CS 6170: Computational Topology, Spring 2019
Lecture 22

Topological Data Analysis for Data Scientists

Dr. Bei Wang

School of Computing
Scientific Computing and Imaging Institute (SCI)
University of Utah
www.sci.utah.edu/~beiwang
beiwang@sci.utah.edu

March 27, 2019

Extended Persistence

(Edelsbrunner and Harer, 2010, VII.3)

Extended filtration

- $f : \mathbb{M} \to \mathbb{R}$
- $a_1 < a_2 < \cdots < a_n$: homological critical values of f
- Find interleaved values b_i:

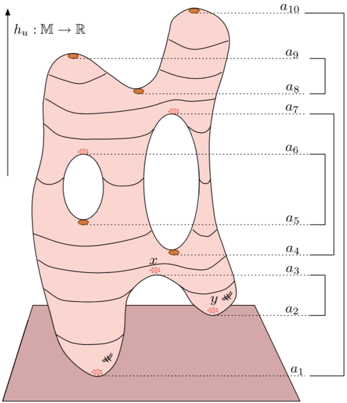
$$b_0 < a_1 < b_1 < \cdots < a_n < b_n$$

- *Sublevel sets* $\mathbb{M}_{b_i} = f^{-1}(-\infty, b_i]$, 2-manifolds with boundary
- *Superlevel sets* $\mathbb{M}^{b_i} = [b_i, \infty)$
- Fix dimension p, sequence of homology groups:

$$\begin{aligned} 0 &= \mathsf{H}_p(\mathbb{M}_{b_0}) \to \mathsf{H}_p(\mathbb{M}_{b_1}) \to \cdots \to \mathsf{H}_p(\mathbb{M}_{b_n}) \\ &= \mathsf{H}_p(\mathbb{M}, \mathbb{M}^{b_n}) \to \mathsf{H}_p(\mathbb{M}, \mathbb{M}^{b_{n-1}}) \to \cdots \to \mathsf{H}_p(\mathbb{M}, \mathbb{M}^{b_0}) = 0 \end{aligned}$$

- Absolute homology, e.g. $\mathsf{H}_p(\mathbb{M}_b)$
- Relative homology, e.g. $\mathsf{H}_p(\mathbb{M}, \mathbb{M}_b)$

Extended persistence on 2-manifold



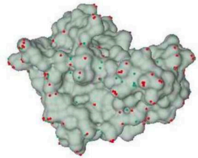
- (dim 0) $[a_2, a_3), [a_1, a_{10})$;
- (dim 1) $[a_4, a_7), [a_5, a_6), [a_6, a_5), [a_7, a_4), [a_8, a_9), [a_9, a_8)$;
- (dim 2) $[a_{10}, a_1), [a_3, a_2)$.

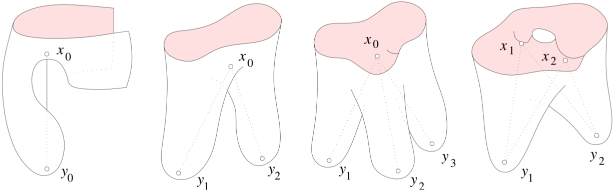
Elevation Functions and Protein Docking

Agarwal et al. (2006); Wang et al. (2011, 2005)

Motivation

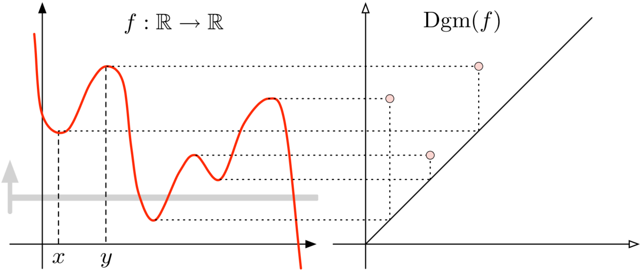
- Identify cavities and protrusions of macromolecules for the purpose of protein docking.
- Goal: compute elevation maxima faster in practice!





