Lecture 23: April 4, 2017

Lecturer: Prof. Bei Wang <beiwang@sci.utah.edu>   *Scribe: Chris Brooks, Will Usher*

Thursday 4/6 Quiz Given a 1 or 2 dimensional function, give reeb graph, contour tree or mapper

Topics: Extended persistence Elevation function Protein Docking

23.1   Extended Persistence

There is something not satisfactory about ordinary persistence. Consider a "deformed torus" fig. 23.1a with a height function f. Going from bottom to top, the sublevel sets $\mathbb{X} = f^{-1}(-\infty, a]$ are a bowl at the first local min a, two bowls at the next local min b, glued bowls at the saddle point c, a bowl with a handle at the second saddle point d (note the tunnel) and so on. As we move up the object, a tunnel is formed at e which is not filled by the addition of any subsequent critical points. This tunnel indicates an important feature about the object, in its structure.
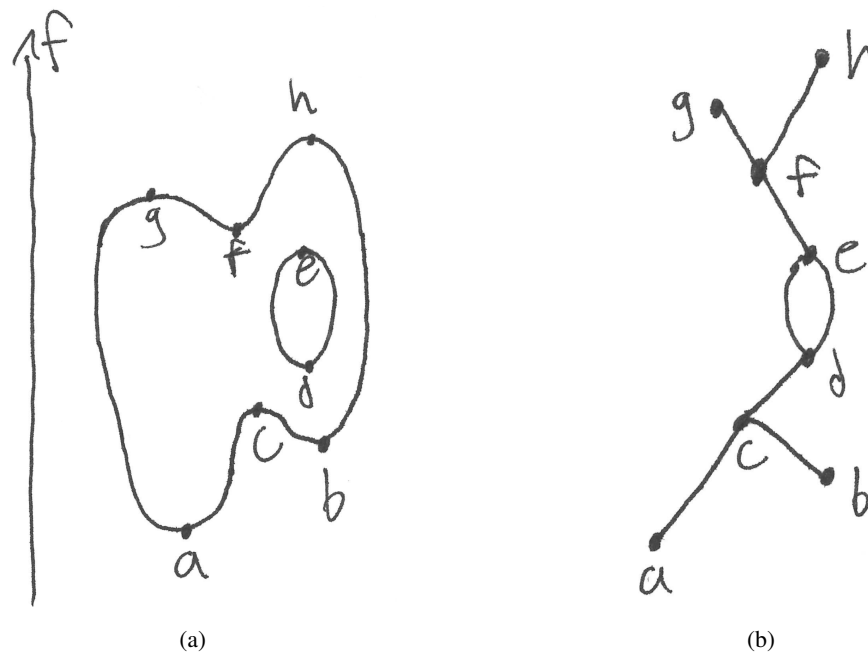

(a)   (b)

Figure 23.1: The deformed torus (a) and its Reeb graph for the height function f

We can pair critical points based on features being born at one, and killed at the next. For example, we can pair (b, c) on the basis of a component being born at b and killed at c. Same for (f, g). We can also pair the global minimum with the global maximum (a, h). This is the first part of "extended persistence" and it captures an *essential feature* of the space. We call (a, h) an *extended persistence pair*. Another essential feature would be the tunnel described by the pair of saddle points (d, e). These pairs are features created but not destroyed during the sub-level set filtration, and

(a) Bottom of deformed torus   (b) Simplify by pushing b up   (c) Simplify by pushing c down   (d) Simplify by moving both

Figure 23.2: Different methods for moving the critical points b and c to remove the lower-persistence feature.

correspond to essential features in the structure. In the Reeb graph for this space (fig. 23.1b), the smallest branches correspond to the ordinary pairs, and the remaining vertices can be associated with the extended persistence pairs. We can also do "persistence simplification". For example, if we push the point b up to, or past c, we remove the two critical points, and the protrusion associated with it. This can also be done by bringing c down to b or moving them both up and down to some equal height fig. 23.2.

