Determing Periodicity in Data

Kelly Spendlove
Department of Mathematics
Rutgers University
110 Frelinghusen Rd, Piscataway NJ 08854
kelly.spendlove@rutgers.edu

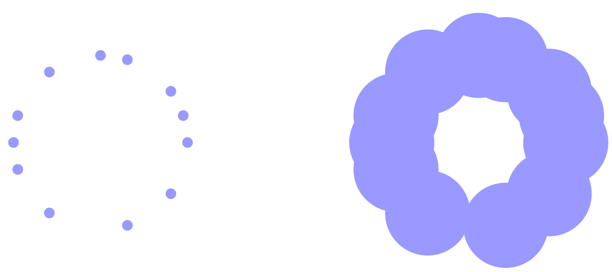
Keywords:
Topological Data Analysis
Takens' Embedding
Diffusion Maps

Background

In the last few years high-throughput technologies have enabled the efficient and inexpensive collection of massive amounts of data. In many cases the data are high dimensional and being generated by some nonlinear system. In such a situation one is interested in both the geometry of the data and the action of the unknown nonlinear system. The analysis of such data remains one of the pressing challenges for the future.

It is important to approach such analysis with the understanding that over the past few decades dynamical systems theory has established that nonlinear systems can exhibit extremely complex behavior *at all scales* with respect to both system variables and parameters. Such complex behavior proven in theoretical work has to be contrasted with the capabilities of application. Experimental data is typically noisy, of limited precision and may be generated by an underlying chaotic system. System parameters will rarely be known exactly and the nonlinearities often cannot be derived from first principles. This contrast suggests that discovering robust features is of greater importance than a detailed understanding of the fine structure, and that when attempting to understand multiscale phenomena one must extract relatively crude invariants of the data to compare to any model. That is, the *resolution* at which one analyzes the problem is of fundamental importance.

The past decade has seen the development of *persistent homology*, a technique for extracting multiscale topological structure from noisy data by considering the effect of resolution upon the topology. Persistent homology attempts to understand a continuous object from a discrete set $\mathbb{X}$ (for instance, an underlying manifold being sampled) by investigating how the topological structure of $\mathbb{X}$ changes as it is progressively thickened.



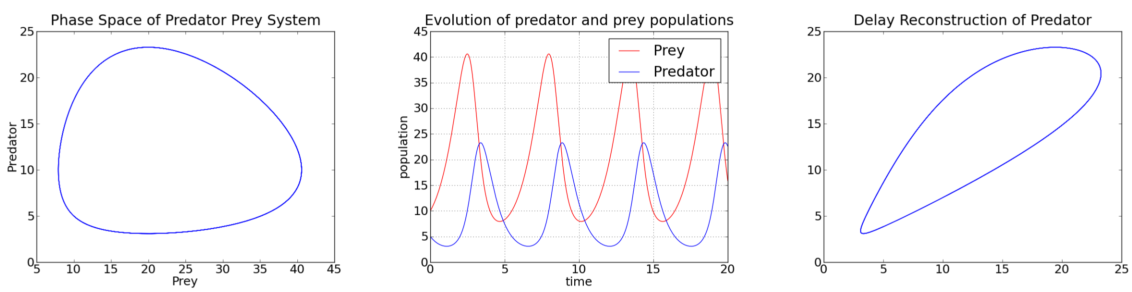
(Left) A sampled circle (Right) For a large thickening we obtain a space with similar topology to a circle

Formally, persistent homology rests on the idea of a defining a filtration $\mathbb{X} = \mathbb{F}_0 \subseteq \mathbb{F}_1 \subseteq \ldots \subseteq \mathbb{F}_n$. In the example above, the sublevel sets $\{f^{-1}(-\infty, a]\}$ of the function $f(x) := \min_{y\in\mathbb{X}} d(x,y)$ form the filtration. The inclusions $\mathbb{F}_i \hookrightarrow \mathbb{F}_j$ induce homomorphisms in homology and persistence captures how the homology changes. For instance, homology classes that exist in $\mathbb{F}_i$ may be trivial in $\mathbb{F}_j$. A homology class appearing $\mathbb{F}_i$ and disappearing at $\mathbb{F}_j$ has *persistence* $j - i$ and is represented as the ordered pair (i, j). Informally, persistence encodes a lifespan for a homology generator, with the ordered pair giving its *birth* and *death*. The collection of these pairs for $\mathbb{X}$ can be represented as a multiset of points in $\bar{\mathbb{R}}^2$, entitled a *persistence diagram*, denoted $\mathrm{Dgm}(\mathbb{F})$. The space of persistence diagrams can be endowed with a computable metric and the resulting metric space is complete [3]. Most significantly, the *stability theorem* [3] shows that if two filtrations are close their persistence diagrams are close; in the case of sublevel sets of f and g this gives $d(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq ||f - g||_\infty$. In this sense, the persistence diagram is a robust tool for analysis.

