Topological State Space Inference for Dynamical Systems

Mishal Assif P K MISHAL2@ILLINOIS.EDU
Department of Electrical and Computer Engineering, Grainger College of Engineering, University of Illinois at Urbana-Champaign, Urbana, IL

Yuliy Baryshnikov YMB@ILLINOIS.EDU
Departments of Mathematics and Electrical and Computer Engineering, Grainger College of Engineering, University of Illinois at Urbana-Champaign, Urbana, IL

Editors: N. Ozay, L. Balzano, D. Panagou, A. Abate

Abstract

We present a computational pipe aiming at recovery of the topology of the underlying phase space from observation of an output function along a sample of trajectories of a dynamical system.

Keywords: State space realization, topological data analysis, nonlinear dynamics.

1. Introduction

The problem of state space realization, that is, the reconstruction of the underlying space of a dynamical system and an observation function, remains foundational in systems theory. At its core, state space realization aims to reconstruct or identify a state space model that accurately describes the internal dynamics of a system based solely on observed data or system outputs. This problem originated with the work of Richard Bellman and Rudolf Kalman, whose pioneering research in the 1960s laid the groundwork for modern systems theory (Schutter, 2000).

The state space realization problem for linear models is well studied. Classical elegant solutions are available in this case relying, in essence on the theory of modules over the polynomial rings. Linearity enables a clear and tractable description of system dynamics, where system dynamics can be modeled as linear transformations, providing well-established methods for state space realization.

Extensions to nonlinear systems introduce, however, substantial complexity. Nonlinear systems often exhibit intricate behaviors, including bifurcations and chaos. Consequently, realizing or reconstructing a state space for such systems is far less straightforward. In this note, we present an attempt to address the nonlinear version of the state space realization problem using a mathematical theory as a guiding light, much as the LTI realization theory relied on the structure theory of modules over polynomial rings in one variable.

1.1. Setup

Consider a dynamical system (M, ϕ) where M is a topological space and $\phi : \mathbb{R} \times M \to M$ is a one-parametric transformation group, that is, ϕ is a continuous function

$$\mathbb{R} \times M \ni (t, m) \to \phi(t, m) := \phi_t(m) \in M \text{ such that } \phi_0(m) = m,\ \phi_t(\phi_s(m)) = \phi_{t+s}(m)$$

In most of our examples, M will be a smooth compact manifold and the transformation group will be generated by flows of a vector field $v : M \to TM$. We will denote by $\pi_x : \mathbb{R} \to M$ the trajectory of v such that

$$\pi_x(0) = x,\ \frac{d\pi_x(t)}{dt} = v(\pi_x(t)). \tag{1}$$

Note that in such cases, the transformation group ϕ is at least as smooth as the vector field v.

The question we investigate here is *whether the topology of the space M can be recovered from a finite number of trajectories $\{\pi_{x_k}\}_{k=1,..,n}$* of such a dynamical system on M. We will see that this is indeed possible and provide a principled computational pipeline and a methodology to extract topological information of M from such trajectories.

1.2. Data Analytic setup

We are interested in building an algorithmic setup towards inference of the topology from observation data. To make the general problem more amenable to such data analysis, we add the following constraints to the problem setting:

1. The data are collected at discrete time intervals, whether it be actual readings from sensors or the result of numerical algorithms or simulations. In addition, such readings are collected only up to a finite time limit. We cannot hope to get entire trajectories π_x starting from a point x. Instead we will get discrete trajectories obtained by performing a finite sampling along the trajectory at regular time intervals. The sampling interval δ will be assumed to be constant across all trajectories and will be dropped from the notation hereafter.

2. One observes only a finite sample of such trajectories, initialized according to some distribution, which may or not be under our control.

3. One observes not the points of the state space M (which is unknown), but rather a smooth output function
$$g : M \to V \cong \mathbb{R}^p$$
such that only $g^n(x) := (g(\pi_x(-l\delta)), g(\pi_x(-(l-1)\delta)), ..., g(\pi_x(k\delta))) \in V^n, n = k+l+1$ is available to us. The dimension of V can be much smaller than that of M, making this problem more challenging.

