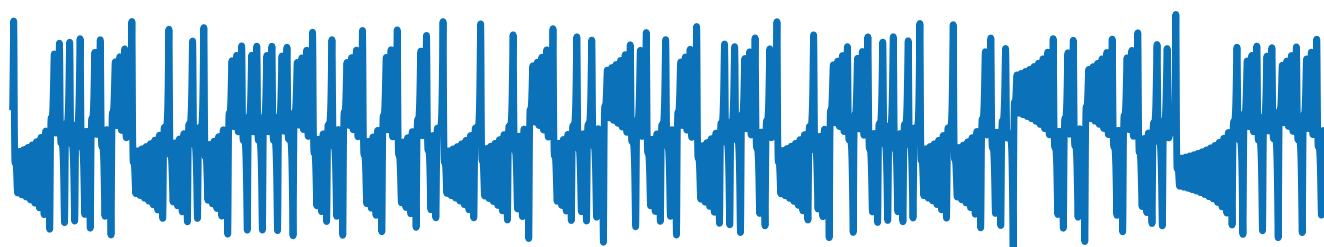
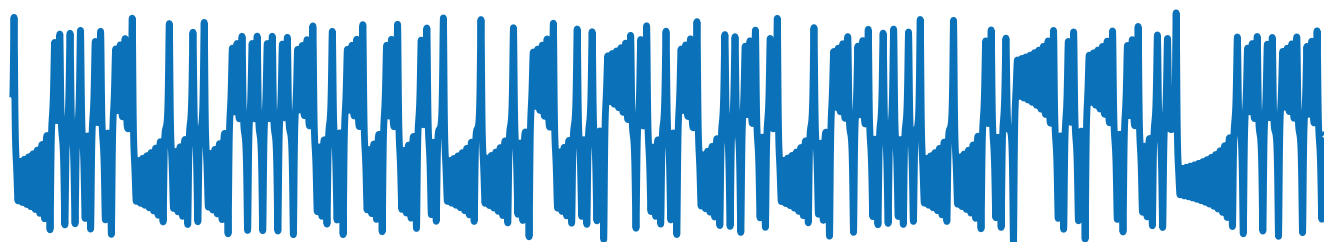
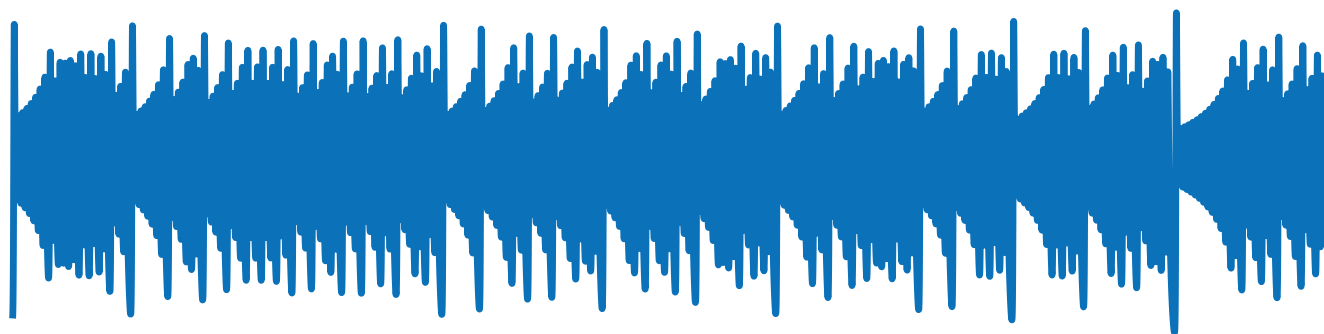

Topological Time Series Analysis



Jose A. Perea

Jose Perea is an assistant professor in the Department of Computational Mathematics, Science & Engineering, and the Department of Mathematics at Michigan State University. His email address is joperea@msu.edu.

Communicated by Notices Associate Editor Emilie Purvine.

For permission to reprint this article, please contact:
reprint-permission@ams.org.

DOI: <https://doi.org/10.1090/noti1869>

Introduction

Time series are ubiquitous in our data rich world. In what follows I will describe how ideas from dynamical systems and topological data analysis can be combined to gain insights from time-varying data. We will see several applications to engineering and the life sciences, as well as some of the theoretical underpinnings.

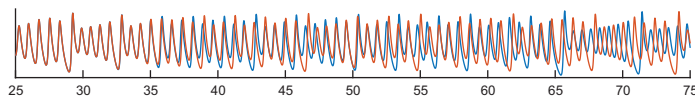


Figure 1. Orange: results from the simulation initialized at \mathbf{v}_0 ; blue: results after manually restarting the simulation from \mathbf{v}_j .

Lorenz and the Butterfly

Imagine you have a project involving a crucial computer simulation. For an initial value $\mathbf{v}_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$, a sequence $\mathbf{v}_0, \dots, \mathbf{v}_n \in \mathbb{R}^3$ is computed in such a way that \mathbf{v}_{j+1} is determined from \mathbf{v}_j for $j = 0, \dots, n-1$. After the simulation is complete you realize that a rerun is needed for further analysis. Instead of initializing at \mathbf{v}_0 , which might take a while, you take a shortcut: you input a value \mathbf{v}_j selected from the middle of the current results, and the simulation runs from there while you go for coffee. Figure 1 is displayed on the computer monitor upon your return; the orange curve is the sequence x_0, \dots, x_n from the initial simulation, and the blue curve is the x coordinate for the rerun initialized at \mathbf{v}_j .

The results agree at first, but then they diverge widely; what is going on? Edward Norton Lorenz, a mathematical meteorologist, asked himself the very same question while studying a simplified model for weather forecasting [19]. In the process of resolving the aforementioned discrepancy, which one could erroneously attribute to software error or a hardware malfunction, Lorenz laid out the foundations for what we know today as chaos theory. The relevant set of differential equations for the simplified model, called the Lorenz system, is shown in equation (1); x, y , and z are real-valued functions of time t , and $\sigma, \rho, \beta \in \mathbb{R}$ are physical constants.

$$\begin{aligned} x'(t) &= \sigma \cdot (y - x) \\ y'(t) &= x \cdot (\rho - z) - y \\ z'(t) &= xy - \beta z \end{aligned} \quad (1)$$

Solving the Lorenz system yields a differentiable function

$$\Phi : \mathbb{R} \times \mathbb{R}^3 \rightarrow \mathbb{R}^3 \quad (2)$$

where $\Phi(t, \mathbf{v}_0) = (x(t), y(t), z(t))$ satisfies (1) for all $t \in \mathbb{R}$, and $\Phi(0, \mathbf{v}_0) = \mathbf{v}_0$ for all $\mathbf{v}_0 \in \mathbb{R}^3$. In fact, the orange curve from Figure 1 corresponds to $\Phi(t, (5, 5, 5))$ when $(\sigma, \rho, \beta) = (10, 28, 8/3)$. The discrepancy between

the orange and blue curves, as elucidated in [19], is a property inherent to the system. Lorenz realized that when manually entering \mathbf{v}_j as input, he only used the first few significant digits instead of the full precision values. In other words, the system (1) can be extremely sensitive to initial conditions in that any errors are compounded exponentially with time.

