CS 7301.003.20F Lecture 26–November 16, 2020

Main topics are #persistent_homology, #stability, and #curves.

- Today, we're going to recast stability of persistence in terms of continuous functions. Then we'll see a surprising application of stability to the length and curvature of two curves.

Sublevel Sets

- Let X be a topological space and g : X ➡ R be a continuous function.
- Given a threshold a in R, the *sublevel set* at a is the points in X with function value at most a, denoted as $X_a = g^{-1}(-\infty, a]$.
- Similar to complexes in a filtration, the sublevel sets are nested and give rise to sequences of homology groups, one per dimension.
- Let $f^{a, b}_k : H_k(X_a) \to H_k(X_b)$ be the map induced by inclusion for homology groups of a and b's sublevel sets. It's image is a *persistent homology group* and the *persistent Betti number* $\beta^{a, b}_k = \text{rank im } f^{a, b}_k$.
- Now, let's find a definition for persistence diagrams.
- We call a in R a *homological critical value* if there is no eps > 0 for which $f^{a - \varepsilon, a + \varepsilon}_k$ is an isomorphism for each dimension p, i.e., homology changes precisely at threshold a.
- Function g is *tame* if there are finitely many homological critical values and every sublevel set has finite rank homology groups.
- Let $a_1 < a_2 < \ldots < a_n$ be the homological critical values and $b_0 < b_1 < \ldots < b_n$ be interleaved values such that $b_{i - 1} < a_i < b_i$ for $1 \le i \le n$.
- All of the k-dimensional homology groups can now be expressed as
  $0 = H^{b_{-1}}_k \to H^{b_0}_k \to H^{b_1}_k \to \ldots \to H^{b_n}_k \to H^{b_{n + 1}}_k = 0.$
- For convenience, let $a_0 = -\infty$ and $a_{n + 1} = \infty$.
- For $0 \le i < j \le n + 1$, the *multiplicity* of the pair (a_i, a_j) can be defined as

$$\mu_p^{i,j} \quad = \quad (\beta_p^{b_i, b_{j-1}} - \beta_p^{b_i, b_j}) - (\beta_p^{b_{i-1}, b_{j-1}} - \beta_p^{b_{i-1}, b_j}).$$

  (except use k instead of p).
- In the *dimension k persistence diagram* of g, we draw each point (a_i, a_j) with multiplicity $\mu^{i, j}_k$.
- Unlike with filtrations, we can have multiplicities other than 0 and 1. We may also have points at infinity. For example, if g is bounded, we may birth 0-dimensional homology classes that never die, meaning they correspond to points with a second coordinate of ∞.
- We can define bottleneck distance exactly as yesterday and get the same stability

theorem. I won't prove this theorem.

- Stability Theorem for Functions: Let X be a triangulable topological space, g, g_0 : X ➡ R two tame functions, and k any dimension. Then d_B(Dgm_k(g), Dgm_k(g_0)) ≤ ||g - g_0||_infty.

Curvature of Cycles

- We're now going to shift gears a bit by discussing curvature of cycles.
- Let gamma : S^1 ➡ R^2 be a smooth cycle with derivative of all orders. It may have self intersections.
- The *speed at a point* gamma(s) is the magnitude of the velocity vector, ||gamma'(s)||. The length of gamma is therefore L(gamma) = int_{s in S^1} ||gamma'(s)|| ds.
- Let's assume gamma has a constant speed parameterization, so v = ||gamma'(s)|| = 1 / (2pi) L(gamma) for all s in S^1.
- The *curvature at a point* gamma(s) is defined as kappa(s) = ||gamma''(s)|| / v^2.
- Imagine the velocity vector as we trace along the curve. It's length is constant, so it sweeps along a circle of radius v.
- The curvature is the speed at which the unit tangent vector sweeps out the unit circle as we move the point with unit speed along the curve. That's why we divide by the speed twice: first to compensate for the length of the velocity vector and second to to compensate for the actual speed we move along gamma.
- The *total curvature* of gamma is the distance swept by the unit tangent vector
  K(gamma) = int_{s in S^1} v kappa(s) ds
- As a simple example, the circle of radius r has total length 2pi r and total curvature 2pi. Edelsbrunner does the actual calculations in his notes.
- There's another way we can derive length and curvature using intersections with lines in the plane and maxima and minima in various directions.
- Consider a unit length segment in the plane. The set of lines crossing the segment at angle phi form a strip of width sin phi. If we integrate over all angles, we get int_{phi = 0}^pi sin phi dphi = [-cos phi]_0^pi = 2. In other words, the total number of intersections with lines is twice the length of the segment.
- Let's extend the idea to a cycle by writing lines as the preimage of a linear function.
- Given a direction u in S^1, let h_u : R^2 ➡ R be defined as h_u(x) = <u, x>. The line with normal direction u and offset z is h^{-1}_u(z).
- The intersections between gamma and this line correspond to the preimage of h^{-1}(z) where composition h = h_u circ gamma.
- Therefore, the length of the curve is the Cauchy-Crofton formula:

$$L(\gamma) \quad = \quad \frac{1}{4} \int_{u \in \mathbb{S}^1} \int_{z \in \mathbb{R}} \operatorname{card}\left(h^{-1}(z)\right) \mathrm{d}z \, \mathrm{d}u.$$

We divide by 2 to account for the doubling of the length discussed earlier and divide by 2 again, because each slope is represented by u and -u.