Hence the overall problem we investigate here is: *What topological information of M can be extracted from a finite set of observations of uniformly sampled trajectories $\{g^n(\pi_{x_\alpha})\}_{\alpha=1,..,N}$ of a dynamical system on M?*

1.3. Takens' Embedding

There is a well-known answer to the question of the state realization due to the Takens embedding theorem (Takens, 2006).

Theorem 1.1 *Let M be a compact smooth manifold, and $n \geq 2\dim(M) + 1$ an integer. Then, for a* generic *pair (ϕ, g), where $\phi : M \to M$ is a C^2-diffeomorphism, and g a function in $C^2(M, \mathbb{R})$, the map*
$$M \ni x \mapsto (g(x), g(\phi(x)), ..., g(\phi^{n-1}(x)) \in \mathbb{R}^k$$
is an embedding *of M.*

Genericity here is understood in the standard way: the pairs (ϕ, g) which satisfy this condition form an open, everywhere dense subset in the Banach manifold of such pairs.

Takens embedding theorem guarantees that for generic dynamical systems and observations, given the infinite collection of observations of uniformly sampled trajectories starting at any point on the manifold M with length more than twice the dimension of M, we can exactly recover M up to a diffeomorphism.

While Takens theorem is encouraging and suggests that our problem is not completely intractable, it does not address our problem, being a purely existence result, not providing a viable strategy for reconstructing M from the trajectories, even in the ideal case where trajectories starting from all possible initial points are available.

In our setting, the nature of the sample one observes, - a finite sample, - is intrinsically different from the smooth manifold M. Nevertheless, *topological invariants*, namely, the homology groups of the state space, can be recovered from finite observation of the trajectories, as we argue below.

1.4. Related Results

The problem of the state space realization of nonlinear dynamical systems from observations has been addressed by many authors. Thus, (Jakubczyk, 1980) develops an analogue of the Taken's embedding theorem for nonlinear control systems, while (van der Schaft, 1986) looks at the action of the dynamics on the functions in a similar manner, leading to, essentially, a version of the Koopman formalism (Budišić et al., 2012).

The recent work (Otto et al., 2023) considers data driven state space recovery using Machine Learning methods. Among other works exploring such model-free approaches one should mention (Schmidt and Lipson, 2009; Brunton et al., 2016; Champion et al., 2019).

2. Mathematical Background

The motivation for our theory comes from the beautiful work of Cohen, Jones, and Segal (Cohen et al., 1995), which we then couple with the well-established apparatus of the Topological Data Analysis (TDA) (Edelsbrunner and Harer, 2010).

2.1. Basic Definitions

Here is their setup. Assume that M is a smooth closed Riemannian manifold and $f : M \to \mathbb{R}$ a smooth Morse function on M. The key idea of (Cohen et al., 1995) is to describe a combinatorial construction that reconstructs the topology of M from the space of flow lines of the *gradient vector field* ∇f.

To this end, they introduced the enriched category[1] $\mathcal{C}_f$. The objects of this category are critical points of f, and the space of morphisms $\mathtt{Mor}(a_-, a_+)$ between two critical points a_-, a_+ is the closure of the space of the flow lines $\gamma : \mathbb{R} \to M$ of the gradient vector field v such that $\gamma(t) \to a_\pm$ as $t \to \pm\infty$.

Concatenating smooth flow lines (which have matching target and source, respectively) allows one to define a continuous composition $\mathtt{Mor}(a, b) \times \mathtt{Mor}(b, c) \to \mathtt{Mor}(a, c)$.

1. The language of category theory (which can be looked up in, say, (Riehl, 2017)) can be avoided, but for the space reasons we just follow the original presentation. Enriched category here just means that the morphisms between any objects carries an extra structure (of the topological space).

