TOPOLOGICAL PERSISTENCE IN JACOBI SETS

DAVID F. SNYDER

1. Background

This report is generated as a result of work done by this author in the IMA's New Directions: Computational Topology course. The purpose of the paper is to discuss results of one of the problems given by the course leaders to the participants. The problem given was to investigate the concept of topological persistence when computing Jacobi sets of a pair of real-valued functions f, g defined on a point-cloud data set approximating a smooth manifold $\mathbb{M}$, so $f : \mathbb{M} \to \mathbb{R}$. Of particular interest is the special case where f is a time-varying Morse function and g is the time elapse function. The specific version of this mentioned is the case $\mathbb{M} = \mathbb{N} \times \mathbb{S}^1$, where $\mathbb{N}$ is a smooth manifold undergoing a smooth periodic motion, $f, g : \mathbb{N} \times \mathbb{S}^1 \to \mathbb{S}^1$ and $g(u) = u$ for all $u \in \mathbb{S}^1$. A real world example of the latter is the electrostatic potential on the molecular skin of a stable molecule. We also addressed the case of affine time $f, g : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ and $g(u) = u$ for all $u \in \mathbb{R}$, though this was not in the original question. Applications to contour cleaning were also suggested.

Two (essentially the same) metrics for topological persistence are proposed. These should allow users of the Edelsbrunner-Harer algorithm a method of filtering from its output topological features which are spurious or likely to have arisen from noise in the data. The method is simple to implement and speedy to compute, but will not necessarily remove all such topological noise.

We are interested in applying these metrics to simulated (or real) data sets to investigate their potential utility. Application areas include molecular motion, spatial time-series analysis, and investigation of the distribution of electrostatic potential (and other forces) on the molecular skin of a molecule undergoing periodic motion, with automatic location of 'hot spots'.

In [1], the notion of Jacobi set is defined by Edelsbrunner and Harer. Let $\mathbb{M}$ be a smooth manifold of dimension n embedded in a Euclidean space and let $f, g : \mathbb{M} \to \mathbb{R}$ be smooth functions defined on $\mathbb{M}$. The

Date: July 29, 2004.

Jacobi set of f and g, denoted $\mathbb{J}(f, g)$ (in fact, Edelsbrunner and Harer define the Jacobi set for any k-tuple ($k > 1$) of smooth real-valued Morse functions defined on $\mathbb{M}$) is the collection of points for which the gradients of f and g are codirectional. We will not describe here the algorithm given for computing it, referring the reader to the course notes or to the reference paper.

Persistence, in computer science, typically refers to a measurement of the lifetime of a set with regards to its existence as a data structure, or as a component thereof. The algorithm for computing Jacobi sets never kills an edge once it is introduced to the Jacobi set. This made it difficult (for this mathematician) to determine, in the general case, what is meant by "the lifetime of a set." Depending on the data set, significant features may not be introduced into the Jacobi set until late in the process. In the case of a time-varying Jacobi set, there is more hope for a definition of persistence, as long as we complete computation of the Jacobi set for a given time before proceeding to the next time slot (this has the disadvantage of requiring simultaneous data measurements across $\mathbb{M}$); in fact, the time-varying case guided our thinking). We think our contribution can be used to address both cases, and hence be applied in a wide variety of applications. Moreover, it is easy to generalize these metrics to the case where there are more than two smooth functions. Still, there is at least one difficulty, described to us by Edelsbrunner in application to computing contours, which our approach cannot address; in essence, our work here addresses the global topological noise in a component, but says nothing of local noise.

So, let the reader be forewarned that by the word 'persistence' we mean a type of topological persistence, which may not be related to persistence in a data structure. It does seem possible to us to tie the concepts together, but we have not done so.

In other words, we may not have addressed the intended question, perhaps, but our interpretation of it.

