

WHAT IS . . .

Persistent Homology?

Shmuel Weinberger

In memory of my friend, Partha Niyogi (1967–2010)

Consider the art of Seurat or a piece of old newsprint. The eye, or the brain, performs the marvelous task of taking the sense data of individual points and assembling them into a coherent image of a continuum—it infers the continuous from the discrete.

Difficult issues of a similar sort occur in many problems of data analysis. One might have samples that are chosen nonuniformly (e.g., not filling a grid), and, moreover, one is constantly plagued by problems of noise—the data can be corrupted in various ways.

Pure mathematicians have problems of this sort as well. One is often interested in inferring properties of an enveloping space from a discrete object within it or, in reverse, seeking commonalities of all the discrete subobjects of a given continuous one. To give one example, this theme is a central one in geometric group theory, in which a typical problem, going back to Furstenberg and Mostow, asks to reconstruct a connected Lie group from a lattice in it.

And, ubiquitously in analysis, one often tries to get information about a function from approximations to it. For instance, any function uniformly close to $z \rightarrow z^n$ on the complex plane necessarily has at least n roots (with multiplicity).

Because topology is essentially a qualitative field, it is perhaps not surprising that there has been a development of some common topological technology for these problems. Needless to say,

Shmuel Weinberger is professor of mathematics at the University of Chicago. His email address is shmuel@math.uchicago.edu.

The author was partially supported by NSF 0805913, NSF-SGER grant DMS-0852227, and STOMP, a DARPA program.



The Seine at La Grande Jatte by Georges Seurat.

the deepest aspects of these problems all have the idiosyncrasies of their particular application domains. The focus here is on what is common to them.

For simplicity in what follows, we will take all homology groups to have coefficients in a field.

Definition. Suppose that we have $X = \{X_r | r \in \mathbf{R}\}$ a nested sequence of spaces (satisfying mild technical conditions), parameterized by the real numbers. We define the k th persistent homology $PH_k(X)$ by the formula:

$$PH_k(X) = \prod H_k(X_r).$$

The product on the right is an awful object: formally it is an uncountable dimensional vector space, but there is a reasonable way to make sense of this, taking into account the fact that

the inclusion $X_r \subset X_s$ for $r < s$ induces a map $H_k(X_r) \rightarrow H_k(X_s)$.

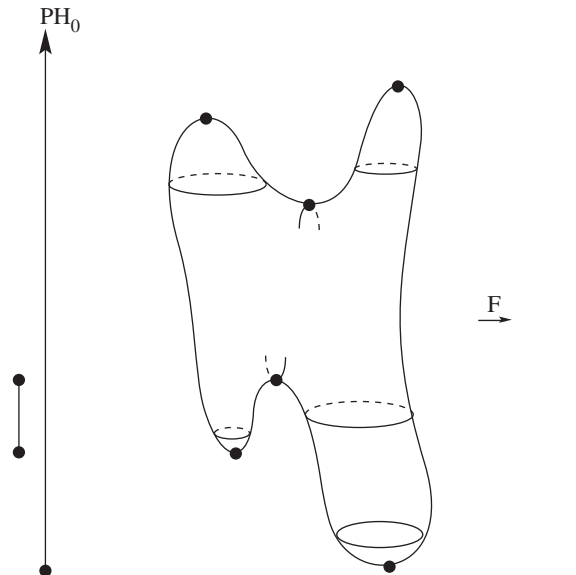
If one considers an element of $H_k(X_r)$, then one can follow it along “further in time” to larger s 's and watch whether or not it dies. If it dies, there is a smallest (or actually infimal) s at which it does. If $H_k(X_r)$ is finite dimensional, then it is possible to give it a basis so that every element in the basis has a well-defined moment of death and a sum of basis elements dies in $H_k(X_s)$ if and only if each of the basis elements with nonzero coefficients in its description dies in $H_k(X_s)$. (This is a consequence of elementary linear algebraic considerations.)

A concrete example is given by a positive real valued function $f : Z \rightarrow \mathbf{R}^+$; we can think of $f^{-1}[0, r]$ as an approximation to Z . From this point of view, there are some homology classes that are “born immediately” at $r = 0$, i.e., are in the image of $H_k(f^{-1}\{0\})$, but others are born somewhat later. Other classes might be born at one moment and die somewhat later.^{1,2}

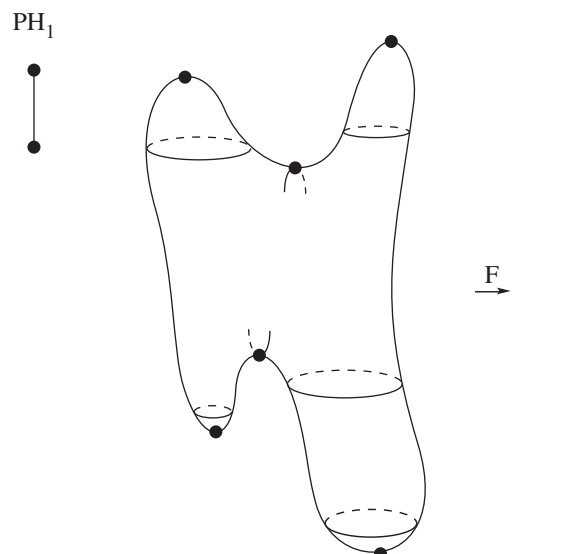
We can summarize this in a “barcode” that encodes these births and deaths. Here is a picture of a function on the 2 sphere and the persistent homology in dimensions 1 and 2 of the associated filtration of the sphere.

