ARTICLE IN PRESS
Physica D ▮ (▮▮▮▮) ▮▮▮–▮▮▮
Contents lists available at ScienceDirect
Physica D
journal homepage: www.elsevier.com/locate/physd

Exploring the topology of dynamical reconstructions

Joshua Garland [a,*], Elizabeth Bradley [a,b], James D. Meiss [c]

[a] Department of Computer Science, University of Colorado, Boulder, CO, USA
[b] Santa Fe Institute, Santa Fe, NM, USA
[c] Department of Applied Mathematics, University of Colorado, Boulder, CO, USA

HIGHLIGHTS

- Time-delay embedding is used to build a simplicial complex from a scalar time series.
- Downsampled data gives a small set of landmarks as vertices in the complex.
- A witness relation determines the connections in the complex between landmarks.
- Dependence of the homology of the complex on parameters is studied.

ARTICLE INFO

Article history:
Received 5 June 2015
Received in revised form
15 February 2016
Accepted 9 March 2016
Available online xxxx

Keywords:
Topology
Delay-coordinate embedding
Nonlinear time-series analysis
Computational homology
Witness complex

ABSTRACT

Computing the state-space topology of a dynamical system from scalar data requires accurate reconstruction of those dynamics and construction of an appropriate simplicial complex from the results. The reconstruction process involves a number of free parameters and the computation of homology for a large number of simplices can be expensive. This paper is a study of how to compute the homology efficiently and effectively without a full (diffeomorphic) reconstruction. Using trajectories from the classic Lorenz system, we reconstruct the dynamics using the method of delays, then build a simplicial complex whose vertices are a small subset of the data: the "witness complex". Surprisingly, we find that the witness complex correctly resolves the homology of the underlying invariant set from noisy samples of that set even if the reconstruction dimension is well below the thresholds for assuring topological conjugacy between the true and reconstructed dynamics that are specified in the embedding theorems. We conjecture that this is because the requirements for reconstructing *homology* are less stringent: a homeomorphism is sufficient—as opposed to a diffeomorphism, as is necessary for the full dynamics. We provide preliminary evidence that a homeomorphism, in the form of a delay-coordinate reconstruction map, may exist at a lower dimension than that required to achieve an embedding.

1. Introduction

Topology is of particular interest in dynamics, since many properties – the existence of periodic orbits, transitivity, recurrence, entropy, etc. – depend only upon topology. This idea is commonly exploited in the computational topology community, often using the Conley index of isolating neighborhoods, to study dynamical invariants [1]. However, computing topology from time series can be a real challenge. First, one typically has only *scalar* data, not the full trajectory, and hence one must begin by reconstructing the full dynamics from that data—e.g., via delay-coordinate reconstruction. Success of this reconstruction procedure depends on several free parameters. In practice, the embedding theorems provide little guidance regarding how to choose these parameters. A number of creative strategies have been developed for doing so, but these methods require good data and input from a human expert. Moreover, the delay-coordinate reconstruction machinery (both theorems and heuristics) targets the computation of dynamical invariants like the correlation dimension and the Lyapunov exponent. If one just wants to extract the topological structure of an invariant set, as we show in this paper, a scaled-back version of that machinery may be sufficient. Nevertheless, there are issues of parameter choice here, as in the standard approach. Moreover, real-world data sets have finite length, nonzero sampling interval, limited precision, and may be contaminated by noise. In the face of these issues, one obviously cannot compute the topology to arbitrary precision, but computations can still be useful to extract information about the large-scale features.

* Corresponding author.
E-mail address: joshua.garland@colorado.edu (J. Garland).

http://dx.doi.org/10.1016/j.physd.2016.03.006
0167-2789/

Coarse-graining the topological analysis of data also addresses another issue: the associated computations are expensive, and that expense grows with the number of simplices in the complexes that one constructs during that process. The pioneering work in this area used cubical complexes and multivalued maps for this purpose [2], and these results can be computationally rigorous even in the face of noise. For more efficiency, one can use a simplicial complex that follows the natural geometry of the data—e.g., the *witness complex* of [3]. To construct a witness complex, one chooses a set of "landmarks", typically a subset of the data, that become the vertices of the complex. The connections between the landmarks are determined by their nearness to the rest of the data—the "witnesses". Two landmarks in the complex are joined by an edge, for instance, if they share at least one witness. As described in Section 2, there are many possible definitions for a witness "relation". The one that we use includes a scale parameter, ε, that is intended to provide a measure of noise immunity. The ideas of persistent homology [4,5] can be used to choose ε, build the complex, and then explore the changes in its topology with changing reconstruction dimension.

Our initial work on this approach suggests that *the witness complex correctly resolves the homology of the underlying invariant set even if the reconstruction dimension is well below the thresholds for which the embedding theorems assure smooth conjugacy between the true and reconstructed dynamics.* This paper reports upon an exploration of that conjecture in the context of the classic Lorenz system and suggests some implications and applications. To set the stage for that discussion, the rest of this section gives a brief review of delay-coordinate reconstruction. The witness complex is covered in more depth in Section 2, which also describes the notion of persistence and demonstrates how that idea is used to choose scale parameters for a complex built from reconstructed time-series data. In Section 3, we explore how the homology of such a complex changes with reconstruction dimension.