This behavior is known today as the *Butterfly Effect*. The metaphor is that even the tiniest change in initial atmospheric conditions, such as a butterfly flapping its wings in Brazil, may change the long-term evolution of weather patterns enough to produce a tornado in Texas. Such unpredictability is one of the hallmarks of a *chaotic dynamical system*, and speaks to the futility of long-term weather prediction. The butterfly metaphor is further amplified by the shape of the solution $t \mapsto \Phi(t, (5, 5, 5)) = (x(t), y(t), z(t))$, $0 \leq t \leq 200$, shown in the left-hand side of Figure 2.

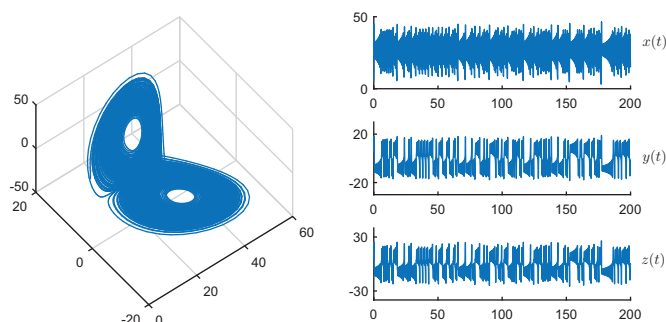


Figure 2. Lorenz's butterfly attractor

Dynamical systems are mathematical abstractions of time-dependent physical processes. Intuitively speaking, a dynamical system consists of two pieces of data: a set of states M — e.g., all possible atmospheric conditions at a given location on earth — along with rules $\Phi = \{\Phi_p : p \in M\}$ describing how each state $p \in M$ changes over time. More specifically,

Definition 0.1. A *global continuous time dynamical system* is a pair (M, Φ) , where M is a topological space and $\Phi : \mathbb{R} \times M \rightarrow M$ is a continuous map so that $\Phi(0, p) = p$, and $\Phi(s, \Phi(t, p)) = \Phi(s + t, p)$ for all $p \in M$ and all $t, s \in \mathbb{R}$.

The typical examples arising from differential equations (e.g., the Lorenz system) have as state space a smooth manifold M (e.g., \mathbb{R}^3), and the dynamics are given by the integral curves (e.g., equation (2)) of a smooth vector field on M (e.g., (1)).

Some subsets of M are especially important since they attract the evolution of states in close proximity. Indeed, a set $A \subset M$ is called an *attractor* (these are the kinds of sets we will focus on) if it satisfies three conditions: (1) it

is compact, (2) it is an invariant set — that is, if $a \in A$ then $\Phi(t, a) \in A$ for all $t \geq 0$ — and (3) it has an open basin of attraction. In other words, there is an invariant open neighborhood $U \subset M$ of A , so that

$$\bigcap_{t \geq 0} \{\Phi(t, p) : p \in U\} = A.$$

An attractor A is called strange — and this is just a name — if (1) there are arbitrarily close points p, p' in a basin of attraction of A for which the distance between $\Phi(t, p)$ and $\Phi(t, p')$ grows exponentially quickly with t , and (2) A has non-integral Hausdorff dimension. Lorenz's butterfly $\Lambda \subset \mathbb{R}^3$ from Figure 2 is one of the most widely known examples of a strange attractor; its Hausdorff dimension is approximately 2.063 [31].

Persistent Homology: Measuring Shape From Finite Samples

The shape of attractors carries a great deal of information about the global structure of a dynamical system. Indeed, attractors with the shape of a circle $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ give rise to periodic processes; non-integral Hausdorff dimension is evidence of chaotic behavior; while high-dimensional tori $\mathbb{T}^n = S^1 \times \cdots \times S^1$ are linked to quasiperiodicity. The latter is a type of recurrence emerging from the superposition of periodic oscillators with *incommensurate* (i.e., \mathbb{Q} -linearly independent) frequencies. Quasiperiodicity appears, for example, in turbulent fluids [26], and the detection of biphonation in high-speed laryngeal videendoscopy [30]. Similarly, the existence of chaos in brain activity is of considerable interest in neuroscience [17], as is the presence of periodic oscillators in biological systems [27].

A data analysis question that has received significant attention in recent years is how to measure the shape of a topological space \mathbb{X} — e.g., an attractor — from a finite set X (the data) approximating it. This is the type of problem driving advances in Topological Data Analysis [7], and where tools like persistent homology [22] — which we will describe shortly — are relevant.

The space \mathbb{X} around which the observed data X accumulates is unknown in practice, so the typical strategy in topological data analysis is to use X to construct another space whose shape approximates that of \mathbb{X} . One of the most widely used constructions is the *Rips complex*. If X (not necessarily finite) comes equipped with a metric ρ , then the Rips complex of X at scale α is the set

$$R_\alpha(X) := \{\{x_0, \dots, x_k\} \subset X : \rho(x_i, x_j) \leq \alpha \text{ for all } 0 \leq i, j \leq k\} \quad (3)$$

This is in fact a *simplicial complex*; points in X can be thought of as vertices, sets with two elements $\{x_0, x_1\} \in R_\alpha(X)$ are edges, $\{x_0, x_1, x_2\} \in R_\alpha(X)$ is a triangular face, and

so on. A theorem of Janko Latschev [18] contends that if \mathbb{X} is a closed Riemannian manifold, then under mild density hypotheses (which unfortunately cannot be checked in practice) the geometric realization of $R_\alpha(X)$ is homotopy equivalent to \mathbb{X} for small $\alpha > 0$.