The key result of (Cohen et al., 1995) is that the *classifying space of the category* $\mathcal{C}_f$, denoted by $\mathcal{BC}_f$, is homotopy equivalent to M. Recall that the classifying space $\mathcal{BC}_f$ of a category is a simplicial space whose spaces of k-simplices are parametrized by k-tuples of composable morphisms in $\mathcal{C}_f$, modulo a natural identification procedure. Altogether, these identification operations yield a combinatorial procedure of gluing together the pieces (topological spaces $\texttt{Mor}(\cdot,\cdot)$), resulting in a topological space, the classifying space of the category $\mathcal{C}_f$ The relation of this space to the original manifold M is given by the theorem:

Theorem 2.1 ((Cohen et al., 1995)) *Let $f : M \to \mathbb{R}$ be a Morse function on the closed Riemannian manifold M. Let $\mathcal{C}_f$ be the topological category whose objects are the critical points of f and whose space of morphisms between two critical points is the space of piecewise flow lines of the gradient vector field of f connecting the two critical points and composition of morphisms given by composition of piecewise flow lines, then*

1. *the classifying space $\mathcal{BC}_f$ is homotopy equivalent to M, $\mathcal{BC}_f \simeq M$.*
2. *the classifying space $\mathcal{BC}_f$ is homeomorphic to M if f is a (generic) Morse function whose gradient flow satisfies the Morse-Smale transversality conditions, $\mathcal{BC}_f \cong M$.*

2.2. Topological Data Analysis

We give a brief introduction to the relevant tools (see (Edelsbrunner and Harer, 2010) for a general reference).

Let X be a compact metric space (say, a manifold). Given a finite sample of points $\mathcal{A} = \{x_\alpha\}, \alpha \in A$ drawn from X (and the induced distances between the points), persistent homology is a tool to recover the topological features (e.g., the homology groups $H_k(X)$ in dimensions $k \in \mathbb{N}$) of the space X from the discrete data $\mathcal{A}$ alone. [2]

A finite collection of open subsets $\{U_\alpha\}_{\alpha\in A}$ covering X is a *nice* cover if any intersection of these subsets is either empty or contractible. In this situation, the classical *Nerve Theorem* implies that a combinatorial object, the simplicial complex formed by the subsets $S \subset A$ for which the intersection of the open sets $\cap_{\alpha\in A} U_\alpha$ is non-empty, is homotopy-equivalent to X, thus opening an effective procedure of computing homologies of X.

A collection of open metric balls $U_\alpha = B^\circ_r(x_\alpha), \alpha \in A$, could serve as such an open collection, if one could overcome two difficulties.

One is that detecting the non-empty intersection of those balls might be difficult. An effective palliative is to work rather than with the nerve with the *Vietoris-Rips* simplicial complex: here one declares a collection $S \subset A$ to form a simplex, if all *pairwise* distances between points $x_\alpha, \alpha \in S$ are at most r. This works, at least for dense enough samples, for appropriate r (Latschev, 2001).

The other difficulty is to find an appropriate scale r: choose it too small, and there are no nontrivial simplices; too large, and the resulting space is one high-dimensional simplex.

To overcome this problem, one considers a *filtered simplicial complex*, an increasing collection of simplicial complexes $\Sigma_r, r \in \mathbb{R}, \Sigma_r \subset \Sigma_{r'}$ for $r < r'$. In our context, Σ_r is the Vietoris-Rips complex corresponding to the cutoff distance r.

2. The homology group in dimension k is a vector space whose dimensions reflects the number of k-dimensional holes in X. Although there are other important topological features of a space, homology groups are particularly easy to compute algorithmically and so a large part of TDA focuses on these groups.

In this setting, one can keep track of the modules of cycles and boundaries in Σ_r as they evolve with r. A *persistent* cycle, i.e., one that emerged for small r but is not patched until much larger r is deemed to represent the true homology of the space X. The theory developed over the past two decades allows one to set up efficient computational procedures for finding them (Edelsbrunner and Harer, 2010).

3. Inferring Topology from Observations

We will be deploying the tools of Persistent Homology for the metrics on the spaces of samples of trajectories we define relying on the intuitive picture generalizing the Cohen-Jones-Segal (CJS) construction. In a contrast with their approach, we allow arbitrary smooth dynamics, not requiring it being gradient.

