]$ a *global embedding*, and we call each SW$[f_i]$ a *singleton embedding*. Our analysis is focused on global embeddings of data. This approach is motivated by the following observation about the persistent homology of singleton embeddings.

Proposition 5.1. *Consider a time series $f\colon \mathbb{R} \to \mathbb{R}^n$ with components $f(t) = (f_1(t), \ldots, f_n(t))$ and a finite subset $D \subset \mathbb{R}$. For any choice of sliding window parameters τ and L, any $1 \leq i \leq n$, any homological dimension d, and real numbers $\alpha, \beta \in \mathbb{R}$, the map $f_i \mapsto \alpha f_i + \beta$ induces*

$$\mathrm{PD}_d(\mathrm{SW}[f_i|_D]) \mapsto |\alpha|\mathrm{PD}_d(\mathrm{SW}[f_i|_D]) := \{(|\alpha|b, |\alpha|d) \mid (b, d) \in \mathrm{PD}_d(\mathrm{SW}[f_i|_D])\}.$$

In other words, the persistent homology of a singleton embedding is invariant with respect to shifting the data by a constant factor, and it is invariant to scaling the data after normalizing persistence diagrams to have maximal persistence equal to one. On the other hand, a global embedding encodes relative scaling, shifting, and time delays amongst a collection of individual time series, which is a desirable property in applications with data exhibiting coordinated dynamics. For example, individual components of neuronal time series data may be real-valued time series obtained from recordings of distinct neuronal activities across the brain.

Along these lines, we are interested in determining the extent to which individual time series contribute to global dynamics as measured by the persistent homology of the sliding window point cloud. In this context, we refer to the problem of assigning scores to individual time series based on their contribution to these global dynamics *feature selection*.

Given two matrices A and B of equal dimension $m \times n$, the *Hadamard product* $A \circ B$ is the $m \times n$ matrix defined by $(A \circ B)_{ij} = A_{ij}B_{ij}$.

Definition 5.2. For a time series $f\colon \mathbb{R} \to \mathbb{R}^n$ and a vector $v \in \mathbb{R}^n$, we define the *v-weighted time series* $v \circ f\colon \mathbb{R} \to \mathbb{R}^n$ by $(v \circ f)(t) := v \circ f(t)$. For a choice of window parameters, we call $\mathrm{SW}[v \circ f]$ a *v-weighted global embedding of f*.

Vector-weighted global embeddings are continuous in an appropriate sense; see Appendix A.

Toward quantifying how singleton time series contribute to the dynamics encoded by a global embedding, we phrase topological feature selection for time series as follows:

> Fix a time series $f\colon \mathbb{R} \to \mathbb{R}^n$, and a finite subset $D \subset \mathbb{R}$. Given an objective function $F\colon \mathrm{PD}_{\mathrm{fin}} \to \mathbb{R}$ on the set of finite persistence diagrams, find $v^* \in \mathbb{R}^n$ such that $\mathrm{PD}_d(\mathrm{SW}[v^* \circ f|_D])$ locally or globally optimizes F.

For example, F may denote the sum of the lifetimes of all points in the persistence diagram (called *total persistence*), or it may denote the birth, death, or lifetime of a given point.

Without constraints on $v \in \mathbb{R}^n$, coordinate time series f_i may scale arbitrarily, and for that reason we restrict to a *feasible region* $U \subset \mathbb{R}^n$. Different choices of feasible regions may be appropriate for different applications. To maintain relationships between the singleton time series f_i leading to coordinated dynamics within the global embedding, we restrict to the set of convex coefficients.

Definition 5.3. We call $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$ a *convex coefficient vector* if $\sum_{i=1}^n \lambda_i = 1$ and $\lambda_i \geq 0$ for all i. The *geometric $(n-1)$-simplex* $|\Delta^{n-1}|$ is the set of convex coefficient vectors in $\mathbb{R}^n$.

For a time series with coordinates $f(t) = (f_1(t), \ldots, f_n(t))$ and $\lambda = (\lambda_1, \ldots, \lambda_n) \in |\Delta^{n-1}|$ locally or globally optimizing an objective function, those singleton time series f_i with coefficients λ_i closer to one are considered to be the most topologically relevant, while those with coefficients closer to zero are considered to be less relevant. That is, we interpret the coefficients λ_i as scores of topological relevancy of individual time series to global dynamics as quantified by persistent homology.

In Section 6, we consider the problem of identifying optimal $\lambda \in |\Delta^{n-1}|$ in the more general setting of linear combinations of certain symmetric matrices.

6. Optimization and algorithms

Fix a time series $f\colon \mathbb{R} \to \mathbb{R}^n$ with components $f(t) = (f_1(t), \ldots, f_n(t))$, a finite subset $D \subset \mathbb{R}$, and a choice of sliding window parameters. Note that a sliding window point cloud naturally carries the total order induced by $\mathbb{R}$; hence, we identify each with the corresponding $|D| \times |D|$ symmetric ℓ_1-distance matrix. For $\lambda \in |\Delta^{n-1}|$, is straightforward to verify the matrix equality

$$\mathrm{SW}[\lambda \circ f|_D] = \sum_{i=1}^n \lambda_i \mathrm{SW}[f_i|_D]. \tag{2}$$

That is, the distance matrix of the λ-weighted global embedding may be expressed as the corresponding convex combination of the distance matrices of the (unweighted) singleton embeddings.

