MATH 529: Lecture 25 (04/14/2026)

Today: * mapper algorithm
* Reeb graph

The Mapper Algorithm

INPUT: A set of points X with a metric

Function(s) $f_j : X \to \mathbb{R}$ (or $\mathbb{R}^d$), a cover $\mathcal{U}$ of $f(X)$ $\to (U_i)_{i\in I}$

$\hookrightarrow$ all f_j together taken as $f: X \to \mathbb{R}^d$

Steps: For each $U_i \in \mathcal{U}$, decompose $f^{-1}(U_i)$ into clusters $C_{i,1}, \dots, C_{i,k_i}$. $k_i \geq 1\ \forall i$

Compute Nrv of cover of X defined by $\{C_{i,1}, \dots, C_{i,k_i}\}, i \in I$.

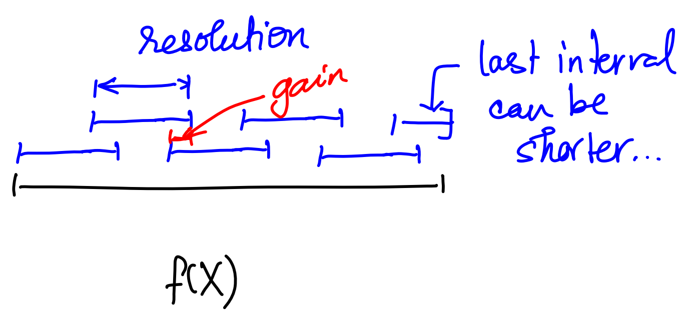
OUTPUT: The simplicial complex, which is the Nrv.

Def The functions f_i are called as **filter** functions. The metric (on X) used for clustering is called the **distance** function.

There are a lot of choices to be made by the user— which filters to use, what covers to use, etc. There are some theoretical results on when such choices are stable. But practitioners have been trying various values heuristically, and mapper has been getting used widely.

Default choice for filter functions:

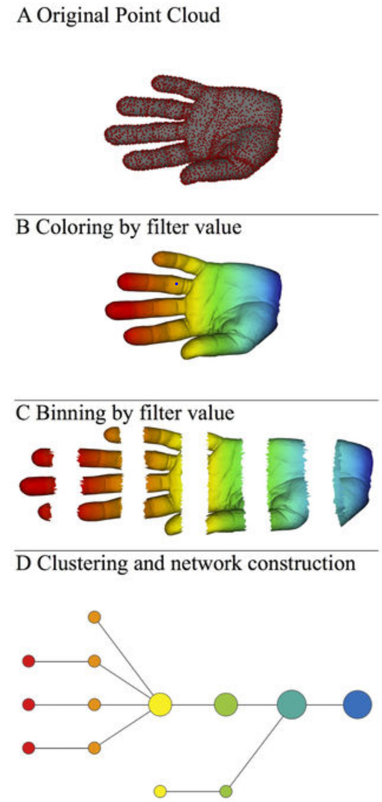
$f: X \to \mathbb{R}$



Length of (uniform) intervals: resolution

% overlap: gain

Here is another (more realistic) illustration from Lum et al. (2013). The input is ~1000 points sampled from the surface of a hand. Distance of a point from the wrist (base of hand) is used as the filter function.



The mapper representation takes us from ~1000 pts in $\mathbb{R}^3$ to a mapper (graph) with 13 nodes and 12 edges.

We now introduce some motivating mathematical concepts.

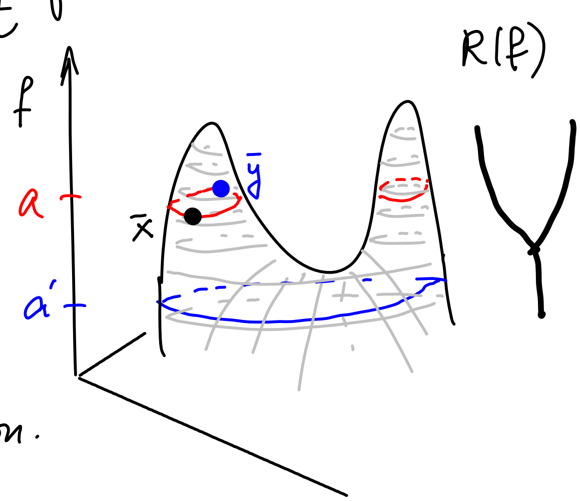
Reeb Graph

We study $\mathbb{X}$ and $f(\mathbb{X})$ and level sets of f on $\mathbb{X}$.

super/sub Level set: $\mathbb{X}_a = \{\bar{x} \in \mathbb{X} \mid f(\bar{x}) \overset{\geq}{\underset{\leq}{=}} a\}$.