The shape of a space (i.e., its homotopy type) refers to those properties that are invariant under continuous deformations; e.g., is it connected? are there holes? Said properties can be formalized and quantified using *homology* [13]. Given an integer $n \geq 0$, the n -th homology of a topological space B with coefficients in a field \mathbb{F} , denoted $H_n(B; \mathbb{F})$, is a vector space over \mathbb{F} . Its dimension $\beta_n(B; \mathbb{F})$ — the n -th *Betti number* of B with coefficients in \mathbb{F} — provides a count for the number of n -dimensional holes in B . Indeed, β_0 counts the number of path-connected components, β_1 is the number of holes bounded by a closed loop in B , β_2 is the number of voids bounded by a closed 2-dimensional region, and so on for $n \geq 3$. Here is an example: the 2-dimensional torus $\mathbb{T}^2 = S^1 \times S^1$ in Figure 3 (left) has Betti numbers $\beta_0(\mathbb{T}^2; \mathbb{F}) = 1$ since it is path-connected, $\beta_1(\mathbb{T}^2; \mathbb{F}) = 2$ since it has a horizontal and a vertical hole, $\beta_2(\mathbb{T}^2; \mathbb{F}) = 1$ since \mathbb{T}^2 itself encloses an empty volume, and $\beta_n(\mathbb{T}^2; \mathbb{F}) = 0$ for all $n \geq 3$. Similarly, the 2-sphere $S^2 = \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| = 1\}$ has Betti numbers $\beta_0(S^2; \mathbb{F}) = \beta_2(S^2; \mathbb{F}) = 1$ and $\beta_n(S^2; \mathbb{F}) = 0$ for $n \geq 1$, for the same reasons as the torus, but $\beta_1(S^2; \mathbb{F}) = 0$ since every closed loop on S^2 bounds a filled-in region.

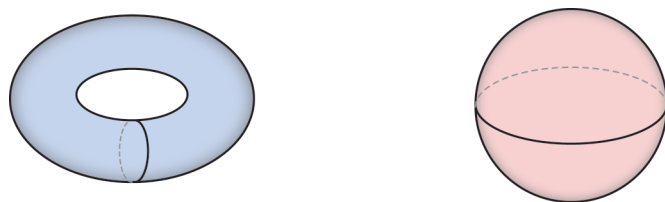


Figure 3. Left: The 2-dimensional torus $\mathbb{T}^2 = S^1 \times S^1$. Right: The 2-sphere $S^2 = \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| = 1\}$.

We remark that $\beta_n(B; \mathbb{F})$ can change with the choice of field \mathbb{F} . Indeed, if B is a closed connected n -dimensional manifold, $p \geq 3$ is a prime, and \mathbb{F}_p denotes the field with p elements, then B is orientable if and only if $\beta_n(B; \mathbb{F}_p) = 1$, and non-orientable if and only if $\beta_n(B; \mathbb{F}_p) = 0$ (see for example 3.26 and 3.28 in [13]).

For a realistic data set X , the Betti numbers $\alpha \mapsto \beta_n(R_\alpha(X); \mathbb{F})$ are expected to be unstable as α varies. Indeed, sampling artifacts or noise in X can produce holes that are present in $R_\alpha(X)$ but not in $R_{\alpha+\delta}(X)$ for small $\delta > 0$ (e.g., see Figure 4). This is where comparing the homology of spaces related via maps is useful. If $\alpha \leq \alpha'$, then $\iota^{\alpha, \alpha'} : R_\alpha(X) \hookrightarrow R_{\alpha'}(X)$ given by $\iota^{\alpha, \alpha'}(\{x_0, \dots, x_k\}) = \{x_0, \dots, x_k\}$ induces a linear transformation

$$\iota_n^{\alpha, \alpha'} : H_n(R_\alpha(X); \mathbb{F}) \longrightarrow H_n(R_{\alpha'}(X); \mathbb{F}). \quad (4)$$

Classes in $H_n(R_\alpha(X); \mathbb{F})$ that are not in the Kernel of $\iota_n^{\alpha, \alpha'}$ for large $\alpha' - \alpha$ can thus be interpreted as being persistent in the data, and suggest true homological features of the underlying space \mathbb{X} . The collection of vector spaces and linear maps resulting from (4) is called the n -dimensional *persistent homology*, with coefficients in \mathbb{F} , of the *Rips filtration* $\mathcal{R}(X) = \{R_\alpha(X)\}_{\alpha \in \mathbb{R}}$. Thus far we have that the Betti numbers capture the homology of a space, yielding succinct shape descriptors for its topology. The persistent homology of $\mathcal{R}(X)$, on the other hand, describes the multiscale evolution of homological features underlying the data. More generally,

Definition 0.2. A persistence vector space \mathbf{V} is a collection of vector spaces V_α , $\alpha \in \mathbb{R}$, and linear transformations $\iota^{\alpha, \alpha'} : V_\alpha \rightarrow V_{\alpha'}$, $\alpha \leq \alpha'$, so that:

1. $\iota^{\alpha, \alpha}$ is the identity of V_α for every $\alpha \in \mathbb{R}$.
2. $\iota^{\alpha', \alpha''} \circ \iota^{\alpha, \alpha'} = \iota^{\alpha, \alpha''}$, whenever $\alpha \leq \alpha' \leq \alpha''$.

Two persistence vector spaces $\mathbf{V} = \{V_\alpha, \iota^{\alpha, \alpha'}\}$ and $\mathbf{W} = \{W_\alpha, \kappa^{\alpha, \alpha'}\}$ are isomorphic, denoted $\mathbf{V} \cong \mathbf{W}$, if there are linear isomorphisms $T_\alpha : V_\alpha \rightarrow W_\alpha$ for all $\alpha \in \mathbb{R}$, so that $\kappa^{\alpha, \alpha'} \circ T_\alpha = T_{\alpha'} \circ \iota^{\alpha, \alpha'}$ whenever $\alpha \leq \alpha'$.