3.1. Approximations

The key difference of CJS setup from Takens embedding is its reliance on the infinitely long trajectories. This forces corresponding adjustments to any computational implementation.

Consider the space $\Pi_T = \{\pi_t(x)(t), x \in M, t \in [-T, T]\}$ of the fragments of trajectories of the length $2T$. They are parameterized by their centers x (so that the space is diffeomorphic to M). Defining the distance between the trajectories in the standard way, using C^0 or similar norm would essentially reproduce Takens (or Koopman) paradigm. Instead, we use the distance that accounts for proximity of long subintervals of the trajectories.

As we intend to recover the state space from observations, we need to measure rather the distances between the trajectories of *observed* values. These considerations lead us to the following

Definition 3.1 (Slack distance) *Fix T, $n = 2k + 1$, and δ such that $T = k\delta$. For a given observation function $g : M \to V \cong \mathbb{R}^d$, let*

$$y^n(x) := (g(\pi_x(-k\delta)), \pi_x(-(k-1)\delta)) \ldots, g(\pi_x((k-1)\delta)), g(\pi_x(k\delta))) \in V^n.$$

Let $y_1^n = y^n(x_1), y_2^n = y^n(x_2)$ be two elements of V^n (i.e., two length n samples of trajectories of the dynamical system on M).

We say that y_1^n and y_2^n are at slack distance at most (t, ϵ) if the subsequences

$$(y_1^n(i), y_1^n(i+1), \ldots, y_1^n(i+n-s)) \text{ and } (y_2^n(j), y_2^n(j+1), \ldots, y_2^n(j+n-s)) \in V^{n-s}$$

are at the sup *distance at most ϵ for some $s \leq t$.*

We say that $x_1, x_2 \in M$ are at slack distance (t, ϵ) if $y_1^n = y^n(x_1), y_2^n = y^n(x_2)$ are.

In other words, $x_1, x_2 \in M$ are at slack distance (t, ϵ) if the measurements along the trajectories π_{x_1}, π_{x_2} are ϵ-close to each other for at least L steps, for some $L \geq n - t$.

Lemma 3.2 *If y_1^n, y_2^n are at the slack distance (s, ϵ), and y_2^n, y_3^n are at the slack distance (s', ϵ'), then y_1^n, y_3^n are at the slack distance at most $(s + s', \epsilon + \epsilon')$.*

Proof There are s indices at the endpoints of y_1^n and y_2^n that are at distance greater than ϵ and s' indices at the endpoints of y_2^n and y_3^n that are at distance greater than ϵ'. Outside the union of these

indices, which has length at most $s + s'$, the elements of pairs of trajectories (y_1^n, y_2^n) and (y_2^n, y_3^n) are within distance ϵ and ϵ' respectively. This means the elements of (y_1, y_3) are within distance $\epsilon + \epsilon'$ outside a set of indices of length $s + s'$. ■

As a corollary, for any convex homogeneous degree 1 function d_π of s and d, $d_\pi(s, d)$ satisfied the triangle inequality. The slack distance naturally gives rise to a bifiltration whose vertices represent trajectories of observations from a dynamical system and an edge appears between two trajectories at parameter (k, ϵ) if the trajectories are at slack distance (k, ϵ).

3.2. Vietoris-Rips Complexes for Slack Distance

Consider a sample $\mathcal{A}$ of N trajectories of the dynamical system (1). For each of these trajectories $\pi_{x_l} : [-T, T] \to M, l = 1, \ldots, N$, we assume that an output function $g : M \to V$ is observed at equispaced intervals, resulting in a collection of observations y_l^n of length n. Our objective of is to recover the topology of M from this set of observables.

The definition of the slack distance (and the Lemma 3.2) provide for a flexible construction of filtrations by Vietoris-Rips complexes with respect to slack distance.