These features of a topological space can be useful in analyzing the interaction of proteins and other molecules since these interactions are dependent on the compatibility of their shapes. The persistence pairs from the previous example give a sense of the size of that feature; for instance, the pair (c, d) gives the size of the corresponding protrusion by computing the difference in height between the points $|f(d) - f(c)|$. By comparing the persistence of the differente features on each protein we wish to dock, we can determine the best docking orientation, or if they are incompatible.

23.2   Elevation Function

An elevation function $E : \mathbb{X} \to \mathbb{R}$ for any point $x \in \mathbb{X}$ gives us $E(x)$, which is the persistence of x when it becomes critical. In the previous examples the critical points appeared where the normal direction aligned with the height direction. As we change the height direction, the corresponding critical points will also change (fig. 23.3).

In the first example (fig. 23.1), the elevation function f is defined by $E(f) = E(g) = |t_g - t_f|$, $E(a) = E(h) = |t_h - t_a|$, etc. If our height function is not Morse, there may be multiple critical points that can be paired with another critical point. This can happen if we sample directions on a shape to do this kind of analysis; a small perturbation can eliminate the ambiguity. As proteins can rotate and dock in different orientations we can use multiple elevation functions in different directions to determine if they are compatible in some orientation This gives us different critical points and new critical pairs which highlight the shape's features along that direcion.

23.3   Application to Protein Docking

To determine if two proteins can dock we can determine if there's a matching cavity in one that can be filled by the other, e.g. a lock and key. To check if such a compatible structure exists in two proteins we can examine their extended persistence along all directions. This would require computing the persistence on all directions of a half-sphere (due to symmetry), which in the continuous case gives an infinite number of directions. In practice, the molecular surface is represented using an *Alpha complex*, which is a tessellation created by the intersection of fixed size balls defined at each point. The radii of these balls can be chosen to be the distance at which different molecular forces have effect.
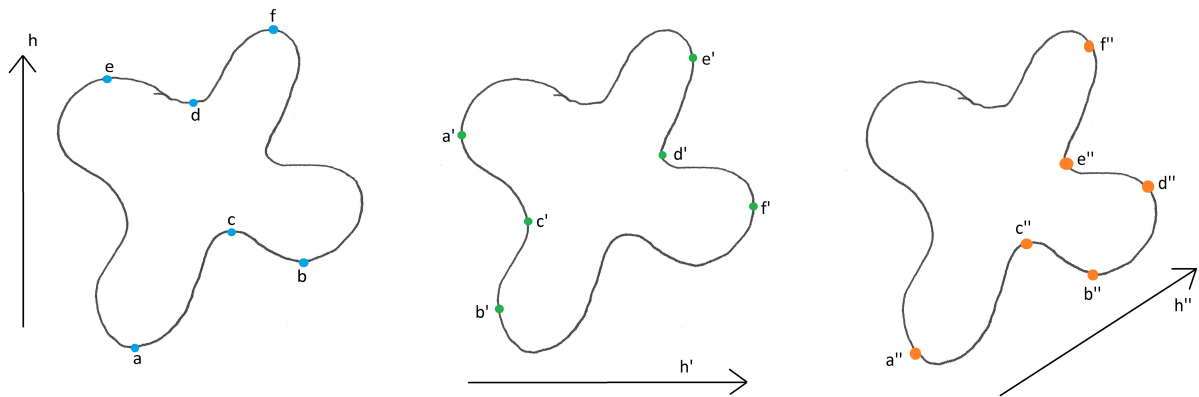
Figure 23.3: As the height function direction changes, the critical points change accordingly.
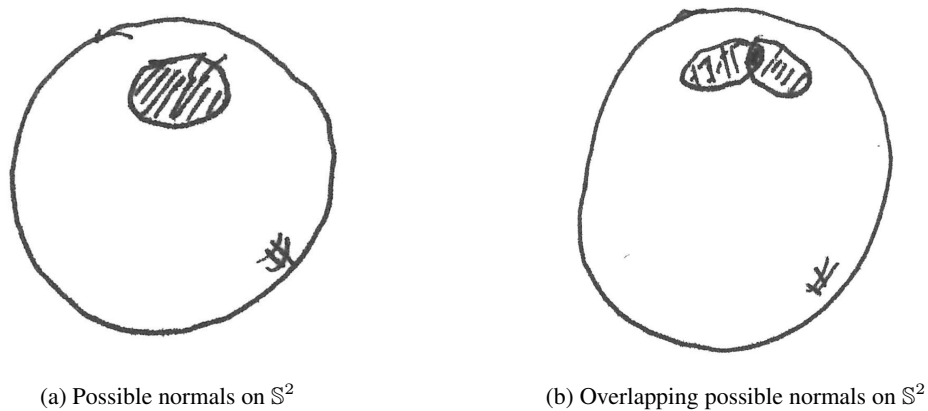