For the remainder of this paper, we use a superscript notation for indexing matrices. With the above observation in mind, we consider the more general setting of linear combinations of an arbitrary set $M = \{M^1, \ldots, M^n\}$ of $k \times k$ positive symmetric matrices with zeros along the diagonal. For what follows, it will be convenient to introduce the following notation.

Definition 6.1. For a set $M = \{M^1, \ldots, M^n\}$ of $k \times k$ matrices and $p, q \in \{1, \ldots, k\}$, define

$$M_{pq} = \left(M_{pq}^1, \ldots, M_{pq}^n\right) \in \mathbb{R}^n.$$

When no confusion may arise, for a vector $w = (w_1, \ldots, w_n) \in \mathbb{R}^n$, we write $w \cdot M$ to denote the $k \times k$ matrix such that

$$(w \cdot M)_{pq} = w \cdot M_{pq}.$$

That is, $w \cdot M = \sum_{i=1}^n w_i M^i$.

6	2	1	4
5	0	3	7
9	3	2	4
5	6	4	8

M^1

1	7	8	8
4	0	6	2
8	7	0	2
3	5	2	6

M^2

4	5	3	8
1	2	9	2
9	4	6	1
0	7	3	7

M^3

3
7
4

M_{32}

8	-1	-1	0
9	-2	-3	3
9	2	-2	7
10	5	5	9

$(2,0,-1) \cdot M$

FIGURE 1. For $k = 4$ and $n = 3$, an example set $M = \{M^1, M^2, M^3\}$, the vector $M_{32} \in \mathbb{R}^3$, and the matrix $(2, 0, -1) \cdot M$.

In light of Equation 2, we are interested in how the persistent homology of $w \cdot M$ may change with respect to varying w. Given a continuous path $p \colon \mathbb{R} \to \mathbb{R}^n$, with components $p(t) = (p_1(t), \ldots, p_n(t))$, the induced map $t \mapsto \mathrm{PD}_d(p(t) \cdot M)$ is an example of a *persistence vineyard* [8], which is a continuous one-parameter family of persistence diagrams.

As t varies, each off-diagonal point in the vineyard traces a curve $\{v(t)\}_t \subset \mathbb{R}_{\leq}^2$ called a *vine*. Each vine has either zero, one, or two endpoints on the diagonal. The *support* of a vine v is $\mathrm{supp}(v) \coloneqq \{t \in \mathbb{R} \mid v(t) \in \mathbb{R}_{<}^2\}$, which is a connected subset of $\mathbb{R}$.
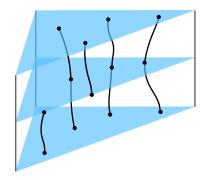

FIGURE 2. An example of a persistence vineyard.

Toward understanding how $\mathrm{PD}_d(p(t) \cdot M)$ may change with respect to varying t, we employ locally constant functions π_b and π_d mapping each point (b_i, d_i) in a persistence diagram to the birth simplex $\pi_\mathrm{b}(b_i, d_i) \coloneqq \sigma_{b_i}$ and the death simplex $\pi_\mathrm{d}(b_i, d_i) \coloneqq \sigma_{d_i}$ of a persistence pair $(\sigma_{b_i}, \sigma_{d_i}) \subset [k] \times [k]$, respectively, where $(b_i, d_i) = (\mathrm{diam}(\sigma_{b_i}), \mathrm{diam}(\sigma_{d_i}))$. As observed in [4], both π_b and π_d can be made to be constant throughout each connected component of $\mathbb{R} \setminus \theta$, where

$$\theta \coloneqq \{t \in \mathbb{R} \mid \text{not all entries of } p(t) \cdot M \text{ above the diagonal are distinct}\}.$$

Indeed, away from θ, there exists a constant total order on the simplices in the Vietoris–Rips filtration compatible with the induced total order on the set of edges. In more detail, fix an arbitrary total order on the set of vertices and break ties between higher-dimensional simplices by listing their edges in descending order and using the lexicographic order on those lists induced by the total order on the edges.

To refer to entries of a $k \times k$ matrix above the diagonal, we define

$$[k]^2_< := \{(p,q) \mid p,q \in \{1,\ldots,k\} \text{ and } p < q\}.$$

We define $[k]^2_\leq$ analogously.

Theorem 2. *Fix a set $M = \{M^i \mid 1 \leq i \leq n\}$ of $k \times k$ positive symmetric matrices with zeros along the diagonal and a continuous path $p \colon \mathbb{R} \to \mathbb{R}^n$. Consider a vine v in the persistence vineyard $\{\mathrm{PD}_d(p(t) \cdot M)\}_{t \in \mathbb{R}}$ and let $t_0 \in \mathrm{supp}(v) \setminus \theta$. Then,*

$$v(t) = \big(p(t) \cdot M_{pq},\ p(t) \cdot M_{p'q'}\big)$$

for all t within the connected component of $\mathrm{supp}(v) \setminus \theta$ containing t_0, where $(p,q), (p',q') \in [k]^2_\leq$ are such that $v(t_0) = \big(p(t_0) \cdot M_{pq},\ p(t_0) \cdot M_{p'q'}\big)$

Proof. Let U_0 denote the connected component of $\mathbb{R} \setminus \theta$ containing t_0. In the case $d = 0$, the birth time b_0 of any feature is zero and we may write $b_0 = \mathrm{diam}\,(\sigma_{b_0}) = p(t_0) \cdot M_{pp} = 0$ for any p.