In the function setup, persistent homology can be thought of as a variant of Morse theory. Each critical point of a Morse function either announces the birth of a homology class or is the death knell of another. Thus critical points are always visible in the barcode of homology either in the index dimension or one lower.

Now let us return to the kinds of examples that motivated our discussion. If one starts with a submanifold M of Euclidean space (say, a human face) and samples many points from it (pointillistically), we can consider the function $f(x) =$ smallest distance from x to any of the sample points. The homology of these sublevel sets can be computed using computer algorithms. If s is larger than the density of the samples, then at that scale one has all of the homology classes of M present, and if s is not too large, the homology will be that of M (although, after a while, the homology classes of M will die).



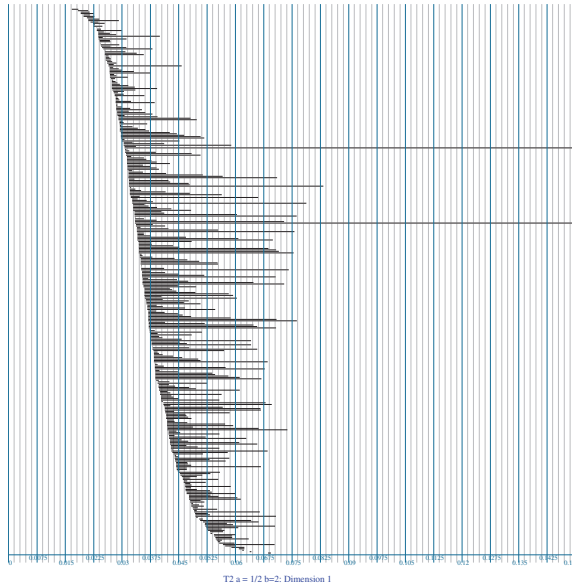
This picture shows, on the left vertical axis, the birth of 0-dimensional homology classes (essentially components) and their deaths (the coalescing of components). For 1-dimensional homology it looks like this:



A typical example of PH_1 taken from 1400 samples of a 2-torus looks like this (with the persistence parameter being drawn horizontally, as is the usual custom):

¹Even if Z is compact, if the function is just assumed continuous, then at particular levels one can have infinite dimensional homology. However, the homology that persists over any positive length interval is necessarily finite dimensional—already a good reason to look at such homology groups.

²If the function f is nice, then the persistent homology can be thought of as a derived pushforward sheaf.



In this example, there was no noise; the short bars are indicative of the irregularities and inefficiencies of the sampling. Much remains to be learned about the shape of barcodes of data, but there are theorems that tell us how to use *PH* to compute genuine homology from samples that are sufficiently dense and not too noisy.

In almost all applications, a key role is played by a stability theorem that asserts that for nearby functions (or filtrations), the barcodes cannot be very far apart, in a precise quantitative sense. The topology on the space of barcodes has two barcodes being close if by ignoring “short intervals” we can match up their long intervals in such a way that corresponding intervals have nearby endpoints. A picture should suffice:



because, aside from the “short” intervals, the barcodes can be closely aligned.

The stability theorem of Edelsbrunner, Cohen-Steiner, and Harer says that if functions are at most C apart, then the long bars of their persistence diagrams correspond—and indeed their initial and terminal points cannot be shifted by more than C , although short intervals of length $< C$ can be arbitrarily different. The reader can think about the diagrams of the functions $y = x^2$ and $x^2 + \sin(10000x)$ on \mathbf{R} to see why this is true.

When we apply these ideas to a discrete group with the word metric,³ we form a different nested

³For any finitely generated group, we define the distance between two group elements to be the number of multi-

sequence of spaces. These spaces are all simplicial complexes. The space associated with the number r (for $r > 1$) has k -simplices spanned by $k + 1$ group elements that are a pairwise distance at most $\log(r)$ apart.

It is easy to check that these simplicial complexes for different generating sets or for different uniform lattices in the same Lie group are close to one another (essentially the distance is determined by the length of the words describing one generating set in terms of the other) and therefore have close persistent homology. This can be used to show that certain homological properties of groups only depend on “coarse quasi-isometry type” and, e.g., agree for different uniform lattices in the same Lie group. As an example, essentially due to Gersten, the last dimension for which there is a long, i.e., infinitely long, interval in the persistent homology detects cohomological dimension.