2. The case $f, g : \mathbb{M} \to \mathbb{R}$

In this section, we describe the general setting of [1] in which we define a measure of persistence. We assume $\mathbb{M}$ to be a smooth, closed manifold embedded in d-dimensional Euclidean space, $\mathbb{R}^d$. We also assume we are given two smooth functions $f, g : \mathbb{M} \to \mathbb{R}$. More precisely, we assume we are given a finite set $V \subset \mathbb{R}^d$ estimating $\mathbb{M}$ and functions (also called, *v*ia abuse of notation, f and g, $f, g : V \to \mathbb{R}$ which are extended to PL functions on $\mathbb{M}$ approximating the original f and g. For any $v \in V$, we use, whenever convenient, subscript notation $f_v = f(v)$ and similarly for g.

Now, presume we have run the Jacobi Set Algorithm and we have the components of $\mathbb{J}(f,g)$. Let $K_1, K_2, \ldots, K_l$ denote these components. For each $i \in \{1, \ldots, l\}$, we define the *persistence of K_i with respect to g* as

$$\pi(i,g) = \max_{v\in V}(g_v) - \min_{v\in V}(g_v).$$

In other words, $\pi(i,g)$ computes the range of the interval $g(K_i)$. The reader can supply a similar definition for persistence of the ith component with respect to f. [Perhaps the more descriptive term 'the range of g over K_i' would be better?]

2.1. **A 'box metric' measure of persistence.** As a first candidate for a definition of persistence of K_i within the system given by f and g, we first look at the value

$$\pi(i) = \max_{1\le i\le l, h\in\{f,g\}}(\pi(i,h)).$$

In other words, select the larger of the two ranges. The higher the persistence, the more interesting the component should be. If a component has a very small persistence with respect to other members of the family of components, then it shows much less fluctuation with respect to f and g and, so, may be due to error in the estimations of the functions at the vertices or, if not, it may reflect that this component does not have a large global impact and may be safely filtered out.

This measure of persistence has the advantage of being relatively inexpensive and fast to compute.

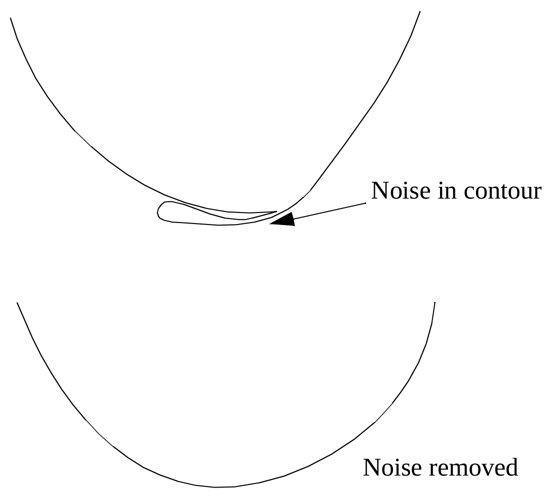
Nonetheless, there are at least two dissatisfactory aspects to this definition. One consideration (the most irritating) is that a component may be of significance and yet have topological noise we would wish to filter out; see the figure on the next page. A second consideration is that the filtering as given is not controllable by an end-user of the algorithm.

2.1.1. *Weighted maximum persistence.* One possible solution to the latter difficulty is to define a weighted version. For example, let λ_1, λ_2 be non-negative real numbers, not both zero. We can define the weighted persistence of the ith component to be

$$\pi(i) = \lambda_1 \cdot \pi(i,f) + \lambda_2 \cdot \pi(i,g)$$

where the user is allowed to vary the parameters λ_1, λ_2.

Still, these measures are very crude and do not take into account that a component may contain quite a bit of action with respect to g (or f) between its minimum and maximum values.



2.2. **Total variation as a measure of persistence.** Let K_i be any component of the Jacobi set, as computed by the algorithm of [1]. Let $K_i^{(1)}$ denote the edges of K_i. Let $e \in K_i^{(1)}$ and let u, v be the vertices of e. Then define

$$\delta_g(e) = |g(u) - g(v)|.$$

Define $\delta_f(e)$ similarly. Define the *total variation in K_i with respect to g* as

$$\Delta_g(K_i) = \sum_{e \in K_i^{(1)}} \delta_g(e).$$

The total variation in K_i with respect to f is defined similarly.