The collection of all level sets forms a partition of $\mathbb{X}$.

Def We say $\bar{x}, \bar{y} \in \mathbb{X}$ are equivalent if they belong to the same connected component of a level set (or contour) of f.



Def The **Reeb graph** of f is the set of contours $R(f)$ together with the quotient topology induced by this equivalence relation.

We consider $\psi : \mathbb{X} \to R(f)$ as the contour map, where $\psi(\bar{x})$ is the contour that contains $\bar{x}$. We can show that ψ maps components to components, but could merge some cycles. Hence we get

$$\beta_0(R(f)) = \beta_0(\mathbb{X}), \text{ and}$$
$$\beta_1(R(f)) \leq \beta_1(\mathbb{X}).$$

Hence if $\mathbb{X}$ has no "holes", $R(f)$ is a truthful representation for sure. But even when $\mathbb{X}$ has holes, Reeb graphs find many applications, e.g., medical image analysis.

Morse theory studies the case where $\mathbb{X} = \mathbb{M}$, a compact manifold, and f is a Morse function (continuous function whose critical points are non-degenerate).

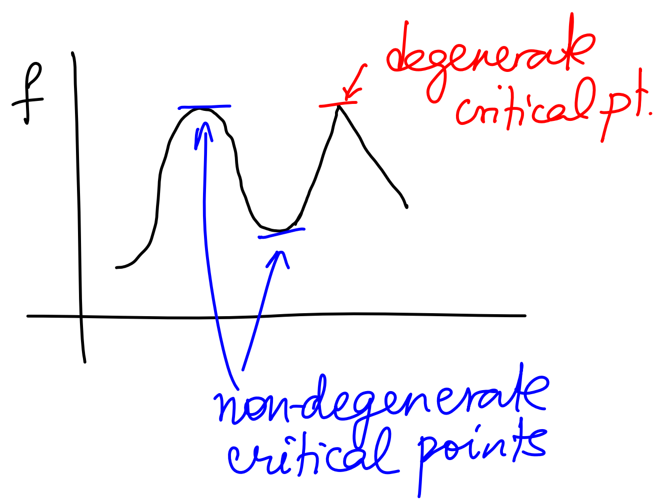
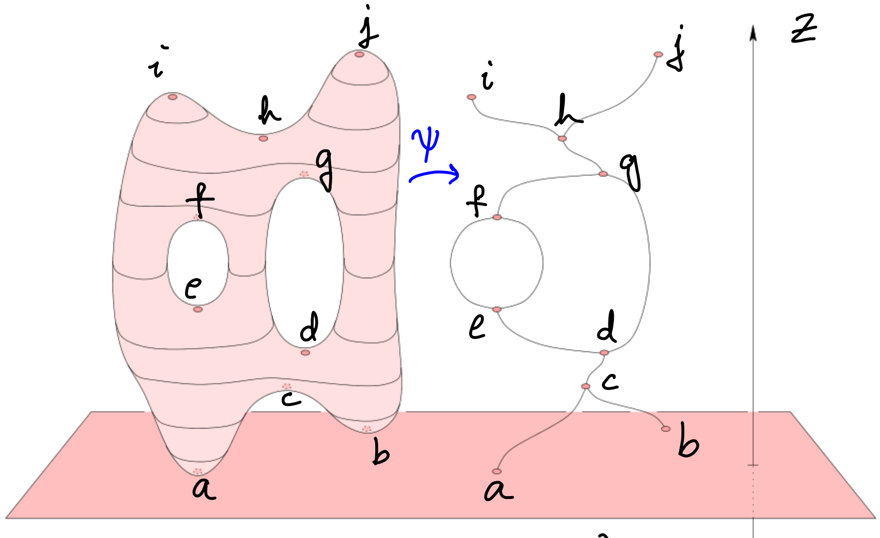


Illustration of Reeb graph

From Edelsbrunner and Harer



$(\mathbb{X}, f) \longrightarrow R(f)$

u is a node in $R(f)$ if it is the image of a critical point (under ψ).