Namely, we define the space $\mathbf{\Sigma}(t, r)$ as the Vietoris-Rips complex on $\mathcal{A}$ corresponding to parameters (t, r). In other words, the simplices of $\mathbf{\Sigma}(t, r)$ are given by the condition

$$\mathbf{\Sigma}(t, r) = \{\sigma = (\alpha_0, \ldots \alpha_d) : \text{ slack distance between } x_{\alpha_i}, x_{\alpha_j} \text{ is at most } (t, \epsilon)\}.$$

The slack distance between two trajectories for fixed slack $t \leq n$ can be calculated in $O(n^2 \log(n))$ time using dynamic programming, and the corresponding bifiltration can be built in $O(n^2 \log(n) N^2)$ time. The algorithm also requires $O(n^2)$ space.

The pseudocode of the algorithm is shown as Algorithm 1 below.

3.3. Slack Distance and Sliding Windows

Our approach to constructing a filtration out of trajectories of observables of a dynamical system for recovering topological information is superficially related to the *sliding window* setup developed in (Perea and Harer, 2015). The key difference that makes our approach more efficient is the *slack* allowing measuring the distance between shifted fragments of two trajectories. In the sliding window approach, there is no slack allowing the trajectories to be aligned after a time shift: in some sense, the distance between points pulled back from the distance between the sampled trajectories is quasi-conformal to the original metric structure on M.

In our approach, the presence of slack that makes the sampled trajectory $y^n(x)$ close to the sampled trajectory $y^n(\phi(\tau, x))$ for a bounded τ in essence adjusts the underlying metric structure to be small *along* the trajectories of v. This anisotropic rescaling of the underlying metric should be compared to the constructions of (Ferry and Okun, 1995), where the loosening of the Riemannian metric along the fibers of a map $M \to N$ leads to the Gromov-Hausdorff convergence of the resulting metric structures on M to those on N under some connectivity assumptions. We conjecture that the surprising efficacy of our algorithm can be explained along those lines, - which is a separate project, to be pursued elsewhere.

Algorithm 1 Algorithm for building the MSD bifiltration

Data: A pair of trajectories x^i, x^j both of length n
Result: List `ms_dist` of length n containing the Matching substring distances between x^i and x^j

Compute distance matrix $d(k,l) = \|x^i_k - x^j_l\|$
`srtd_indices` $\leftarrow argsort(d(k,l))$
`ms_dist`$[i] = \infty$ for $i \in \{0, 1, ..., n-1\}$
for $i, j \in$ `srtd_indices` **do**
  `init_indices`$[i,j] = i, j$
  `final_indices`$[i,j] = i, j$
  `len` $= 1$
  if $i+1, j+1 \in$ `init_indices` **then**
    `init_indices`$[i,j] =$ `init_indices`$[i+1, j+1]$
    `final_indices`[`init_indices`$[i,j]$] = `final_indices`$[i,j]$
    `len`$+=$ `init_indices`$[i+1, j+1][0] - i$
  if $i-1, j-1 \in$ `final_indices` **then**
    `final_indices`$[i,j] =$ `final_indices`$[i-1, j-1]$
    `init_indices`[`final_indices`$[i,j]$] = `init_indices`$[i,j]$
    `len`$+= i -$ `final_indices`$[i-1, j-1][0]$
  `ms_dist`[`len`] $= \min($`ms_dist`[`len`]$, d(i,j))$
end

4. Computational Examples

The rest of the note deals with the computational examples, where we attempt to reconstruct the topology of the underlying phase space, using the finite samples (of size N) of trajectories, finite observations of length n along the trajectories, and slack distance for some slack parameter t. The results are shown as persistence diagrams with respect to the allowable distance r between the fragments of the trajectories.

4.1. Gradient Dynamical System: Height Function on the 2-Sphere

We apply our computational pipeline to a gradient dynamical system on the 2-sphere $\boldsymbol{S}^2 \subset \mathbb{R}^3$ embedded as a submanifold of $\mathbb{R}^3$ in the standard sense. Consider the height function

$$\boldsymbol{S}^2 \ni (x, y, z) \mapsto h(x, y, z) = z \in \mathbb{R}$$

taking each point on the sphere to its z coordinate, or the point's *height* assuming the z-axis is aligned in the vertical direction.