Otherwise, assume $d > 0$. Then, the birth simplex $\pi_\mathrm{b}(b_0, d_0) = \sigma_{b_0}$ is constant throughout U_0, and because all simplex edge lengths are distinct, it follows that the unique pair $p, q \in \sigma_{b_0}$ such that

$$b_0 = \mathrm{diam}\,(\sigma_{b_0}) = \max\{p(t_0) \cdot M_{rs} \mid r, s \in \sigma_{b_0}\} = p(t_0) \cdot M_{pq}$$

is constant throughout U_0. Similarly, the unique pair $p', q' \in \sigma_{d_0}$ such that

$$d_0 = \mathrm{diam}\,(\sigma_{d_0}) = \max\{p(t_0) \cdot M_{r's'} \mid r', s' \in \sigma_{d_0}\} = p(t_0) \cdot M_{p'q'}$$

is constant throughout U_0. $\qquad\square$

Corollary 6.2. *Fix a set $M = \{M^i \mid 1 \leq i \leq n\}$ of $k \times k$ positive symmetric matrices with zeros along the diagonal. Given an affine function $\ell \colon \mathbb{R} \to \mathbb{R}^n$, it is generically true that each vine in the persistence vineyard $\{\mathrm{PD}_d(\ell(t) \cdot M)\}_{t \in \mathbb{R}}$ is piecewise linear, and each line segment has the form*

$$\big(\ell(t) \cdot M_{pq},\ \ell(t) \cdot M_{p'q'}\big)_{t \in I}$$

for some $(p,q), (p',q') \in [k]^2_\leq$ and open interval $I \subset \mathbb{R}$.

Proof. Not all entries of $\ell(t) \cdot M$ above the diagonal are distinct if and only if $\ell(t) \cdot M_{pq} = \ell(t) \cdot M_{p'q'}$ for $(p,q) \neq (p',q') \in [k]^2_<$. That is, the entries of $\ell(t) \cdot M$ are distinct whenever $\ell(t)$ is not contained in the set

$$\Theta := \bigcup_{(p,q) \neq (p',q') \in [k]^2_<} \big\{w \in \mathbb{R}^n \mid w \cdot (M_{pq} - M_{p'q'}) = 0\big\}$$

Generically, all vectors M_{pq} are distinct, in which case Θ is a union of hyperplanes. The claim now follows from Theorem 2 and the fact that $\ell \colon \mathbb{R} \to \mathbb{R}^n$ generically intersects a hyperplane transversely. $\qquad\square$
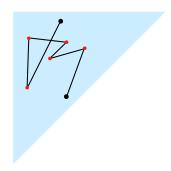
FIGURE 3. A piecewise linear image of a vine in a vineyard $\{\mathrm{PD}_d(\ell(t) \cdot M)\}_{t \in I \subset \mathbb{R}}$. Red dots indicate where the persistence pairing changes.

Remark 6.3. The set Θ defined above is generically a union of $\binom{(k^2-k)/2}{2} = \frac{1}{8}(k^4 - 2k^3 - k^2 + 2k)$ hyperplanes, so the image of each vine consists of at most $\frac{1}{8}(k^4 - 2k^3 - k^2 + 2k)$ distinct line segments. However, in practice, the persistence pairing often does not change as $\ell(t)$ passes through a hyperplane; therefore, one expects each vine to consist of far fewer line segments. See [17] for related ideas in a more general setting.

A version of Corollary 6.2 holds in greater generality. In that setting, given any continuous function $p \colon \mathbb{R} \to \mathbb{R}^n$, it is generically true that each vine decomposes as piecewise linear segments of the form $\big(p(t) \cdot M_{pq}, \ p(t) \cdot M_{p'q'}\big)_{t \in I}$ for some p, q, p', q' and open interval $I \subset \mathbb{R}$.

6.1. **Algorithms.** Corollary 6.2 states that each vine in a persistence vineyard arising from linear combinations of symmetric matrices is entirely determined by the matrix entries corresponding to the diameters of birth and death simplices. In this section, we use this fact to describe an algorithm for feature selection across the geometric simplex.

Recall that the metric structure of a vector-weighted global embedding of a time series may be expressed as a linear combination of distance matrices (Equation 2). Also, recall that we restrict weight vectors to the feasible region $|\Delta^{n-1}|$ of convex coefficients.

As before, we consider the more general setting of linear combinations of positive symmetric $k \times k$ matrices $M = \{M^i \mid 1 \leq i \leq n\}$ with zeros along the diagonal. For the remainder of this section, we will assume the following genericity conditions:

(1) The vectors M_{pq} for $(p, q) \in [k]^2_<$ are distinct.
(2) The initial vector $\lambda \in |\Delta^{n-1}|$ is not contained in any hyperplane $w \cdot (M_{pq} - M_{p'q'}) = 0$ for $(p, q) \neq (p', q') \in [k]^2_<$.

We may implicitly infinitesimally perturb some of the M^i so that these conditions are met. By condition (2) above, note that the entries of $\lambda \cdot M$ are distinct; hence, each point in the persistence diagram $\mathrm{PD}_d(\lambda \cdot M)$ appears with multiplicity one.