We will concentrate on three quantities for a nonzero element $y \in V_\alpha$:

$$\text{birth}(y) := \inf \left\{ \tilde{\alpha} \leq \alpha : y \in \text{Image} \left(\iota^{\tilde{\alpha}, \alpha} \right) \right\} \quad (5)$$

$$\text{death}(y) := \sup \left\{ \alpha' \geq \alpha : y \notin \text{Kernel} \left(\iota^{\alpha, \alpha'} \right) \right\}$$

$$\text{persistence}(y) := \text{death}(y) - \text{birth}(y)$$

When each V_α is finite dimensional, the isomorphism type of \mathbf{V} can be completely described via a simple invariant called the *barcode* [9]:

Theorem 0.3. Let $\mathbf{V} = \{V_\alpha, \iota^{\alpha, \alpha'}\}$ be a persistence vector space so that $\dim(V_\alpha)$ is finite for all $\alpha \in \mathbb{R}$. Then, there exists a multiset (i.e., a set whose elements may appear with repetitions) of intervals $I \subset [-\infty, \infty]$ called the *barcode* of \mathbf{V} , denoted $\text{bcd}(\mathbf{V})$, and so that:

1. $\text{bcd}(\mathbf{V})$ subsumes the Betti numbers: If $\alpha \in \mathbb{R}$, then $\dim(V_\alpha)$ is exactly the number of intervals $I \in \text{bcd}(\mathbf{V})$, counted with repetitions, so that $\alpha \in I$.
2. $\text{bcd}(\mathbf{V})$ encodes persistence: For every $I \in \text{bcd}(\mathbf{V})$ and every $\alpha \in I$, there exists $y \in V_\alpha$ so that the left and right end-points of I are $\text{birth}(y)$ and $\text{death}(y)$, respectively.
3. $\text{bcd}(\mathbf{V})$ is an invariant: If \mathbf{W} is a persistence vector space with $\dim(W_\alpha)$ finite for all $\alpha \in \mathbb{R}$, then $\text{bcd}(\mathbf{V}) = \text{bcd}(\mathbf{W})$ if and only if $\mathbf{V} \cong \mathbf{W}$.

We will use $\text{bcd}_n^{\mathcal{R}}(X; \mathbb{F})$ to denote the barcode for the n -dimensional persistent homology of $\mathcal{R}(X)$. Below in

Figure 4 we show an example for $X \subset \mathbb{R}^2$ sampled with noise around the unit circle S^1 , the Rips complex $R_\alpha(X)$ at scales $\alpha = 0, 0.36, 0.6, 1.21$, and the intervals (i.e., the horizontal blue lines) that comprise the barcode $\text{bcd}_1^{\mathcal{R}}(X; \mathbb{F}_2)$. The computations were performed using the C++ library *Rips* [4]. The single long interval is indicative of a persistent 1-dimensional hole in the data, which is consistent with X being sampled around S^1 ; indeed, $\beta_1(S^1; \mathbb{F}) = 1$. The shorter intervals, on the other hand, are due to noise and sampling artifacts.

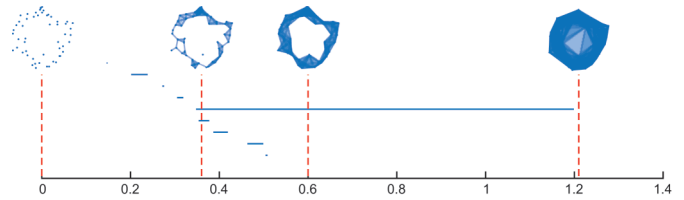


Figure 4. Barcode for the Rips filtration on $X \subset \mathbb{R}^2$ near S^1 .

As we have seen thus far, persistent homology can be used to infer the topology of a space \mathbb{X} given a finite sample X : the number of (comparatively) longer intervals in $\text{bcd}_n^{\mathcal{R}}(X; \mathbb{F})$ suggests a value for $\beta_n(\mathbb{X}; \mathbb{F})$. The barcodes $\text{bcd}_n^{\mathcal{R}}(X; \mathbb{F})$ can also be used to quantify/identify properties of dynamical systems: if \mathbb{X} is an attractor, a barcode like the one in Figure 4 would be indicative of periodicity, while the barcodes in Figure 5 would point to quasiperiodicity. Indeed, as we alluded to at the beginning of this section, the superposition of incommensurate oscillators (i.e., quasiperiodicity) is tied to attractors with the topology of a high-dimensional torus [21]. In addition to periodicity and quasiperiodicity, measures of Hausdorff dimension can also be derived from persistent homology; see for instance [20, 25].

Attractor Reconstruction: Time Series Data, Takens' Theorem, and Sliding Window Embeddings

In practice it is exceedingly rare to have an explicit mathematical description of a dynamical system of interest. Instead, one can often gather measurements of relevant quantities for each state $p \in M$ — e.g., in weather prediction one can estimate temperature, pressure, etc. A way of measuring can be thought of as a continuous map $F : M \rightarrow \mathbb{R}$ — called an *observation function* — and given an initial state $p \in M$, one obtains the time series

$$\begin{aligned} \varphi_p : \mathbb{R} &\rightarrow \mathbb{R} \\ t &\mapsto F \circ \Phi(t, p) \end{aligned} \quad (6)$$

The blue and orange curves from Figure 1 are examples of time series from the Lorenz system. A single time series may appear to be a complete oversimplification of the

underlying dynamics. However, Takens' embedding theorem, due to Floris Takens [28], implies that they can actually be very useful. For smooth manifolds M and N , and for a nonnegative integer k , let $C^k(M, N)$ denote the set of functions from M to N whose derivatives up to degree k exist and are continuous. $C^k(M, N)$ can be endowed with a topology, the (strong) Whitney topology, in which, roughly speaking, two functions are close if and only if the functions and all their derivatives up to degree k are close on compact subsets of M . Here is Takens' theorem:

Theorem 0.4. *Let M be a smooth, compact, Riemannian manifold; let $\tau > 0$ be a real number; and let $d \geq 2 \dim(M)$ be an integer. Then, for generic $\Phi \in C^2(\mathbb{R} \times M, M)$ and $F \in C^2(M, \mathbb{R})$, and for $\varphi_p(t)$ defined by (6), the delay map*

$$\begin{aligned} \varphi : M &\rightarrow \mathbb{R}^{d+1} \\ p &\mapsto (\varphi_p(0), \varphi_p(\tau), \varphi_p(2\tau), \dots, \varphi_p(d\tau)) \end{aligned} \quad (7)$$

is an embedding (i.e., φ is injective and its derivative has full-rank everywhere).

Generic means that the set of functions Φ, F for which (7) is an embedding is open and dense in the Whitney topology. In fact, if $A \subset M$ is a strange attractor, then φ restricted to A will be (generically) an embedding whenever d is at least twice the Hausdorff dimension of A . Takens' theorem motivates the following definition.