(a) Possible normals on $\mathbb{S}^2$ (b) Overlapping possible normals on $\mathbb{S}^2$

Figure 23.4: Patches on $\mathbb{S}^2$ representing possible normals for a critical point. In (b) it's possible for the two points to be paired as their normals could be the same, but additional computation is required to determine if this is actually the case.

As a result of this discretization we have a finite number of directions to test for each critical point, given by the normals of the triangles for which it is a face. To find the correct normal for the point we interpolate the normals of the different triangles, finding a patch on $\mathbb{S}^2$ (fig. 23.4a). If the patches of two critical points overlap (fig. 23.4b) they could potentially pair with each other, and additional refinement may be needed to determine if they can actually pair.

In the case of the elevation function on proteins, it is possible the function is not Morse, making the analysis more challenging. However, these degeneracies capture important information about the protein structure which would be lost if perturbed, and must be handled separately. The four possible degenerate cases are shown in fig. 23.5.

References

[WABER05]   Y. WANG, P.K. AGARWAL, P. BROWN, H. EDELSBRUNNER and J. RUDOLPH "Coarse and Reliable Geometric Alignment for Protein Docking" *Pacific Symposium on Biocomputing*, 2005
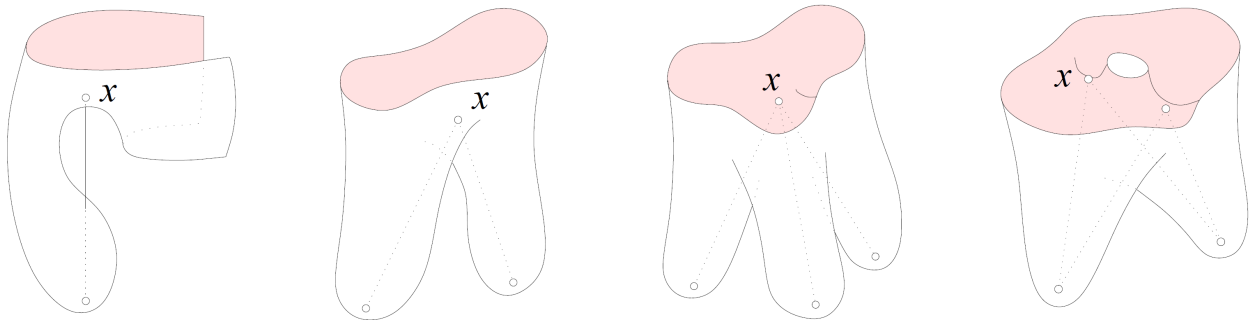
Figure 23.5: The possible special cases for protein docking, where critical points have equal elevation. From left to right are the one-, two-, three- and four-legged cases [WABER05]. In the two-legged case the saddle x projects on to the line between the two minima. In the three-legged case it projects to the triangle formed by the three minima. In the four-legged case the line between the two saddle points can be projected down and will intersect the line between the two minima