We then define the *(total variational) persistence of K_i* to be

$$\pi(i) = \max(\Delta_g(K_i), \Delta_f(K_i)).$$

Or, we can define a weighted version (a definite linear combination of Δ_g and Δ_f, in a similar fashion as before.

Again, the operations required are simple, and fast to compute.

2.2.1. *Generalizations.* We could generalize the above construction by using a different metric than the l^1 metric to define Δ_g. For example,

$$\Delta_g(K_i) = \sqrt{\sum_{e \in K_i^{(1)}} \delta_g(e)^2}.$$

However, computation costs need to be considered, and there appears to be no advantage to the increase in mathematical operations required.

3. Applications

In this section, we briefly discuss these measures of topological persistence in the context of applying them to contours and time-varying Morse systems.

3.1. **Contours.** In this situation, we have $\mathbb{M}$ as a surface in $\mathbb{R}^3$, a viewing direction $\alpha \in \mathbb{S}^2$, and two linearly independent vectors $\beta, \gamma \in \mathbb{S}^2$ orthogonal to α, with $f(x) =< x, \beta >$ and $g(x) =< x, \gamma >$. The Jacobi set is the contour of $\mathbb{M}$. We refer to the value of f as height and the value of g as width, to keep the discussion transparent.

Our first measure of persistence of a component gives us the larger of the width or height of the component. If both are very small relative to viewing scale, the component will not be contributing much to the contour. However, if either width or height is relatively large (imagine the contour of a cubical box), the contribution should be real. The weighted version would allow for differences in scale units for f and g.

Our second candidate measure for persistence of a component should catch even more features. Imagine looking down the axis of a very long spiral staircase, but from very far above. The contour of the spiral will have small width and height (relative to our view), but the total variation in length is significant. This our total variation metric should pick up, at least in the scaled version.

3.2. **Affine time-evolution.** Assume $f, g : \mathbb{M} \times \mathbb{R} \to \mathbb{R}$ and $g(x,t) = t$. As usual, $f_t(x) = f|_g^{-1}(t)$, for any $t \in \mathbb{R}$. To keep our discussion intuitive, we call f a height function for a landscape and think of the height as varying with time. Then our first measure of topological persistence, for any component of the Jacobi set, returns either the range in heights over the lifetime (with respect to t) of the component or it returns the lifetime of the component, whichever is larger (the weighted version allows us to adjust for multi-scales). If this measure is small, then the component has very little height change and does not live for very long, and could likely be noise-generated. This still allows us to catch short-lived peaks in the evolution of f, as well as small features in the evolution of the landscape height that persist for long times.

Our second measure of persistence is slightly better in that even if the range in values of f and g over a component is small, if the height oscillates significantly, this feature will be captured by the measure.

3.2.1. *Periodic time-evolution.* Here we assume $f, g : \mathbb{M} \times \mathbb{S}^1 \to \mathbb{S}^1$ and that $g(x, u) = u$.

In this case, we have to modify our constructions, since $\mathbb{S}^1$ is not a totally ordered set.

4. Questions

There are still many interesting questions remaining. How to eliminate large errors within a component we wish to keep? The structure of the Jacobi set (birth-death pairs, saddles) should yield additional information. Also, it would be interesting to derive a curvature estimate at a vertex based on its lower link, and use this to detect where errors in edge selection may occur. Also, does our result tell us anything about persistent or partially persistent subsets of the data structure itself?

References

1. Herbert Edelsbrunner and John Harer, *Jacobi sets of multiple morse functions*, Foundations of Computational Mathematics, ed. F. Cucker (London), Cambridge Univ. Press, 2004, p. to appear.

Department of Mathematics, Texas State University, San Marcos, Texas 78666

E-mail address: dsnyder@txstate.edu