- Now for curvature, again let u in S^1, but overload h_u : S^1 ➔ R as the height function in that direction h_u(s) = <u, gamma(s)>.
- For generic u, this height function has a finite number of local maxima and minima.
- We can define curvature as the length of circle traveled by the outward unit normal vector (as opposed to tangent vector). The number of maxima of h_u is the number of times the unit normal passes over u in S^1, and the minima is the number of times it passes over -u in S^1.
- Let c(h_u) be the number of maxima and minima of h_u. We have

$$K(\gamma) \quad = \quad \frac{1}{2} \int_{u \in \mathbb{S}^1} c(h_u)\, \mathrm{d}u.$$

Relating Length to Total Curvature

- Suppose gamma fits entirely within the unit disk in the plane, i.e. im gamma subseteq B^2. Then gamma must turn a lot to fit within the disk. Fáry's Theorem states L(gamma) ≤ K(gamma) in this case.
- There's a generalization of this theorem involving two smooth cycles gamma, gamma_0 : S^1 ➔ R^2 (not necessarily in B^2).
- We define (for this lecture) the *Fréchet distance* between them as the infimum over all homeomorphisms the largest distance between corresponding points.
  - d_F(gamma, gamma_0) = inf_eta max_u ||gamma(u) - gamma_0(eta(u))|| where eta : S^1 ➔ S^1 is a homeomorphism and u in S^1.
- Generalized Fáry Theorem: Let gamma, gamma_0 : S^1 ➔ R^2 be two smooth cycles. Then |L(gamma) - L(gamma_0)| ≤ [K(gamma) + K(gamma_0) - 2pi] d_F(gamma, gamma_0).
- For example, if we take gamma_0 to be an arbitrarily tiny circle centered at the origin and im gamma subset B^2, then L(gamma_0) approaches 0, K(gamma_0) = 2pi, and d_F(gamma, gamma_0) approaches at most 1. We get the original Fáry Theorem.
- Surprisingly, we'll prove this theorem by using everything we know about persistent homology.
- Fix a direction u in S^1 and let h = h_u circ gamma and h_0 = h_u circ gamma_0 be the restrictions to the height functions to the curves.
- Almost all level sets of the form h^{-1}(z) consist of an even number of points decomposing gamma into the same number of arcs. Half of these live in h^{-1}(-infty, z].
- Let Chi(z) be the Euler characteristic of the sublevel set for z. It's equal to number of points in h^{-1}(z) minus the number of arcs in h^{-1}(-infty, z], which again is half that number of points. Therefore,

$$\int_{z\in\mathbb{R}} \operatorname{card}\left(h^{-1}(z)\right) \mathrm{d}z \quad = \quad 2\int_{z\in\mathbb{R}} \chi(z)\, \mathrm{d}z$$

- So if we want to study the difference of lengths, we can study the difference of integrals over card(h^{-1}) or the difference in these Euler characteristics.
- By the Euler-Poincaré formula, the Euler characteristic of a sublevel set is the number of components (beta_0) minus number of loops (beta_1).
- Therefore, Chi(z) is the number of points in Dgm_0(h) within quadrant Q_z = (-infty, z] x (z, infty] minus the number of points in Dgm_1(h) within the same quadrant.
- Let Chi_0 be the Euler characteristic for gamma_0. We can say the same things here.
- Now let's compare those persistence diagrams. Let eps = d_F(gamma, gamma_0) and eta : S^1 ➜ S^1 a homeomorphism such that the Euclidean distance between points gamma(s) and gamma_0(eta(s)) is at most eps + delta.
- The stability of persistence diagrams implies the bottleneck distance between Dgm_0(h) and Dgm_0(h_0) is at most eps + delta, and the same holds for the Dmg_1. Let psi_0 and psi_1 be the corresponding bijections between the bottleneck diagrams.
- Now observe that the contributions of points v in Dgm_k(h) and psi_k(v) in Dgm_k(h_0) cancel each other out except for values of z for which one point lies inside the quadrant Q_z and the other lies outside. For pairs of finite points, the integral over such values of z is at most 2(eps + delta).
- Note finite points in one diagram may be matched to the fake points in the diagonal of the other diagram. Each coordinate of a finite point contributes at most eps + delta to the integral.
- However, points at infinity must be matched to each other and we have exactly four; one each in the two diagrams of h and the two diagrams of h_0. The integral over values of z for which one point of a pair is in the quadrant here is at most eps + delta. So the total contribution from these finite coordinates is 2(eps + delta), which is (eps + delta) times the number of their finite coordinates minus 2.
- In other words, we get a contribution of at most eps + delta for each minimum and maximum of h and h_0, except for two.
- Taking delta to 0 and using the Cauchy-Crofton formula, we get

$$\begin{aligned} |L(h) - L(h_0)| &\leq \frac{1}{4}\int_{u\in\mathbb{S}^1}\int_{z\in\mathbb{R}} \left|\operatorname{card} h^{-1}(z) - \operatorname{card} h_0^{-1}(z)\right| \mathrm{d}z\, \mathrm{d}u \\ &\leq \frac{1}{2}\int_{u\in\mathbb{S}^1}\int_{z\in\mathbb{R}} |\chi(z) - \chi_0(z)|\, \mathrm{d}z\, \mathrm{d}u \\ &\leq \frac{\varepsilon}{2}\int_{u\in\mathbb{S}^1} \left[c(h) + c(h_0) - 2\right] \mathrm{d}u. \end{aligned}$$

- Finally, plug in the formula we saw for total curvature to reduce the right hand side to eps[K(gamma) + K(gamma_0) - 2pi].