Delay-coordinate reconstruction [6,7] is arguably the most well-established technique for obtaining the dynamics of a system from scalar time-series data. Suppose that $\vec{Y}$ is a point on a compact invariant set $M \subset \mathbb{R}^d$, and $\vec{Y}(t)$ represents its trajectory. A smooth *measurement function* $h : M \to \mathbb{R}$ gives rise to a scalar time series, $x(t) = h(\vec{Y}(t))$, from that trajectory. Then the delay-coordinate map, $F : M \to \mathbb{R}^m$

$$F(\vec{Y}(t); h, m, \tau) = (x(t), x(t-\tau), \dots, x(t-(m-1)\tau)), \qquad (1)$$

is almost always a diffeomorphism whenever $\tau > 0$ and m is large enough, i.e., $m > 2d_{box}$, where d_{box} is the box-counting dimension of M [8]. When these conditions are met, the reconstructed attractor and the true attractor are diffeomorphic, and thus certainly have the same topology. The left panel of Fig. 1 shows an example: a trajectory from the classic Lorenz system [9]. The middle panel shows the corresponding time series of the x coordinate of that trajectory (i.e., $h(x, y, z) = x$), and the right panel shows a delay-coordinate reconstruction using $\tau = 174T$, where T is the interval between points in the time series. Note that a reconstruction dimension of five ($m = 5$) is required in order to satisfy the $m > 2d_{box}$ requirement for this attractor, since $d_{box} \approx 2.06$. Of course, it is not easy to display the 5D picture; Fig. 1(c) shows a 3D projection of this reconstruction.

In practice, one is presented with a scalar time series so that the dimension, d, of the original state space is unknown, and one cannot compute d_{box} without first embedding the data. Thus, choosing the reconstruction dimension m is a challenge. There are a number of heuristics for doing so. Perhaps the most well-known is the family of false near-neighbor methods pioneered in [10]. The basic idea behind this class of methods is to increase the reconstruction dimension until the geometry of the neighbor relationships stabilizes; this is taken to indicate that any false crossings created by the measurement function h have been eliminated and the dynamics are properly unfolded. The choice of the delay τ also plays a role in this unfolding. Though the theorems only require $\tau > 0$, in practice one needs to ensure that τ is large enough to make the coordinates numerically independent, but not so large that the coordinates become causally unrelated [11]. The standard approach for this – which we used to select the τ value in Fig. 1(c) – is to calculate the time-delayed average mutual information of the time series and choose τ at the first minimum of that curve [12]. There are many other methods for estimating both m and τ; see [13] for a deeper discussion. All of these methods are subtle and subjective. Invoking them and interpreting their results requires good data and expert knowledge; the false-near neighbor method, for instance, typically overestimates the embedding dimension when noise is present in the time series—something that is unavoidable in experimental data.

In this paper, we adopt the philosophy that one might only desire knowledge of the topology of the invariant set, and we conjecture that this might be possible with a lower reconstruction dimension than that needed to obtain a true "embedding". That is, the reconstructed dynamics might be *homeomorphic* to the original dynamics at a lower dimension than that needed for a diffeomorphically correct embedding. We will return to this idea below.

2. Witness complexes for dynamical systems

To compute the topology of data that sample an invariant set of a dynamical system, we need a complex that captures the shape of the data but is robust with respect to noise and other sampling issues. To do so *efficiently*, the complex should have as few simplices as possible while still accurately representing the topology, i.e., it should be parsimonious. A witness complex is an ideal choice for these purposes. Such a complex is determined by the reconstructed time-series data, $W \subset \mathbb{R}^m$ – the *witnesses* – and an associated set $L \subset \mathbb{R}^m$, the *landmarks*, which can (but need not) be chosen from among the witnesses. The landmarks form the vertex set of the complex; the connections between them are dictated by the geometric relationships between W and L. In a general sense, a witness complex can be defined through a relation $R(W, L) \subset W \times L$. As Dowker noted [14], any relation gives rise to a pair of simplicial complexes. We will use one: a point $w \in W$ is a witness to an abstract k-dimensional simplex $\sigma = \langle l_{i_1}, l_{i_2}, \dots, l_{i_{k+1}} \rangle \subset L$ whenever $\{w\} \times \sigma \subset R(W, L)$. The collection of simplices that have witnesses is a complex relative to the relation R. For example, two landmarks are connected if they have a common witness—this is a one-simplex. Similarly, if three landmarks have a common witness, they form a two-simplex, and so on.

There are many possible definitions for a witness relation R. One very natural construction is to use the matrix $D(W, L)$ of distances $D_{ij} = \|w_i - l_j\|$ to define R. Sorting each row of this matrix from smallest to largest determines the set of landmarks that are closest to the ith witness. One relation corresponds to assigning a cut-off, which thereby determines the simplices witnessed by w_i. For example, one can choose a fixed number (viz., k-nearest neighbors), a strict size (neighbors within some distance), or an increment. The first concept gives the "weak witness complex" of de Silva and Carlsson [3], but suffers from the problem that there is no limit on the distance to the nearest neighbors and thus a simplex might be too spread out. The second notion seems too restrictive: a portion of the invariant set M that has a low density may not be covered enough to be represented in the complex. The third idea is a compromise and gives the notion of an ε-weak witness [15], or what we call a "fuzzy" witness [16]: a point witnesses a simplex if all the landmarks in that simplex are within ε of the closest landmark to the witness:

Please cite this article in press as: J. Garland, et al., Exploring the topology of dynamical reconstructions, Physica D (2016), http://dx.doi.org/10.1016/j.physd.2016.03.006