Definition 0.5. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a function, $\tau > 0$ a real number, and $d > 0$ an integer. The *sliding window embedding* of f , with parameters d and τ , is the vector-valued function

$$\begin{aligned} SW_{d,\tau}f : \mathbb{R} &\rightarrow \mathbb{R}^{d+1} \\ t &\mapsto (f(t), f(t+\tau), f(t+2\tau), \dots, f(t+d\tau)) \end{aligned} \quad (8)$$

The integer $d+1$ is the dimension, τ is the delay, and the product $d\tau$ is the window size. For $T \subset \mathbb{R}$, the set

$$SW_{d,\tau}f = \{SW_{d,\tau}f(t) : t \in T\} \quad (9)$$

is the *sliding window point cloud* associated to the *sampling set* T .

Hence, given time series data $f(t) = \varphi_p(t)$ observed from a potentially unknown dynamical system (M, Φ) , Takens' theorem implies that (generically) the sliding window point cloud $SW_{d,\tau}f$ provides a topological copy of $\{\Phi(t, p) : t \in T\} \subset M$. In particular this will reconstruct attractors. The underlying shape of $SW_{d,\tau}f$ can be then quantified with persistent homology, and the associated barcodes $bcd_n^R(SW_{d,\tau}f; \mathbb{F})$ can be used as features in inference, classification, and learning tasks [12, 25, 32]. We will see shortly several applications of these ideas to science and engineering, as well as a theoretical discussion

on sliding window persistence and parameter choices. For now, here is an instantiation of the pipeline:

Example. Let $\omega \in \mathbb{R}$ be irrational; we will use $\omega = \sqrt{3}$ for computations but any other choice would do. Consider the dynamics Φ and the observation function F on the torus $\mathbb{T}^2 = S^1 \times S^1 \subset \mathbb{C}^2$, given by

$$\begin{aligned} \Phi : \mathbb{R} \times \mathbb{T}^2 &\rightarrow \mathbb{T}^2 \\ (t, (z_1, z_2)) &\mapsto (e^{it}z_1, e^{i\omega t}z_2) \\ F : \mathbb{T}^2 &\rightarrow \mathbb{R} \\ (z_1, z_2) &\mapsto \operatorname{Re}(z_1 + z_2) \end{aligned}$$

If $p \in \mathbb{T}^2$, then $\{\Phi(t, p) : t \in \mathbb{R}\}$ is dense in \mathbb{T}^2 , and hence \mathbb{T}^2 is the only attractor; e.g., see [6], page 86, Example 6.15. For $p = (1, 1)$ we obtain the quasiperiodic time series $f(t) = F \circ \Phi(t, (1, 1)) = \cos(t) + \cos(\omega t)$, and we show in Figure 5 the dynamics $t \mapsto \Phi(t, (1, 1))$ on the torus (left), the resulting time series $f(t)$ (center), and the barcodes $bcd_n = bcd_n^R(SW_{d,\tau}f; \mathbb{F}_2)$ for $d = 4, \tau = \frac{3}{4}\sqrt{3}\pi, n = 0, 1, 2$ (right).

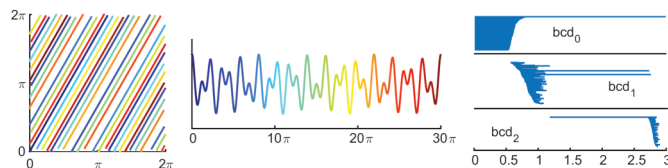


Figure 5. Left: The dynamics Φ on the torus. The colors, blue through red, indicate the time t for $t \mapsto \Phi(t, (1, 1)) \in \mathbb{T}^2$. Center: The time series $f(t) = F \circ \Phi(t, (1, 1)) = \cos(t) + \cos(\sqrt{3}t)$. The colors indicate the time variable $t \in [0, 30\pi]$ and are coordinated with the planar torus on the left panel. Right: Barcodes for the Rips filtration $\mathcal{R}(SW_{d,\tau}f)$; the number of long intervals recovers the Betti numbers of the attractor: $\beta_0(\mathbb{T}^2; \mathbb{F}_2) = \beta_2(\mathbb{T}^2; \mathbb{F}_2) = 1$ and $\beta_1(\mathbb{T}^2; \mathbb{F}_2) = 2$.

Some Applications of Sliding Window Persistence

Wheeze detection. A wheeze is an abnormal whistling sound produced while breathing. It is often associated with obstructed airways and lung diseases such as asthma, lung cancer, and congestive heart failure. In [11], Emrani, Gentimis, and Krim show that the 1-dimensional barcode $bcd_1^R(SW_{d,\tau}f; \mathbb{F})$, and particularly the length of its longest interval (i.e., its maximum persistence)

$$\begin{aligned} \operatorname{mp}_1^R(SW_{d,\tau}f; \mathbb{F}) &= \max \{ \operatorname{length}(I) : I \in bcd_1^R(SW_{d,\tau}f; \mathbb{F}) \} \end{aligned} \quad (10)$$

is an effective feature for wheezing detection when f is a recorded breathing sound. Indeed, the presence of wheezing in the sound signal f leads to circular sliding window point clouds. When testing on a large database of sound

recordings, Emrani et al. show that (10) leads to a higher detection accuracy than that of competing methods.

Periodicity quantification in gene expression data. Many biological processes, including the cell cycle, cell division, and the circadian clock, are periodic in nature. An important problem in systems biology is to describe these mechanisms at a genetic level [27], and biologists approach this by first collecting data. Specifically, how gene expression changes across time in a given model organism — e.g., yeast, mice, etc. In [24] databases of time series of gene expression data from the yeast cell cycle, the yeast metabolic cycle, and the mouse circadian clock are used to show that measures similar to (10) can outperform state-of-the-art methods for periodicity quantification, leading to the discovery of novel clock-regulated genes.

Segmentation of dynamic regimes. Complex high-dimensional systems can exhibit abrupt changes in qualitative behavior. For instance, the Earth's climate has undergone several sudden transitions to and back from a "snowball Earth" [14]. Identifying markedly different regimes in a system's evolution can thus be used for warning, modeling, and parameter estimation purposes [10]. Berwald et al. show in [5] that, given time series data, effective classifiers can be trained on features from the barcodes of sliding window point clouds, with the goal of automatically segmenting a system into different behavioral regimes. Some applications of their methodology include the detection of bifurcations on stochastic and chaotic systems, as well as the analysis of temperature and CO₂ levels in the Earth's ice cores.