Define the *convex hull of M* to be

$$\mathrm{conv}(M) \coloneqq \big\{\lambda \cdot M \mid \lambda \in |\Delta^{n-1}|\big\}.$$

We are interested in how the persistent homology of elements of $\mathrm{conv}(M)$ may change with respect to varying λ. Toward that end, given an initial vector $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathrm{Int}(|\Delta^{n-1}|)$, and $1 \leq j \leq n$,

define the affine function

$$\ell_{\lambda,j} \colon \mathbb{R} \to \mathbb{R}^n$$

$$t \mapsto \left(\lambda_1 - \tfrac{t}{n-1}, \ldots, \lambda_{j-1} - \tfrac{t}{n-1}, \lambda_j + t, \lambda_{j-1} - \tfrac{t}{n-1}, \ldots, \lambda_n - \tfrac{t}{n-1}\right).$$

We restrict the domain of $\ell_{\lambda,j}$ to the interval

$$[-\min\{\min_{i \neq j}(n-1)(1-\lambda_i), \lambda_j\}, \min\{\min_{i \neq j}(n-1)\lambda_i, 1-\lambda_j\}],$$

so the image is a closed line segment contained in $|\Delta^{n-1}|$ with endpoints on the boundary. Note that $\ell_{\lambda,j}|_{t=0} = \lambda$.
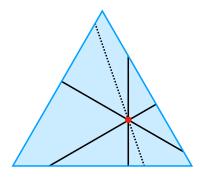
FIGURE 4. The images of $\ell_{\lambda,j}$ for $\lambda = (\tfrac{1}{5}, \tfrac{1}{2}, \tfrac{3}{10})$ and $j = 1, 2, 3$ are drawn as solid lines. The image of the function $\tfrac{1}{4}\ell_{\lambda,1} + 0\ell_{\lambda,2} + \tfrac{3}{4}\ell_{\lambda,3}$ is drawn as a dotted line.

6.1.1. *Derivatives of birth and death coordinates.* Given an initial vector of weights $\lambda \in \mathrm{Int}(|\Delta^{n-1}|)$ and a point $(b_0, d_0) \in \mathrm{PD}_d(\lambda \cdot M)$, let $(p, q), (p', q') \in [k]_<^2$ be such that

$$(b_0, d_0) = \left(\lambda \cdot M_{pq}, \ \lambda \cdot M_{p'q'}\right).$$

Let $\mathbf{1} \in \mathbb{R}^n$ denote the vector of ones. By Corollary 6.2, along $\ell_{\lambda,j}$ we have

$$\frac{db_0}{dt}\Big|_{t=0} = -\tfrac{1}{n-1}M_{pq}^1 - \cdots - \tfrac{1}{n-1}M_{pq}^{j-1} + M_{pq}^j - \tfrac{1}{n-1}M_{pq}^{j+1} - \cdots - \tfrac{1}{n-1}M_{pq}^n$$

$$= \tfrac{n}{n-1}M_{pq}^j - \tfrac{1}{n-1}\mathbf{1} \cdot M_{pq}.$$

Similarly,

$$\frac{dd_0}{dt}\Big|_{t=0} = \tfrac{n}{n-1}M_{p'q'}^j - \tfrac{1}{n-1}\mathbf{1} \cdot M_{p'q'}.$$

The functions $\ell_{\lambda,j}$, $1 \leq j \leq n$, do not encode all possible directions through λ within the simplex. However, birth- and death-derivatives are maximized along these directions.

Proposition 6.4. *Let $\lambda \in \mathrm{Int}(|\Delta^{n-1}|)$ and consider a point $(b_0, d_0) \in \mathrm{PD}_d(\lambda \cdot M)$. Restricted to affine functions $\ell \colon \mathbb{R} \to |\Delta^{n-1}|$ such that $\ell(0) = \lambda$, the magnitude of the derivative $\frac{db_0}{dt}|_{t=0}$ is maximized along $\ell_{\lambda,j}$ for some $1 \leq j \leq n$. Similarly, the magnitude of $\frac{dd_0}{dt}|_{t=0}$ is maximized along $\ell_{\lambda,j'}$ for some $1 \leq j' \leq n$.*

Proof. Let $(p, q) \in [k]^2_<$ be such that $b_0 = \lambda \cdot M_{pq}$. Locally, any affine function within the simplex and passing through λ may be expressed as a convex combination of the functions $\ell_{\lambda, j}$ (see Figure 4). Along such a function $\sum_{i=1}^n \gamma_i \ell_{\lambda, i}$, for some convex coefficients $\gamma = (\gamma_1, \ldots, \gamma_n)$, it follows that

$$
\begin{aligned}
\tfrac{db_0}{dt}\big|_{t=0} &= \sum_{i=1}^n \gamma_i \left(\tfrac{n}{n-1} M_{pq}^i - \tfrac{1}{n-1} \mathbf{1} \cdot M_{pq} \right) \\
&= \tfrac{n}{n-1} \sum_{i=1}^n \gamma_i M_{pq}^i - \tfrac{1}{n-1} \mathbf{1} \cdot M_{pq} \\
&= \tfrac{n}{n-1} \gamma \cdot M_{pq} - \tfrac{1}{n-1} \mathbf{1} \cdot M_{pq} \\
&= \tfrac{1}{n-1} (n\gamma - \mathbf{1}) \cdot M_{pq}
\end{aligned}
$$