Problem Statement

As described in the introduction, there are many challenges inherent in the analysis of dynamical systems. Given these challenges, we wish to start in one of the simplest situations: determining whether the data is periodic.

By now, the classical approach to the analysis of data given by nonlinear systems is Takens' embedding theorem [5], which gives conditions under which one may reconstruct the topological properties of the system (such as a periodic orbit). In [4], they use persistent homology to determine whether there are any holes with significant lifespan in embedded data, which would suggest the data is periodic. However, Takens' theorem only guarantees the orbit will look *topologically* like a circle; there is no reason to believe the orbit should embed into a 'nice' circle (such as in our example). Therefore there may not be a generator with significant lifespan in the embedded data.



(Left) Phase space of a predator-prey system which provides us with (Middle) a time series of a predator prey system (Right) By using Takens' theorem we can use observations of the predator time series to reconstruct the periodic orbit in $\mathbb{R}^2$

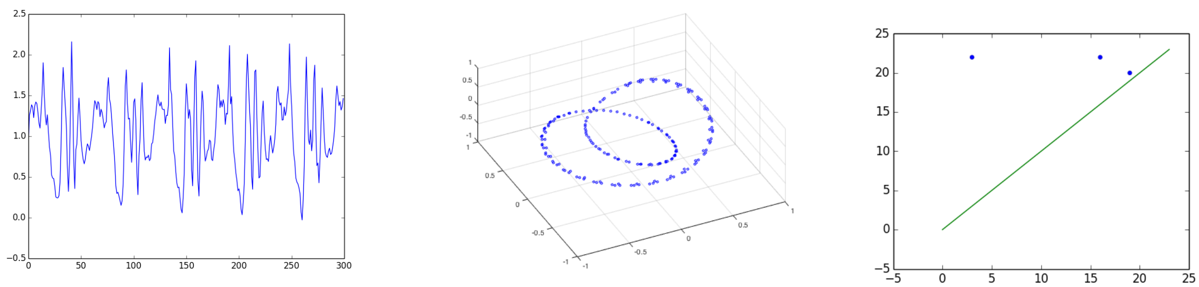
Recently, Berry et al. have had insight into the geometry of the Takens' embedding, showing that the higher the embedding dimension the more the manifold collapses to the most stable dynamical variables [1]. Roughly, this translates to the higher the dimension in which we embed the data, the easier it becomes to capture the essential dynamical behavior. With this result, in combination with the nonlinear dimensionality reduction, one may extract dynamics on different time scales [1]. It is with these results that I propose a technique to amplify the topological signal of the periodic orbit.

More formally, we will embed the data in a very high dimensional space, then define a diffusion operator on the data and use the eigenvalues and eigenvectors to compute the *diffusion distance* [2]. We will use this distance as the metric with which we compute persistent homology. The diffusion distance should allow the homological features to be more effectively extracted, even in the presence of noise. Furthermore, while computing the 1st dimensional persistent diagram does depend on the particular embedding, it is independent of the ambient dimension in which we embed. In a fashion, this implies that embedding into higher dimensions does not affect the time for computation.

Broader Impacts

One broader impact I am very interested in exploring is within the realm of the Quantified Self movement [6]. 'Quantified selfers' use technology to collect a multitude of aspects of their daily lives: macro/micronutrients consumed, blood oxygen levels, blood glucose concentrations, sleep quality, physical performance. Many of these individuals have already collected vast amounts of data although many lack the techniques to analyze such data. The method presented above addresses a very crude question one could ask: is my data periodic?

As an example application, I performed a simple experiment. Using my iPhone, I recorded acceleration data, which is a time series of 3-dimensional vectors. For simplicity, I chose a single coordinate as a time series to embed using Takens' theorem.



(Left) Time series of coordinate of acceleration (Middle) Data after using Takens' theorem to embed in $\mathbb{R}^{99}$ and embedding with the first three eigenvectors (Right) The 1st dimensional persistence diagram for the embedded data using the diffusion distance as the metric. The point furthest off the diagonal is a strong indication of the presence of a periodic orbit.

References

[1] T. Berry, R. Cressman, Z. Greguric-Ferencek, and T. Sauer, Time-scale separation from diffusion-mapped delay-coordinates, *SIAM Journal of Applied Dynamical Systems*, 12 (2013): 618-649.

[2] R. Coifman and S. Lafon, Diffusion maps. *Applied and Computational Harmonic Analysis*, 21 (2006): 5-30.

[3] H. Edelsbrunner, and J. Harer, *Computational topology: an introduction.* American Mathematical Society, 2010.

[4] J. Perea, and J. Harer, Sliding windows and persistence: an application of topological methods to signal analysis, *Foundations of Computational Mathematics*, 15 (2015): 799-838.

[5] F. Takens, Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence*, 898 (1981): 366-381.

[6] G. Wolf. (2010, June). The quantified self. [Video file]. Retrieved from `http://ed.ted.com/lessons/gary-wolf-the-quantified-self`.