Persistence

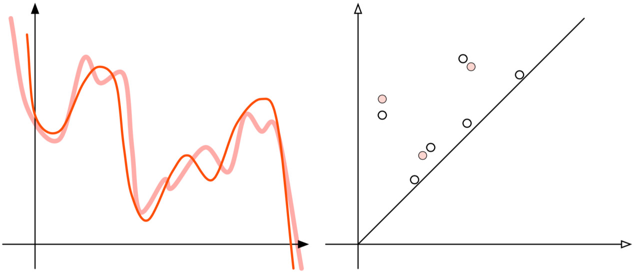
- Persistence for a single variable function $f : \mathbb{R} \to \mathbb{R}$
- f can be extended to $f : \mathbb{M} \to \mathbb{R}$
- Connectivity of sub level set changes at a critical value
- Topological feature has persistence at $f(y) - f(x)$.



Stability of persistence diagrams

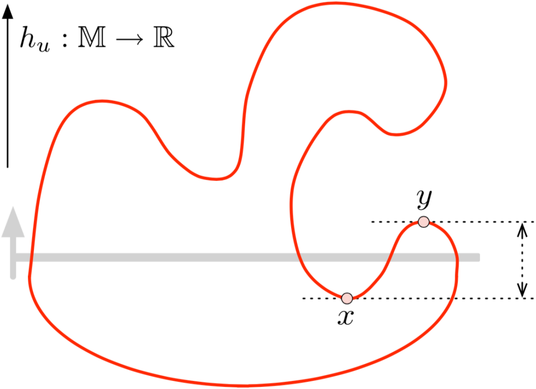
- Given $f, g : \mathbb{X} \to \mathbb{R}$, define $||f - g||_\infty = sup_{x \in \mathbb{X}} |f(x) - g(x)|$, and $d_B(\mathrm{Dgm} f, \mathrm{Dgm} g) = \inf_\gamma \sup_x ||x - \gamma(x)||_\infty$, then we have

$$d_B(\mathrm{Dgm} f, \mathrm{Dgm} g) \leq ||f - g||_\infty.$$

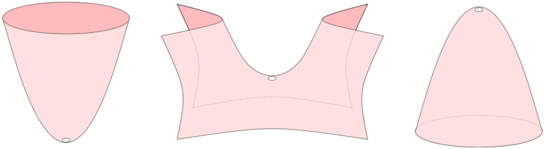


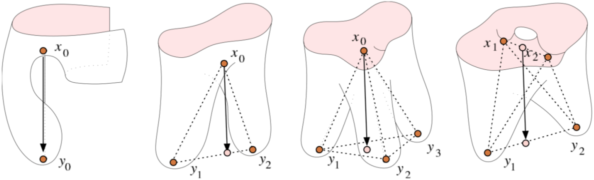
Elevation on 1-manifold

- For $u \in \mathbb{S}^1$, $p(x) = p(y) = |h_u(x) - h_u(y)|$
- $E : \mathbb{M} \to \mathbb{R}$, s.t. $E(x) = p(x)$.



Critical points and elevation maxima

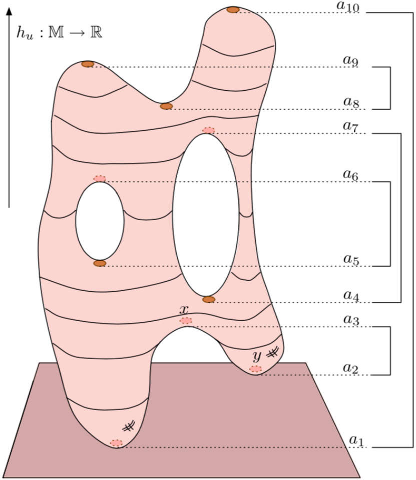




- Top: critical points.
- Bottom: elevation local maxima are sets of critical points that are paired by the persistence algorithm in a given height direction.