Chatter detection and classification in machining. Turning and milling are cutting processes used extensively in industrial manufacturing. Chatter, or machining vibrations, are wide oscillations of the cutting tool with respect to the metal workpiece; these undesired undulations leave surface flaws on the production piece during turning and milling. Khasawneh et al. show in [15, 16] that, given a time series f describing the undulations of the cutting piece, $\text{bcd}_1^R(\text{SW}_{d,\tau}f; \mathbb{F})$ can be used as an input feature in classification algorithms for chatter detection.

Periodicity and quasiperiodicity in video data. A video can be thought of as a multi-dimensional time series: the result of sampling a function $f: \mathbb{R} \rightarrow \mathbb{R}^k$. It follows that computing $\text{bcd}_n^R(\text{SW}_{d,\tau}f; \mathbb{F})$ for the sliding window point cloud $\text{SW}_{d,\tau}f \subset \mathbb{R}^{k(d+1)}$ can be used to detect recurrent behavior (e.g., periodicity and quasiperiodicity) in video data, without the need for tracking or surrogate signals. See for instance [30] for applications of these ideas to the problem of quantifying periodicity in video data, the detection of biphonation in high-speed laryngeal videodendoscopy, as well as the synthesis of slow-motion videos from recurrent movements [29].

Theoretical Investigations of Sliding Window Persistence

The barcodes $\text{bcd}_n^R(\text{SW}_{d,\tau}f; \mathbb{F})$ have been used successfully in several applications as described above. In practice, there are two main challenges that must always be addressed. The first is parameter selection: appropriate values for $d \in \mathbb{N}$, $\tau > 0$, \mathbb{F} and $T \subset \mathbb{R}$ need to be determined, given the application and computational resources at hand. The second challenge is validating the results. In tasks where the barcodes $\text{bcd}_n^R(\text{SW}_{d,\tau}f; \mathbb{F})$ are used as features, one must quantify the likelihood that positive results are due to random fluctuations in the data. These challenges bring into stark focus the need for a theoretical understanding of $\text{bcd}_n^R(\text{SW}_{d,\tau}f; \mathbb{F})$ as a function of all the parameters and the time series involved. The main difficulty here is that our knowledge of the homotopy type/homology of the Rips complex $R_\alpha(X)$, for arbitrary α , is rather limited: planar circles [1] and gluings thereof [3] are essentially the only spaces where we have complete answers. I would like to highlight next some of what we do know.

Sliding window persistence of periodic functions. Let $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$. As a warm-up example for understanding the sliding window persistence of $f \in C^1(\mathbb{T}, \mathbb{R})$, let $L \in \mathbb{N}$, $\phi \in \mathbb{R}$, and $\zeta(t) = \sin(Lt + \phi)$. A bit of trigonometry shows that

$$\text{SW}_{d,\tau}\zeta(t) = \sin(Lt + \phi)\mathbf{u} + \cos(Lt + \phi)\mathbf{v}$$

where $\mathbf{u} = \text{SW}_{d,\tau} \cos(Lt)|_{t=0}$ and $\mathbf{v} = \text{SW}_{d,\tau} \sin(Lt)|_{t=0}$. It readily follows that if $d \geq 1$ and the set $\{\mathbf{u}, \mathbf{v}\}$ is linearly independent, then $\text{SW}_{d,\tau}\zeta = \text{SW}_{d,\tau}\zeta(\mathbb{T})$ is a planar ellipse. The semi-major and semi-minor axes can be computed explicitly as

$$a = \sqrt{\frac{(d+1) + \left| \frac{\sin(L(d+1)\tau)}{\sin(L\tau)} \right|}{2}} \quad (11)$$

and

$$b = \sqrt{\frac{(d+1) - \left| \frac{\sin(L(d+1)\tau)}{\sin(L\tau)} \right|}{2}}.$$

The persistent homology of the Rips filtration on ellipses with small eccentricity, i.e. when $b < a < \sqrt{2}b$, has been recently studied by Adamaszek et al. [2]. In particular, their work implies that if

$$\alpha_1 = \frac{4\sqrt{3}ab^2}{a^2 + 3b^2} \quad \text{and} \quad \alpha_2 = \frac{4\sqrt{3}a^2b}{3a^2 + b^2}$$

then the homotopy type of $R_\alpha(\text{SW}_{d,\tau}\zeta)$ for $0 < \alpha \leq \alpha_2$ is either that of the circle, or that of a wedge of 2-dimensional

spheres as follows:

$$R_\alpha(\mathbb{S}W_{d,\tau}\zeta) \simeq \begin{cases} S^1 & \text{for } 0 < \alpha < \alpha_1 \\ S^2 & \text{for } \alpha = \alpha_1 \\ \sqrt[5]{S^2} & \text{for } \alpha_1 < \alpha < \alpha_2 \\ \sqrt[3]{S^2} & \text{for } \alpha = \alpha_2 \end{cases}$$

The range $\alpha_1 < \alpha < \alpha_2$ is especially interesting since only one of the five 2-dimensional classes persists. In other words, the linear transformation

$$H_2(R_\alpha(\mathbb{S}W_{d,\tau}\zeta); \mathbb{F}) \rightarrow H_2(R_{\alpha'}(\mathbb{S}W_{d,\tau}\zeta); \mathbb{F})$$

has rank one for every $\alpha_1 < \alpha < \alpha' < \alpha_2$. We readily obtain the following theorem.

Theorem 0.6. *Let $\zeta(t) = \sin(Lt + \phi)$ for $L \in \mathbb{N}$, $\phi \in \mathbb{R}$, and for $d \in \mathbb{N}$, $\tau > 0$ let $a, b \geq 0$ be as in (11). If $a < \sqrt{2}b$ and \mathbb{F} is a field, then the maximum persistence (10) in dimensions 1 and 2 satisfies*

$$\begin{aligned} \text{mp}_1^{\mathcal{R}}(\mathbb{S}W_{d,\tau}\zeta; \mathbb{F}) &= \frac{4\sqrt{3}ab^2}{a^2 + 3b^2} \\ \text{mp}_2^{\mathcal{R}}(\mathbb{S}W_{d,\tau}\zeta; \mathbb{F}) &\geq \frac{4\sqrt{3}a^2b}{3a^2 + b^2} - \frac{4\sqrt{3}ab^2}{a^2 + 3b^2}. \end{aligned}$$