Therefore, to maximize the magnitude of $\frac{db_0}{dt}$, we solve

$$
\text{maximize} \quad |(n\gamma - \mathbf{1}) \cdot M_{pq}|
$$
$$
\text{subject to} \quad \gamma \in |\Delta^{n-1}|.
$$

Letting $\{e_1, \ldots, e_n\} \subset \mathbb{R}^n$ denote the standard basis vectors, it is clear the solution γ^* occurs at a vertex of the simplex, that is, $\gamma^* = e_j$ such that $j = \arg\max_i |nM_{pq}^i - \mathbf{1} \cdot M_{pq}|$. In particular, either $j = \arg\min_i M_{pq}^i$ or $j = \arg\max_i M_{pq}^i$. Hence, the magnitude of the derivative $\frac{db_0}{dt}$ is maximized along $\sum_{i=1}^n \gamma_i \ell_{\lambda, i} = \ell_{\lambda, j}$. The argument for $\frac{dd_0}{dt}$ is identical. $\square$

6.1.2. *Optimizing for a lifetime.* With Proposition 6.4 in mind, we now summarize a gradient descent method to optimize for the lifetime of a given point in the persistence diagram:

(1) Given an initial vector of weights $\lambda \in |\Delta^{n-1}|$, compute the persistence diagram of the corresponding convex combination of matrices, that is, $\mathrm{PD}_d(\lambda \cdot M)$.

(2) For $(b_0, d_0) \in \mathrm{PD}_d(\sum_{i=1}^n \lambda_i M^i)$, let $(p, q), (p', q') \in [k]^2_\leq$ be such that

$$
(b_0, d_0) = \left(\lambda \cdot M_{pq}, \ \lambda \cdot M_{p'q'} \right).
$$

(3) Define $j := \arg\max_i |n(M_{p'q'}^j - M_{pq}^j) - \mathbf{1} \cdot (M_{p'q'} - M_{pq})|$.

(4) The derivative at λ is maximized along $\ell_{\lambda, j}$ and is given by the formula

$$
\tfrac{d(d_0 - b_0)}{dt} = \tfrac{1}{n-1}(ne_j - \mathbf{1}) \cdot (M_{p'q'} - M_{pq}).
$$

(5) For a (potentially variable) learning rate $c > 0$, update λ by the rule

$$
\lambda \mapsto \ell_{\lambda, j}\left(c \cdot \tfrac{d(d_0 - b_0)}{dt} \right).
$$

Recompute persistence and update (b_0, d_0).

(6) Iterate steps $(2) - (5)$ until a stopping condition is met.

Note that this algorithm is efficient: after computing persistence, we simply identify the entries in the matrix $\lambda \cdot M$ having the coordinates of a given point in the persistence diagram before updating λ in linear time. In greater generality, we may optimize for an arbitrary linear combination of birth and death times of various points in a similar fashion.

Remark 6.5. The algorithm described above may be modified to incorporate stochasticity, which may be appropriate for applications to noisy or large data sets. In particular, each index i may be assigned a probability determined by the magnitude of $|n(M_{p'q'}^i - M_{pq}^i) - \mathbf{1} \cdot (M_{p'q'} - M_{pq})|$ relative to the sum of these magnitudes; we then update index j in step (3) according to these probabilities.

7. Examples

In this section, we return to the setting of vector-weighted global embeddings of time series. We consider synthetic and biological data sets and apply the methods described in Section 6 to assign scores to individual time series based on their contribution to global dynamics.

7.1. **Synthetic data.** The following is a low-dimensional example meant to illustrate the methods described in Section 6. Consider the time series $f \colon \mathbb{R} \to \mathbb{R}^3$ with components

$$f_i(t) = \sin\left(\tfrac{2\pi}{50}(t - K_i)\right) + b_i(t), \qquad i = 1, 2, 3,$$

where each $K_i \in \mathbb{R}$ is a random phase shift and $b_i(t)$ is sampled uniformly from the interval $[-5, 5]$ for each $t \in \mathbb{R}$. These functions, together with the underlying sine functions, are displayed in Figure 5 (top). Restricted to the finite domain $D = \{1, \ldots, 300\}$ with window length $L = 200$, we obtain the sliding window point cloud $\mathrm{SW}[f|_D]$ consisting of 101 points in $\mathbb{R}^{600}$. The projection of the point cloud onto the first two principle components is shown in Figure 5 (bottom left). Despite the addition of noise, circularity due to the underlying periodicity is apparent.
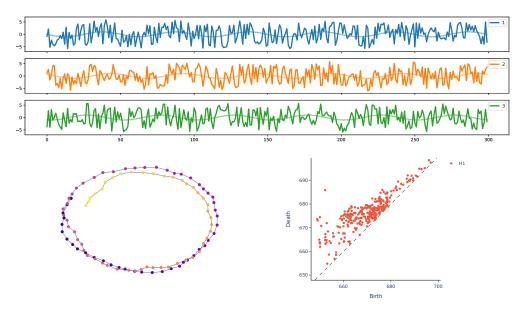

FIGURE 5. Three noisy sine waves, the PCA projection of the sliding window embedding, and the persistence diagram of the high-dimensional embedded data.

