Topological Data Analysis

Kevin P. Knudson

Department of Mathematics
University of Florida
kknudson@honors.ufl.edu
http://kpknudson.com/

June 19, 2013

Discrete Morse Theory

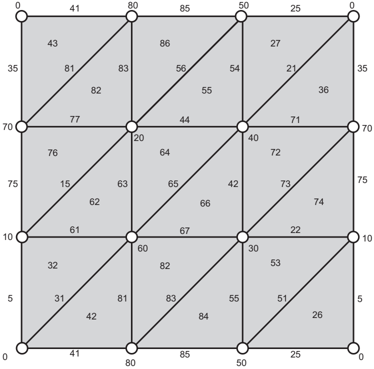
Given a triangulated space M (built with vertices, edges, triangles, tetrahedra, etc.), a *discrete Morse function* is a function f that assigns a number to each simplex in M with the following constraints for every p-simplex $\alpha^{(p)}$:

1. $\#\{\beta^{(p+1)} > \alpha^{(p)} | f(\beta) \leq f(\alpha)\} \leq 1$;
2. $\#\{\tau^{(p-1)} < \alpha^{(p)} | f(\tau) \geq f(\alpha)\} \leq 1$.

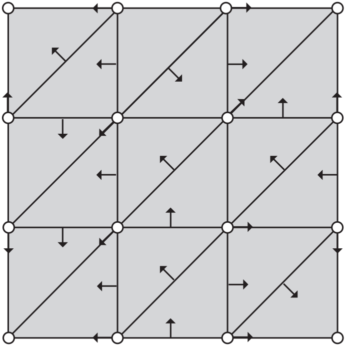
Think: the function values increase (generically) as the dimension of the simplices increase.

A simplex α is *critical* if both of the sets above are empty. So a critical vertex corresponds to a local minimum, a critical n-simplex ($n =$ top dimension) is a local maximum, and the other critical simplices are saddles of various indicies.

Example: the torus



The critical cells are $f^{-1}(0)$, $f^{-1}(42)$, $f^{-1}(44)$, and $f^{-1}(86)$.



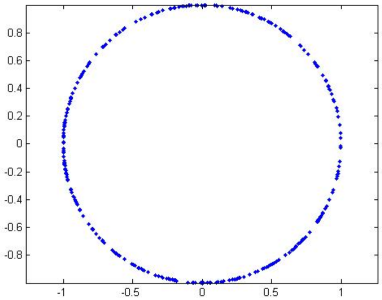
The associated gradient field.

In joint work with H. King and N. Mramor, we develop an algorithm which takes a function sampled at points in some space and produces a discrete Morse function on the space.

This has obvious applications to data analysis–often one has a set of measurements over a region (e.g., temperature, barometric pressure, etc.) and wishes to understand how the function changes in the region. Where are the local maxima and minima, for example? This algorithm finds all the critical points.

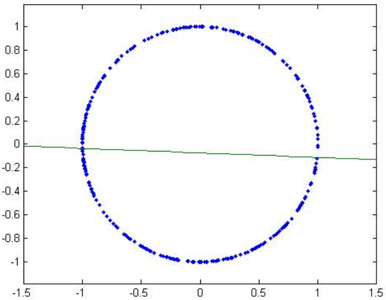
An implementation for regions in $\mathbb{R}^3$ is available at `http://www.math.umd.edu/~hking/MorseExtract.html`

Suppose we have a collection of points:



It's pretty clear that this is a circle, but how could we convince ourselves of that?

Here's the best linear fit:



The field of algebraic topology has many techniques to find geometric structures in a space. In particular, the *homology groups* measure the number of holes of various dimensions in a space. More accurately, the i-dimensional homology group measures the number of i-dimensional objects in the space that cannot be filled in *inside the space* by an $(i+1)$-dimensional object.

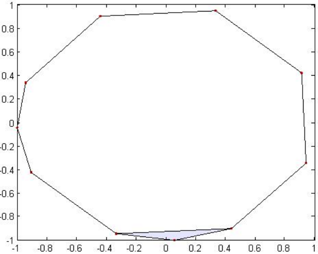
There is a lot of heavy machinery that has been developed over the last 100 or so years to calculate these things, provided one has a good model for the space in question. More recently, however, new techniques have emerged to handle large point clouds (i.e., data sets).

The problem with point clouds is that it is often not obvious that they have been sampled from a reasonable space. But we go ahead and assume that they have, try to reconstruct the underlying space, and then compute its homology.

Question: How do we know we have the right space?
Answer: We don't. So we build a nested sequence of spaces, compute the homology of each, and see how it evolves over time. Homology classes that persist for long intervals may be real, significant geometric objects in our point cloud. We capture this information with an object called a *barcode*.

Example

Here is a collection of complexes built from 10 of the 100 data points on the circle I showed earlier:



and here are the associated barcodes