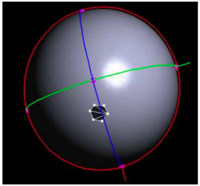
Elevation on 2-manifold

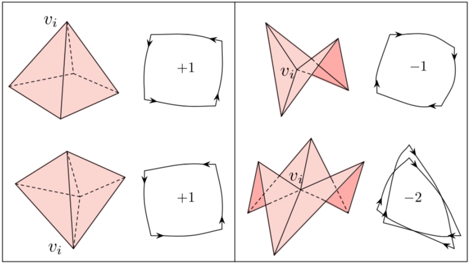
- For $u \in \mathbb{S}^2$, $p(x) = p(y) = |h_u(x) - h_u(y)|$.
- $E : \mathbb{M} \to \mathbb{R}$, s.t., $E(x) = p(x)$



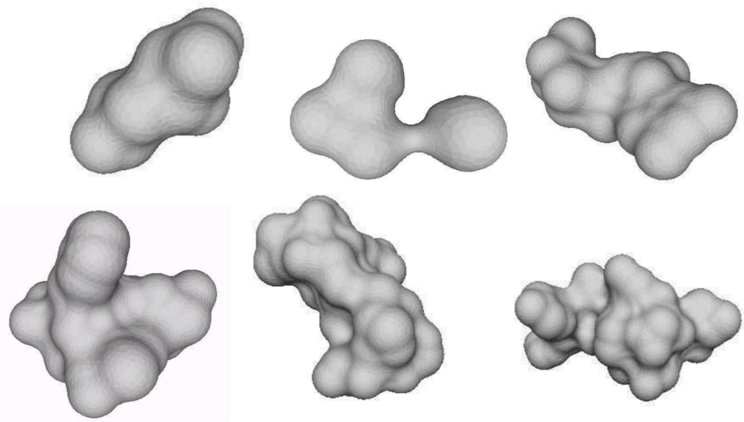
Critical region of a vertex on the Gauss sphere

- Smooth case, a vertex is critical in two directions, u and $-u$
- PL approximation, a vertex is generally critical for an entire region of directions
- The critical region of a vertex is the closure of the set of directions along which it is critical

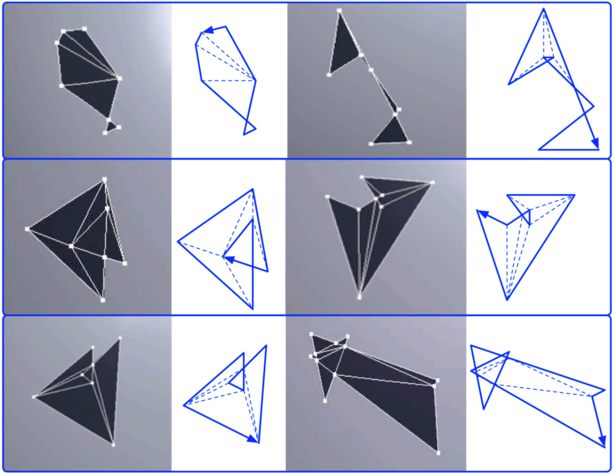




Protein surfaces



Examples of critical regions



Protein docking using elevation function

- Generate rigid motions from feature sets ϕ_A and ϕ_B obtained by analyzing the shapes of two proteins A and B
- A feature consists of two points u and its partner v with common surface normals $n_v = n_u$ and common elevation $E(v) = E(u)$.
- Its length is the Euclidean distance between them $||u - v||$.
- Each maximum of the elevation function is defined by $k = \{2, 3, 4\}$ points and give rise to $\binom{k}{2}$ features.

```
for every α ∈ Φ_A and every β ∈ Φ_B do
  if α, β form a plausible alignment then
    μ = Align(α, β);
    compute the contact and collision numbers for (A, μ(B));
    if (A, μ(B)) is valid then add μ to Γ endif
  endif
endfor;  sort Γ by contact number.
```

Wang et al. (2005)

References I

Agarwal, P. K., Edelsbrunner, H., Harer, J., and Wang, Y. (2006). Extreme elevation on a 2-manifold. *Discrete & Computational Geometry*, 36(4):553–572.

Edelsbrunner, H. and Harer, J. (2010). *Computational Topology: An Introduction*. American Mathematical Society, Providence, RI, USA.

Wang, B., Edelsbrunner, H., and Morozov, D. (2011). Computing elevation maxima by searching the gauss sphere. *Journal of Experimental Algorithmics (JEA)*, 16(1-13).

Wang, Y., Agarwal, P. K., Brown, P., Edelsbrunner, H., and Rudolph, J. (2005). Coarse and reliable geometric alignment for protein docking. *Pacific Symposium on Biocomputing*, pages 64–75.