Next, we compute the persistent homology of $\mathrm{SW}[f|_D]$ and plot the resulting persistence diagram in Figure 5 (bottom right). We apply the algorithm described in Subsection 6.1.2 to optimize for the lifetime of the longest-lived point in the diagram (that is, *maximal persistence*). The path through the geometric 2-simplex leading to a local maximum is visualized in Figure 6. After 15 steps, maximal persistence is found to occur near weight vector $\lambda^* = (0, 0.453, 0.547)$. For comparison, we compute maximal persistence for each singleton embedding, that is, at each vertex of the simplex:

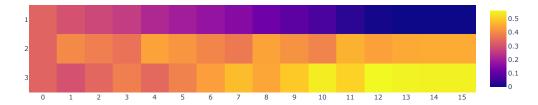| weight vector	$(1,0,0)$	$(0,1,0)$	$(0,0,1)$	$\lambda^* = (0, 0.453, 0.547)$
maximal persistence	39.391	42.972	42.406	46.296

FIGURE 6. Rows correspond to singleton time series, and colors indicate weights on each. The algorithm begins at the barycenter $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ of the 2-simplex, reaches an edge at step 13, then converges near the barycenter of that edge.

7.2. **Neuronal data.** The example in this section is based on GCaMP6s activity traces of 72 neurons within the model organism *C. elegans* obtained by Chaudhary et. al [5]. This data set consists of a 278-second long recording at a resolution of one frame per second. Periodic stimulus (five seconds on, five seconds off) was applied starting at 100 seconds and ending at 180 seconds. Figure 7 shows a heatmap of the data set.

The sliding window embedding with window length $L = 15$ yields a point cloud consisting of 264 points in $\mathbb{R}^{1080}$. The persistence diagram and PCA projection of this point cloud is displayed in Figure 9 (left). There are two large loops detected by persistence; indeed, the periodic stimulus applied throughout the middle third of the recording dramatically affects the trajectory of the orbit, forming the second loop. A stochastic version of the algorithm described in Subsection 6.1.2 was iterated 500 times to find weights increasing the lifetimes of both loops, and the resulting path through the geometric simplex is visualized in Figure 8. The resulting persistence diagram and PCA projection of the optimally-weighted data is shown in Figure 9 (right).
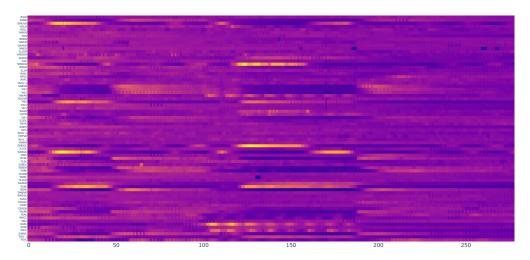

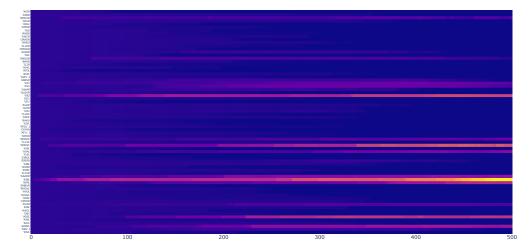FIGURE 7. Heatmap of the *C. elegans* neuronal data set.

FIGURE 8. In analogy with Figure 6, a visualization of the path through the geometric 71-simplex leading to weights increasing the lifetimes of both loops. After 500 iterations, 14 of the 72 neurons have nonzero weight; the persistence diagram and PCA projection of this weighted embedding is shown in Figure 9 (right).
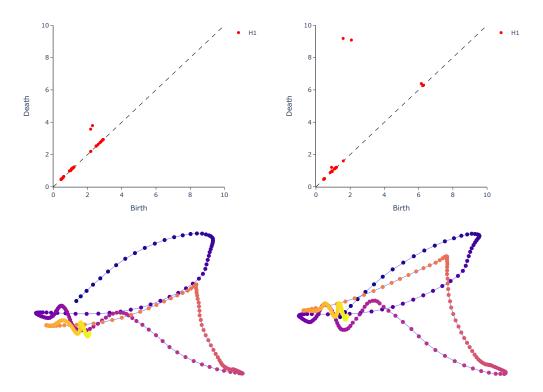

FIGURE 9. Persistence diagrams and PCA projections of the sliding window embeddings of the equally-weighted time series (left) and the optimally-weighted time series (right).

Acknowledgements

References

[1] Jesse J Berwald, Marian Gidea, and Mikael Vejdemo-Johansson. Automatic recognition and tagging of topologically different regimes in dynamical systems. *Discontinuity, Nonlinearity, and Complexity*, 3(4):413–426, 2014.

[2] Rickard Brüel-Gabrielsson, Vignesh Ganapathi-Subramanian, Primoz Skraba, and Leonidas J Guibas. Topology-aware surface reconstruction for point clouds. In *Computer Graphics Forum*, volume 39, pages 197–207. Wiley Online Library, 2020.

[3] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33. American Mathematical Society, 2022.

[4] Mathieu Carriere, Frédéric Chazal, Marc Glisse, Yuichi Ike, Hariprasad Kannan, and Yuhei Umeda. Optimizing persistent homology based functions. In *International conference on machine learning*, pages 1294–1303. PMLR, 2021.

[5] Shivesh Chaudhary, Sol Ah Lee, Yueyi Li, Dhaval S Patel, and Hang Lu. Graphical-model framework for automated annotation of cell identities in dense cellular images. *Elife*, 10:e60321, 2021.

[6] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J Guibas, and Steve Y Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 237–246, 2009.