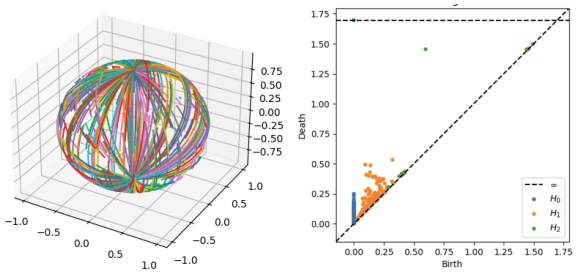

Figure 1: Segments of trajectories of the gradient of the height function on the sphere. Right: Persistence diagram generated from these trajectories on the sphere. Here $N = 400, n = 15, t = 10, T = 1.5$.

The flow lines of this dynamical system are just the meridians, connecting the poles, the critical points of the function z on the

2-sphere. We take $N = 400$ segments of trajectories on the sphere, that start at randomly drawn points. These trajectories are plotted in Figure 1.

The persistence diagram generated by the MSSD algorithm for this set of trajectories is also shown in Figure 1. The persistence diagram clearly reflects the fact that the homologies of a sphere $\boldsymbol{S}^d$ have rank 1 in dimensions 0 and d, and vanish elsewhere.

4.2. Gradient Dynamical System: Random Trigonometric Function on the Torus

We consider the example of a gradient dynamical system on the 2-torus $\mathbb{T}^2 \subset \mathbb{R}^3$ embedded in $\mathbb{R}^3$. We can parametrize the torus with two parameters $(\theta, \phi) \in [0, 2\pi] \times [0, 2\pi]$.

We used a random finite Fourier polynomial $f(\theta, \phi) = \sum_{i=0}^{k} \sum_{j=0}^{k} a_{i,j} \cos((i+1)\theta + \theta_{i,j}) cos((j+1)\phi + \phi_{i,j})$ where the coefficients $a_{i,j}$ are sampled from the standard normal distribution. We use $k = 2$ in this article.

We generated $N = 400$ trajectories, sampled from random initial conditions. Along each trajectory, $n = 25$ samples were taken, for $T = 2.5$. In this experiment, we took $V = \mathbb{R}^3$, with the observables given by an embedding of the torus into the 3-dimensional Euclidean space.

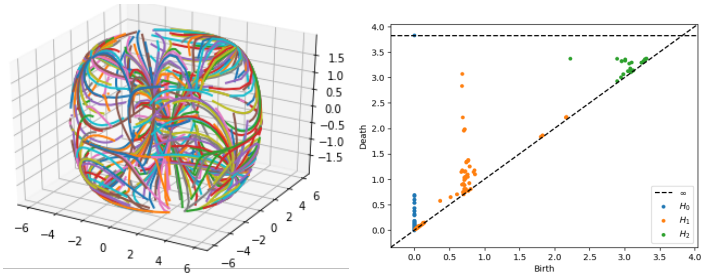

Figure 2: Segments of trajectories of random vector field on 2-torus. Right: Persistence diagram generated from these segments. Here $N = 400, n = 25, t = 3, T = 2.5$

The resulting trajectories are plotted in the left display of the Figure 2. The persistence diagram corresponding to $t = 3$ shown on the right display of the Figure 2.

The homology groups of the 2-torus are of rank 1 in the dimensions 0 and 2, and 2 in dimension 1. This is clearly captured in the persistence diagram Figure 2.

To check our algorithm for scalar output, we considered a random observation function given by $g(\theta, \phi) = b_0 x(\theta, \phi) + b_1 y(\theta, \phi) + b_2 z(\theta, \phi)$ where the coefficients (b_0, b_1, b_2) are sampled from a Gaussian distribution and the functions x, y, z are Cartesian coordinates for our embedding of the 2-torus in $\mathbb{R}^3$.