The cases $\alpha > \alpha_2$ for $a < \sqrt{2}b$ and $\alpha > 0$ for $a \geq \sqrt{2}b$ are currently open. The case $a = b$, i.e. when $\mathbb{S}W_{d,\tau}\zeta$ is a circle, is much better understood. This happens when the window size $d\tau$ is equal to (an integer multiple of) $\frac{d}{d+1} \frac{2\pi}{L}$, which is a little bit under $\text{Period}(\zeta) = \frac{2\pi}{L}$. Since the homotopy type of $R_\alpha(S^1)$ is known for all $\alpha > 0$ [1], we get that

$$R_\alpha(\mathbb{S}W_{d,\tau}\zeta) \simeq \begin{cases} S^{2k+1} & \text{for } \sin\left(\frac{\pi k}{2k+1}\right) < \frac{\alpha}{\sqrt{2(d+1)}} < \sin\left(\frac{\pi(k+1)}{2k+3}\right) \\ \bigvee^{|\mathbb{R}|} S^{2k} & \text{for } \alpha = \sqrt{2(d+1)} \sin\left(\frac{\pi k}{2k+1}\right) \end{cases}$$

where the linear transformation

$$H_{2k+1}(R_\alpha(\mathbb{S}W_{d,\tau}\zeta); \mathbb{F}) \rightarrow H_{2k+1}(R_{\alpha'}(\mathbb{S}W_{d,\tau}\zeta); \mathbb{F})$$

is an isomorphism (with rank one) for every

$$\begin{aligned} \sqrt{2(d+1)} \sin\left(\frac{\pi k}{2k+1}\right) &< \alpha \leq \alpha' \\ &< \sqrt{2(d+1)} \sin\left(\frac{\pi(k+1)}{2k+3}\right). \end{aligned}$$

Therefore,

Theorem 0.7. *Let $\zeta(t) = \sin(Lt + \phi)$, let \mathbb{F} be a field, and let $\tau = \frac{2\pi}{L(d+1)}$. Then, for every integer $k \geq 1$,*

$$\text{mp}_{2k}^{\mathcal{R}}(\mathbb{S}W_{d,\tau}\zeta; \mathbb{F}) = 0 \text{ and}$$

$$\text{mp}_{2k-1}^{\mathcal{R}}(\mathbb{S}W_{d,\tau}\zeta; \mathbb{F})$$

$$= \sqrt{2(d+1)} \cdot \left(\sin\left(\frac{\pi k}{2k+1}\right) - \sin\left(\frac{\pi(k-1)}{2k-1}\right) \right).$$

One strategy to understand $\text{bcd}_n^{\mathcal{R}}(\mathbb{S}W_{d,\tau}f; \mathbb{F})$ for a function $f \in C^1(\mathbb{T}, \mathbb{R})$ is to first approximate f by its truncated Fourier series

$$S_N f(t) = \sum_{|n| \leq N} \hat{f}(n) e^{int}, \quad \hat{f}(n) = \frac{1}{2\pi} \int_{\mathbb{T}} f(t) e^{-int} dt$$

and then investigate the asymptotic behavior of the sequence of barcodes

$$\text{bcd}_n^{\mathcal{R}}(\mathbb{S}W_{d,\tau}S_N f; \mathbb{F}) \quad , \quad N \in \mathbb{N}. \quad (12)$$

Indeed, the analysis of $\zeta(t) = \sin(Lt + \phi)$ presented earlier can be bootstrapped to trigonometric polynomials, and the Stability Theorem [8] can be used to study $\text{bcd}_n^{\mathcal{R}}(\mathbb{S}W_{d,\tau}f; \mathbb{F})$ via the behavior of (12) as $N \rightarrow \infty$. This line of reasoning was explored in [23]. In particular, it yields insights for the choice of window size (it should be approximately $\frac{d}{d+1}$ times the period length $\text{Period}(f)$), the embedding dimension (larger than twice the number of relevant harmonics), and the choice of field of coefficients (one whose characteristic does not divide $\frac{2\pi}{\text{Period}(f)}$). We end with a theorem [23, 6.8] relating the sliding window persistence of f to its harmonic content. In particular, it provides an estimate of the effectiveness of maximum persistence as a measure of periodicity, and highlights how densely the sliding window point cloud needs to be sampled (i.e. the choice $T \subset \mathbb{T}$), given how irregular the function is, and the localization of its spectral power density.

Theorem 0.8. *Let $f \in C^1(\mathbb{T}, \mathbb{R})$ be so that $f(t + \frac{2\pi}{L}) = f(t)$ for all $t \in \mathbb{T}$, and assume (for simplicity) that f has been centered and normalized:*

$$\hat{f}(0) = 0 \quad , \quad \|f\|_2 := \left(\frac{1}{2\pi} \int_{\mathbb{T}} |f(t)|^2 dt \right)^{1/2} = 1$$

Let $\tau_d = \frac{2\pi}{L(d+1)}$, $T \subset \mathbb{T}$, and $\mathbb{S}W_{d,\tau_d}f = SW_{d,\tau_d}f(T)$. If $f' = \frac{df}{dt}$ and d_H is the Hausdorff distance, then

$$\begin{aligned} \sup_{d \in \mathbb{N}} \text{mp}_1^{\mathcal{R}}(\mathbb{S}W_{d,\tau_d}f; \mathbb{Q}) \\ \geq 2\sqrt{3} \sup_{n \in \mathbb{Z}} |\hat{f}(n)| - 2\sqrt{2} \|f'\|_2 d_H(T, \mathbb{T}). \end{aligned}$$

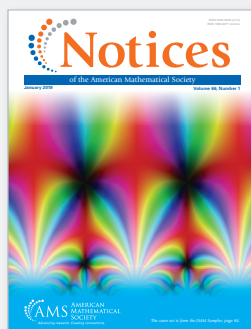
Beyond periodicity. Very little is known about the sliding window persistence of other families of functions. There are some results for quasiperiodic time series [21], but the rest of the landscape is essentially uncharted territory. There is also recent work in identifying families of functions whose sliding window point clouds recover other spaces [32] — e.g., Klein bottles, projective spaces, etc. It would be very interesting to see these models in naturally occurring phenomena, and perhaps in future applications of topology to the analysis of complex time varying data.