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

[8] David Cohen-Steiner, Herbert Edelsbrunner, and Dmitriy Morozov. Vines and vineyards by updating persistence in linear time. In *Proceedings of the 22nd annual Symposium on Computational Geometry (SoCG)*, pages 119–126, 2006.

[9] Mary-Lee Dequeant, Sebastian Ahnert, Herbert Edelsbrunner, Thomas MA Fink, Earl F Glynn, Gaye Hattem, Andrzej Kudlicki, Yuriy Mileyko, Jason Morton, Arcady R Mushegian, et al. Comparison of pattern detection methods in microarray time series of the segmentation clock. *PLoS One*, 3(8):e2856, 2008.

[10] Saba Emrani, Thanos Gentimis, and Hamid Krim. Persistent homology of delay embeddings and its application to wheeze detection. *IEEE Signal Processing Letters*, 21(4):459–463, 2014.

[11] Hitesh Gakhar and Jose A Perea. Sliding window persistence of quasiperiodic functions. *Journal of Applied and Computational Topology*, pages 1–38, 2023.

[12] Marian Gidea, Daniel Goldsmith, Yuri Katz, Pablo Roldan, and Yonah Shmalo. Topological recognition of critical transitions in time series of cryptocurrencies. *Physica A: Statistical mechanics and its applications*, 548:123843, 2020.

[13] Alec Helm, Ann S Blevins, and Danielle S Bassett. The growing topology of the *C. elegans* connectome. *arXiv preprint arXiv:2101.00065*, 2020.

[14] Saul Kato, Harris S Kaplan, Tina Schrödel, Susanne Skora, Theodore H Lindsay, Eviatar Yemini, Shawn Lockery, and Manuel Zimmer. Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans*. *Cell*, 163(3):656–669, 2015.

[15] Jacob Leygonie, Steve Oudot, and Ulrike Tillmann. A framework for differential calculus on persistence barcodes. *Foundations of Computational Mathematics*, pages 1–63, 2021.

[16] Michael Moor, Max Horn, Bastian Rieck, and Karsten Borgwardt. Topological autoencoders. In *International conference on machine learning*, pages 7045–7054. PMLR, 2020.

[17] Arnur Nigmetov and Dmitriy Morozov. Topological optimization with big steps. *arXiv preprint arXiv:2203.16748*, 2022.

[18] Jose A Perea. Topological time series analysis. *Notices of the American Mathematical Society*, 66(5):686–694, 2019.

[19] Jose A Perea, Anastasia Deckard, Steve B Haase, and John Harer. Sw1pers: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data. *BMC Bioinformatics*, 16(1):1–12, 2015.

[20] Jose A Perea and John Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15:799–838, 2015.

[21] Adrien Poulenard, Primoz Skraba, and Maks Ovsjanikov. Topological function optimization for continuous shape matching. In *Computer Graphics Forum*, volume 37, pages 13–25. Wiley Online Library, 2018.

[22] Floris Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80*, pages 366–381. Springer, 2006.

[23] Ashleigh Thomas, Kathleen Bates, Alex Elchesen, Iryna Hartsock, Hang Lu, and Peter Bubenik. Topological data analysis of *C. elegans* locomotion and behavior. *Frontiers in Artificial Intelligence*, 4:668395, 2021.

[24] Christopher Tralie. High dimensional geometry of sliding window embeddings of periodic videos. In *Proceedings of the 32st International Symposium on Computational Geometry (SOCG)*, 2016.

[25] Boyan Xu, Christopher J Tralie, Alice Antia, Michael Lin, and Jose A Perea. Twisty Takens: A geometric characterization of good observations on dense trajectories. *Journal of Applied and Computational Topology*, 3:285–313, 2019.

[26] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 347–356, 2004.

Appendix A. Continuity of global embeddings

In this section, we address continuity of the global embedding defined by $v \mapsto \mathrm{SW}[v \circ f]$. In more detail, fix a time series $f \colon \mathbb{R} \to \mathbb{R}^n$, a finite subset $D \subset \mathbb{R}$, window parameters L and τ, and consider the map

$$\mathrm{SW}[- \circ f|_D] \colon (\mathbb{R}^n, \|\cdot\|_p) \to (\mathrm{Met}_{\mathrm{fin}}, d_{\mathrm{GH}})$$
$$v \mapsto (\mathrm{SW}[v \circ f|_D], \|\cdot\|_q).$$

Here, $(\mathrm{Met}_{\mathrm{fin}}, d_{\mathrm{GH}})$ denotes the collection of finite metric spaces equipped with the Gromov–Hausdorff distance. After defining d_{GH}, we prove $\mathrm{SW}[- \circ f|_D]$ is Lipschitz.

The Gromov–Hausdorff distance d_{GH} is an extended pseudo-metric in general, but it restricts to a metric on the set of isometry classes of compact metric spaces. In particular, it is a metric on the collection of finite metric spaces. We define the Gromov–Hausdorff distance in terms of the notion of the Hausdorff distance.