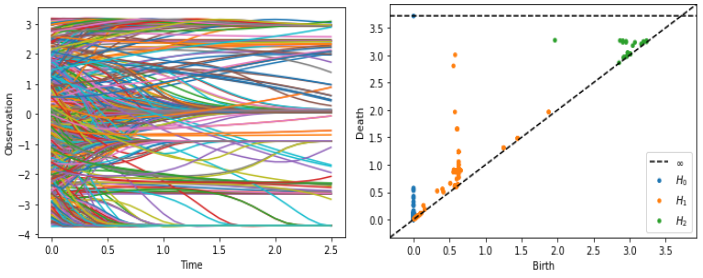

Figure 3: Segments of observations of trajectories of random vector field on 2-torus. Right: Persistence diagram generated from these segments. Here $N = 650, n = 25, t = 1, T = 2.5$

As a result, we get a set of time series plotted on the left display of the Figure 3. The persistence diagram is shown on the right, still clearly matching the homology of $\mathbb{T}^2$.

4.3. Chaotic Dynamical System: Lorenz System

We now look at the standard example of a chaotic dynamical system given by the Lorenz dynamical system in $\mathbb{R}^3$ (Gilmore, 1998):

$$\begin{aligned}\dot{x} &= \sigma(y - x)\\ \dot{y} &= x(\rho - z) - y\\ \dot{z} &= xy - \beta z\end{aligned}$$

Here σ, ρ, β are parameters which we set to $10, 28, \frac{8}{3}$ respectively.

It is known that the system exhibits chaotic behaviour for these values of parameters. We simulate one long trajectory from this dynamical system and split it into 100 pieces (the hyperbolicity of the system makes this essentially equivalent to randomly sampling shorter trajectories). The resulting trajectories are plotted on the left display of the Figure 4.

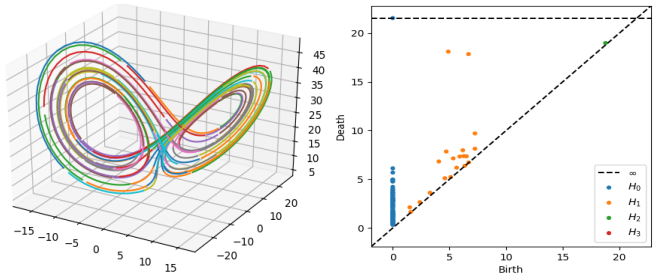

Figure 4: Segments of trajectories of the Lorenz system. Right: Persistence diagram generated from these segments. Here $N = 100, n = 25, t = 20, T = 0.25$

For the slack distance derived from the tautological output function $g = (x, y, z)$, the persistence diagram is shown as the right display of the Figure 4.

These results require some interpretation. While the ambient space carrying the vector field is $\mathbb{R}^3$, the trajectory is essentially supported by a 2-dimensional *template* (Birman and Williams, 1983), homotopy equivalent to the wedge of two circles. As we do not sample trajectories starting far away from this template, it is not surprising we recover its topology: indeed, the ranks of the homology groups should be 1 in the dimension 0, 2 in the dimension 1, and vanish otherwise, in agreement with the right display on the Figure 4.

We also repeat the experiments after passing the trajectories through a random polynomial output of degree 3 with coefficients sampled from a Gaussian distribution. This gives us a set of time series observations as shown in Figure 5. The corresponding persistence diagram still matches the homology of a wedge of two circles.

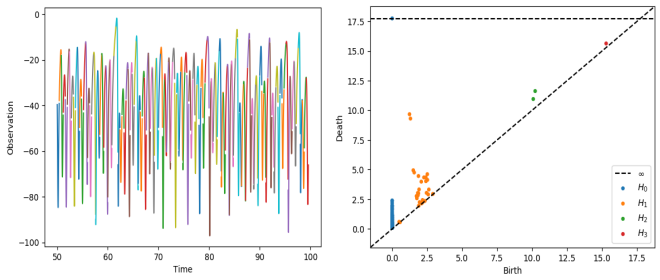

Figure 5: Segments of observations of trajectories of the Lorenz system. Right: Persistence diagram generated from these segments. Here $N = 150, n = 25, t = 10, T = 0.25$

The results become even more interesting when we increase the length T of the fragments of trajectories. When this time T is large enough, the segments of trajectories will hit the template's barrier (where the flow bifurcates to two different branches in the template) twice. Conceptually, this would lead to a restructured template that will now have the homotopy type of the wedge of 4 circles.