For our last example, consider a Riemannian manifold M . Our space will be $X = \Lambda M = \{\gamma : S^1 \rightarrow M\}$, the space of smooth loops in M . Our function is given by the “log energy”

$$\log E(\gamma) = \log \int \langle \gamma'(t), \gamma'(t) \rangle dt,$$

where \langle, \rangle denotes the Riemannian inner product. Note that while the log energy depends on the metric, for two metrics on the same compact manifold, the difference between these functions on X is bounded. As a result, by stability, the persistence homology of this loop space—up to finite distance—is an invariant of the manifold, i.e., is independent of the metric. We are most of the way toward proving the following theorem, essentially due to Gromov:

Theorem. *Let M be a compact Riemannian manifold. The question of whether there is a universal constant C so that every closed nullhomotopic geodesic of length L can be contracted through curves of length at most CL is independent of the metric on M ; indeed, it only depends on the fundamental group of M . This condition is equivalent to the nonexistence of arbitrarily long persistence intervals in PH_0 of the component of the constant loops in ΛM .*

The final statement of the theorem explains the clause immediately preceding it. The geometric condition about geodesics depends only on the fundamental group (and not on the manifold) because this is true, almost by definition, for the 0 dimensional homology of the space of loops. Examples for which this condition does not hold are finitely presented groups with unsolvable

plications by generators it takes to go from one element to the other. Although this depends on the generating set, many of the large-scale properties of this metric space do not, as is explained in Roe’s article [3].

word problem (for a bound on the size of the persistence intervals could be used to give an algorithm for solving the word problem). For manifolds with such fundamental groups, the theorem asserts the existence of many interesting closed nullhomotopic geodesics. More information can be found in Alex Nabutovsky's talk at the 2010 ICM.

The terminology of persistence homology can be viewed as an example of applied applied math. The needs of applied math have given us a very convenient vocabulary for expressing certain questions and arguments in pure mathematics.

Acknowledgments

I would like to thank Jonathan Block, Gunnar Carlsson, Frederick Chazal, Sasha Dranishnikov, Herbert Edelsbrunner, Benson Farb, Steve Ferry, Rob Ghrist, Alex Nabutovsky, Partha Niyogi, and Steve Smale for many very helpful conversations on these matters.

References

- [1] G. CARLSSON, Topology and data, *Bull. AMS* 46 (2009), no. 2, 255–308.
- [2] H. EDELSBRUNNER and J. HARER, Persistent homology: A survey, *Surveys on Discrete and Computational Geometry. Twenty Years Later*, 257–282 (J. E. Goodman, J. Pach, and R. Pollack, eds.), Contemporary Mathematics 453, Amer. Math. Soc., Providence, Rhode Island, 2008.
- [3] J. ROE, What is a coarse space? *Notices AMS* 53 (2006), 668–669.



香港城市大學
City University
of Hong Kong





Worldwide Search for Talent

City University of Hong Kong is a dynamic, fast developing university distinguished by scholarship in research and professional education. As a publicly funded institution, the University is committed to nurturing and developing students' talent and creating applicable knowledge to support social and economic advancement. Currently, the University has six Colleges/Schools. Within the next few years, the University aims to recruit **200 more scholars** in various disciplines from all over the world, including **science, engineering, business, social sciences, humanities, law, creative media, energy, environment,** and other strategic growth areas.

Applications are invited for:

Associate Professor/Assistant Professor
Department of Mathematics [Ref. A/621/49]

The Department of Mathematics has a strong mission to conduct first-class research in applied mathematics and provide high quality education in mathematics.

Duties : Conduct research in areas of Applied Mathematics, teach undergraduate and postgraduate courses, supervise research students, and perform any other duties as assigned.

Requirements : A PhD in Mathematics/Applied Mathematics/Statistics with an excellent research record.

Salary and Conditions of Service
Remuneration package will be driven by market competitiveness and individual performance. Excellent fringe benefits include gratuity, leave, medical and dental schemes, and relocation assistance (where applicable). Initial appointment will be made on a fixed-term contract.

Information and Application
Further information on the posts and the University is available at <http://www.cityu.edu.hk>, or from the Human Resources Office, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong [Fax : (852) 2788 1154 or (852) 3442 0311/email : hrojob@cityu.edu.hk].

Please send the application with a current curriculum vitae to Human Resources Office. **Applications will be considered until positions are filled.** Please quote the reference of the post in the application and on the envelope. The University reserves the right to consider late applications and nominations, and not to fill the positions. Personal data provided by applicants will be used for recruitment and other employment-related purposes. The University is an equal opportunity employer.

City University of Hong Kong is an equal opportunity employer and we are committed to the principle of diversity. We encourage applications from all qualified candidates, especially applicants who will enhance the diversity of our staff.