Definition A.1. Let A and B denote two nonempty subsets of a metric space (Z, d_Z). The *Hausdorff distance* between A and B, denoted $d_{\mathrm{H}}(A, B)$, is defined by

$$d_{\mathrm{H}}(A, B) := \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b) \right\}.$$

If A and B are finite subsets of a metric space (Z, d_Z) of equal cardinality, note that

$$d_{\mathrm{H}}(A, B) = \min_{\substack{\text{bijections} \\ f \colon A \to B}} \max_{a \in A} d_Z(a, f(a)). \tag{3}$$

Definition A.2. Given two metric spaces (X, d_X) and (Y, d_Y), the *Gromov–Hausdorff distance* between them is defined by

$$d_{\mathrm{GH}}(X, Y) := \inf_{\substack{\text{metric spaces} \\ (Z, d_Z)}} \{d_{\mathrm{H}}(i_X(X), i_Y(Y)) \mid i_X \colon X \hookrightarrow Z \text{ and } i_Y \colon Y \hookrightarrow Z \text{ isometric embeddings}\}.$$

Remark A.3. Definition A.2 contains an abuse of notation in the sense that the collection of all metric spaces is a proper class. However, the infimum is achieved because it is sufficient to restrict to a set of certain pseudo-metrics on the disjoint union $X \sqcup Y$; see [3] for details.

Lemma A.4. *Given a time series $f \colon \mathbb{R} \to \mathbb{R}^n$ with components $f(t) = (f_1(t), \ldots, f_n(t))$, a finite subset $D \subset \mathbb{R}$, and window parameters L and τ, the map*

$$\mathrm{SW}[- \circ f|_D] \colon (\mathbb{R}^n, \|\cdot\|_p) \to (\mathrm{Met}_{\mathrm{fin}}, d_{\mathrm{GH}})$$

$$v \mapsto (\mathrm{SW}[v \circ f|_D], \|\cdot\|_q).$$

is K-Lipschitz with Lipschitz constant $K \le \displaystyle\max_{1 \le i \le n} \max_{t \in D} \|f_i(t)\|_q L^{1/p}$.

Proof. As an intermediate step, consider the map $v \mapsto \{v \circ f(t) \mid t \in D\}$. Given $v, w \in \mathbb{R}^n$, observe

$$d_{\mathrm{GH}}\left(\{v \circ f(t) \mid t \in D\}, \{w \circ f(t) \mid t \in D\}\right)^q \le d_{\mathrm{H}}\left(\{v \circ f(t) \mid t \in D\}, \{w \circ f(t) \mid t \in D\}\right)^q$$

$$\le \max_{t \in D} \left\| v \circ f(t) - w \circ f(t) \right\|_q^q \qquad \text{(by Equation 3)}$$

$$= \max_{t \in D} \left\| \begin{bmatrix} v_1 f_1(t) - w_1 f_1(t) & \cdots & v_n f_n(t) - w_n f_n(t) \end{bmatrix} \right\|_q^q$$

$$= \max_{t \in D} \sum_{j=1}^n \|v_j - w_j\|_q^q \|f_j(t)\|_q^q$$

$$\le \max_{1 \le i \le n} \max_{t \in D} \|f_i(t)\|_q^q \sum_{j=1}^n \|v_j - w_j\|_q^q$$

$$= \max_{1 \le i \le n} \max_{t \in D} \|f_i(t)\|_q^q \|v - w\|_q^q.$$

Hence, $v \mapsto \{v \circ f(t) \mid t \in D\}$ is $\displaystyle\max_{1 \le i \le n} \max_{t \in D} \|f_i(t)\|_q$-Lipschitz. Finally, it is straightforward to check that the sliding window embedding is $L^{1/p}$-Lipschitz. $\square$

The collection of persistence diagrams PD is commonly equipped with the *bottleneck distance* d_{B}. Given persistence diagrams P and Q, let π_P and π_Q denote the projections onto the first and second factors of $P \times Q$, respectively. A *partial matching* between P and Q is a subset $M \subset P \times Q$ such that $|\pi_P^{-1}(p)| \le 1$, and $|\pi_Q^{-1}(q)| \le 1$ for every $p \in P$ and every $q \in Q$. Given a point $r = (b, d)$ in either diagram, define $\overline{r} := \left(\frac{b+d}{2}, \frac{b+d}{2}\right)$, which is the point on the diagonal closest to r. Finally, define

$$M_P := \{(p, \overline{p}) \mid \pi_P^{-1}(p) = \varnothing\} \qquad \text{and} \qquad M_Q := \{(q, \overline{q}) \mid \pi_Q^{-1}(q) = \varnothing\}$$

and call the set $\overline{M} := M \cup M_P \cup M_Q$ the *induced matching*.

Definition A.5. The *bottleneck distance* between persistence diagrams P and Q is

$$d_{\mathrm{B}}(P, Q) := \inf_M \sum_{(p,q) \in \overline{M}} \|p - q\|_\infty,$$

where the infimum is taken over all partial matchings $M \subset P \times Q$ and $\overline{M}$ denotes the induced matching.

Vietoris–Rips persistent homology, as a map $\mathrm{PH}_d \circ \mathrm{VR} \colon (\mathrm{Met}_{\mathrm{fin}}, d_{\mathrm{GH}}) \to (\mathrm{PD}, d_{\mathrm{b}})$, is known to be 2-Lipschitz [6]. Hence, the composition $\mathrm{PH}_d \circ \mathrm{VR} \circ \mathrm{SW}[- \circledcirc f|_D]$ is Lipschitz and is consequently differentiable almost everywhere.

(PB) Department of Mathematics, University of Florida, Gainesville, FL 32611, USA
Email address: peter.bubenik@ufl.edu

(JB*) Department of Mathematics, University of Florida, Gainesville, FL 32611, USA
Email address: bush.j@ufl.edu

*Corresponding author.