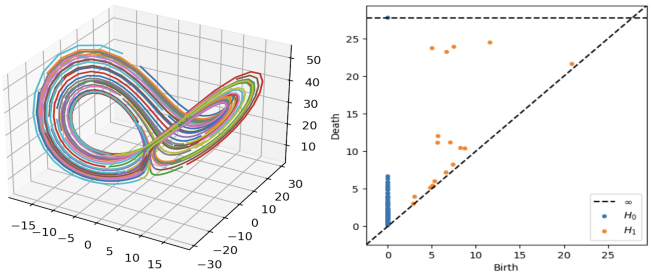

Figure 6: Segments of longer trajectories of the Lorenz system. Right: Persistence diagram generated from these segments. Here $N = 100, n = 25, t = 20, T = 0.5$

And indeed, this is exactly what we observe in the Figure 6.

Perhaps the most intuitive way to think about this phenomenon is to compare with the picture a topological data analysis would paint of a Cantor set $X \subset [0, 1]$ obtained by iteratively removing the middle third from the intervals, starting with the unit one: at the low resolution $1/18 \leq r < 1/6$, the union of r-balls with centers at the points of X are homotopy equivalent to two points; at the resolution $1/54 \leq r < 1/18$, that is 4 points and so on. Given that the attractor of the the dynamical systems of Lorentz type are cantori, this doubling of the corresponding homologies is not unexpected after all.

Acknowledgments

The authors were partially supported by the AFOSR MURI *Hybrid Dynamics - Deconstruction and Aggregation*.

References

Joan S Birman and Robert F Williams. Knotted periodic orbits in dynamical systems i: Lorenz's equations. *Topology*, 22(1):47–82, 1983.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113:3932–3937, 2016.

Marko Budišić, Ryan Mohr, and Igor Mezić. Applied Koopmanism. *Chaos*, 22(4):047510, December 2012. ISSN 1054-1500. doi: 10.1063/1.4772195. URL https://doi.org/10.1063/1.4772195.

Kathleen Champion, Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116: 22445–22451, 2019.

Ralph L Cohen, John DS Jones, and Graeme B Segal. Morse theory and classifying spaces. *preprint*, 1995.

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

Steven C Ferry and Boris L Okun. Approximating topological metrics by riemannian metrics. *Proceedings of the American Mathematical Society*, 123(6):1865–1872, 1995.

Robert Gilmore. Topological analysis of chaotic dynamical systems. *Reviews of Modern Physics*, 70(4):1455–1529, October 1998. ISSN 0034-6861, 1539-0756. doi: 10.1103/RevModPhys.70.1455. URL https://link.aps.org/doi/10.1103/RevModPhys.70.1455.

Bronisław Jakubczyk. Existence and uniqueness of realizations of nonlinear systems. *SIAM Journal on Control and Optimization*, 18(4):455–471, 1980. doi: 10.1137/0318034.

J. Latschev. Vietoris-Rips complexes of metric spaces near a closed Riemannian manifold:. *Archiv der Mathematik*, 77(6):522–528, December 2001. ISSN 0003-889X. doi: 10.1007/PL00000526. URL http://link.springer.com/10.1007/PL00000526.

Samuel E. Otto, Gregory R. Macchio, and Clarence W. Rowley. Learning nonlinear projections for reduced-order modeling of dynamical systems using constrained autoencoders. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33, 2023.

Jose A. Perea and John Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15(3):799–838, 2015.

Emily Riehl. *Category theory in context*. Courier Dover Publications, 2017.

Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324:81–85, 2009.

B. De Schutter. Minimal state-space realization in linear system theory: an overview. *Journal of Computational and Applied Mathematics*, 121(1):331–354, September 2000. ISSN 0377-0427. doi: 10.1016/S0377-0427(00)00341-1. URL https://www.sciencedirect.com/science/article/pii/S0377042700003411.

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.

A. J. van der Schaft. On realization of nonlinear systems described by higher-order differential equations. *Mathematical systems theory*, 19(1):239–275, December 1986. ISSN 1433-0490. doi: 10.1007/BF01704916. URL https://doi.org/10.1007/BF01704916.