References

- [1] Adamaszek M, Adams H. The vietoris–rips complexes of a circle. *Pacific Journal of Mathematics*, 290(1):1–40, 2017. [MR3673078](#)
- [2] Adamaszek M, Adams H, Reddy S. On vietoris–rips complexes of ellipses. *Journal of Topology and Analysis*, pages 1–30, 2017. [MR3673078](#)
- [3] Adamaszek M, Adams H, Gasparovic E, et al. Vietoris-rips and cech complexes of metric gluings. In *34th International Symposium on Computational Geometry (SoCG 2018)*, volume 99 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 3:1–3:15, 2018. [MR3824247](#)
- [4] Bauer U. Ripser: a lean c++ code for the computation of vietoris-rips persistence barcodes. 2017. Software: <https://github.com/Ripser/ripser>
- [5] Berwald JJ, Gidea M, Vejdemo-Johansson M. Automatic recognition and tagging of topologically different regimes in dynamical systems. *Discontinuity, Nonlinearity, and Complexity*, 3(4):413–126, 2014.
- [6] Boothby WM. *An introduction to differentiable manifolds and Riemannian geometry*, volume 120. Academic Press, 1986. [MR861409](#)
- [7] Carlsson G. Topological pattern recognition for point cloud data. *Acta Numerica*, 23:289, 2014. [MR3202240](#)
- [8] Chazal F, Cohen-Steiner D, Glisse M, Guibas LJ, Oudot SY. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual symposium on computational geometry*, pages 237–246. ACM, 2009.
- [9] Crawley-Boevey W. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and its Applications*, 14(05):1550066, 2015. [MR3323327](#)
- [10] Dakos V, Scheffer M, van Nes EH, et al. Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences*, 105(38):14308–14312, 2008.
- [11] Emrani S, Gentimis T, Krim H. Persistent homology of delay embeddings and its application to wheeze detection. *IEEE Signal Processing Letters*, 21(4):459–463, 2014.
- [12] Garland J, Bradley E, Meiss JD. Exploring the topology of dynamical reconstructions. *Physica D: Nonlinear Phenomena*, 334:49–59, 2016. [MR3545969](#)
- [13] Hatcher A. *Algebraic Topology*. Cambridge University Press, 2002. [MR1867354](#)
- [14] Hoffman PE, Kaufman AJ, Halverson GP, Schrag DP. A neoproterozoic snowball earth. *Science*, 281(5381):1342–1346, 1998.
- [15] Khasawneh FA, Munch E. Chatter detection in turning using persistent homology. *Mechanical Systems and Signal Processing*, 70:527–541, 2016.
- [16] Khasawneh FA, Munch E, Perea JA. Chatter classification in turning using machine learning and topological data analysis. In *14th IFAC Workshop on Time Delay Systems TDS 2018*, volume 51, pages 195–200. International Federation of Automatic Control, 2018.
- [17] Korn H, Faure P. Is there chaos in the brain? ii. experimental evidence and related models. *Comptes Rendus Biologies*, 326(9):787–840, 2003.
- [18] Latschev J. Vietoris-rips complexes of metric spaces near a closed riemannian manifold. *Archiv der Mathematik*, 77(6):522–528, 2001. [MR1879057](#)
- [19] Lorenz EN. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.
- [20] MacPherson R, Schweinhart B. Measuring shape with topology. *Journal of Mathematical Physics*, 53(7):073516, 2012. [MR2985256](#)
- [21] Perea JA. Persistent homology of toroidal sliding window embeddings. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6435–6439. IEEE, 2016.
- [22] Perea JA. A brief history of persistence. *preprint arXiv:1809.03624*, 2018. <https://arxiv.org/abs/1809.03624>.
- [23] Perea JA, Harer J. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15(3):799–838, 2015. [MR3348174](#)
- [24] Perea JA, Deckard A, Haase SB, Harer J. SW1PerS: Sliding windows and 1-persistence scoring: discovering periodicity in gene expression time series data. *BMC Bioinformatics*, 16(1):257, 2015.
- [25] Robins V. Towards computing homology from finite approximations. In *Topology Proceedings*, volume 24, pages 503–532, 1999. [MR1876386](#)
- [26] Ruelle D, Takens F. On the nature of turbulence. *Communications in mathematical physics*, 20(3):167–192, 1971. [MR0284067](#)
- [27] Rustici G, Mata J, Kivinen K, et al. Periodic gene expression program of the fission yeast cell cycle. *Nature Genetics*, 36(8):809, 2004.
- [28] Takens F. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981. [MR654900](#)
- [29] Tralie CJ, Berger M. Topological eulerian synthesis of slow motion periodic videos. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3573–3577, 2018.
- [30] Tralie CJ, Perea JA. (quasi) periodicity quantification in video data, using topology. *SIAM Journal on Imaging Sciences*, 11(2):1049–1077, 2018. [MR3796363](#)
- [31] Viswanath D. The fractal property of the lorenz attractor. *Physica D: Nonlinear Phenomena*, 190(1-2):115–128, 2004. [MR2043795](#)
- [32] Xu B, Tralie CJ, Antia A, Lin M, Perea JA. Twisty takens: A geometric characterization of good observations on dense trajectories. *arXiv preprint arXiv:1809.07131*. <https://arxiv.org/pdf/1809.07131>.

AMS MEMBERS,? ARE YOU RELOCATING?

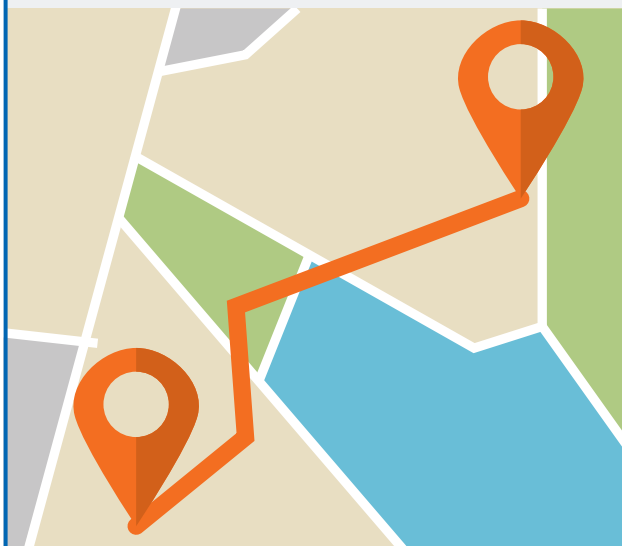
Please make sure that *Notices*
and *Bulletin* find their new home.



Update your address at
www.ams.org/member-directory.

You can also send address changes
to amsmem@ams.org or:

Sales and Member Services
American Mathematical Society
201 Charles Street
Providence, RI 02904-2213 USA



 **AMS** AMERICAN
MATHEMATICAL
SOCIETY
Advancing research. Creating connections.



Jose A. Perea

Credits

Photo of Jose A. Perea and all article figures are courtesy of
the author.