Mapper is motivated by the Reeb graph construction, but aspires to be much more general. We will highlight some aspects of the cover choices and their (refined) pullbacks here.

Typical setting in Mapper: $f: \mathbb{X} \to Z$ a parameter space. Let $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ be an open cover of Z. We study cover of $\mathbb{X}$ by $f^{-1}(U_\alpha)$.

If $\mathcal{V} = \{V_\beta\}_{\beta \in B}$ is another cover of Z, a **map of coverings** is function $g: A \to B$ s.t. $\forall \alpha \in A$, $U_\alpha \subseteq V_{f(\alpha)}$.

Given such a map of coverings g, there is an induced map of the simplicial complexes

$$N(g): Nrv(\mathcal{U}) \longrightarrow Nrv(\mathcal{V}).$$

<u>Example 1</u> Let $\mathbb{X} = [0,n] \subseteq \mathbb{R}$, and $\epsilon > 0$. Then $I_l^{\epsilon} = (l-\epsilon, l+1+\epsilon) \cap \mathbb{X}$ for $l = 0, 1, \dots, n-1$ forms an open cover $\mathcal{I}_{\epsilon}$ of $\mathbb{X}$.

For different choices of ϵ, we get different covers. The map $g: A \to B$ is identity when $I_l^{\epsilon} \subseteq I_l^{\epsilon'}$ for $\epsilon \leq \epsilon'$.

<u>Example 2</u> Let $\mathbb{X} = [0, 2n] \subseteq \mathbb{R}$, and $\varepsilon > 0$. We consider

$$I_l^{\varepsilon} = (l-\varepsilon, l+1+\varepsilon) \cap \mathbb{X} \quad \text{for } l = 0, 1, \dots, 2n-1, \text{ and}$$

$$J_m^{\varepsilon} = (2m-\varepsilon, 2m+2+\varepsilon) \cap \mathbb{X} \quad \text{for } m = 0, \dots, n-1.$$

$\mathcal{I}_{\varepsilon}$: covering of $\mathbb{X}$ by I_l^{ε}, $l = 0, \dots, 2n-1$.

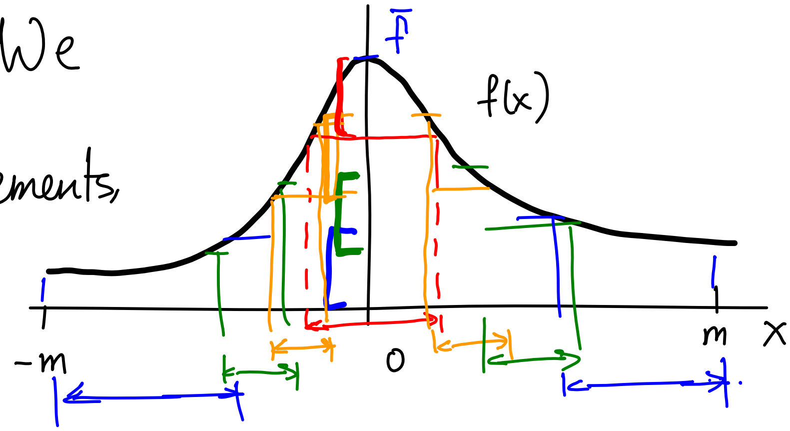
$\mathcal{J}_{\varepsilon'}$: covering of $\mathbb{X}$ by $J_m^{\varepsilon'}$, $m = 0, \dots, n-1$.

Then $g : \{0, \dots, 2n-1\} \longrightarrow \{0, \dots, n-1\}$ is $g(l) = \lfloor \frac{l}{2} \rfloor$. This function induces the map between $\mathcal{I}_{\varepsilon}$ and $\mathcal{J}_{\varepsilon'}$ for $\varepsilon \leq \varepsilon'$.

<u>Example 3</u> We can extend to 2D easily. Consider $\mathbb{X} = [0,n] \times [0,n]$ ($\mathbb{X} \subset \mathbb{R}^2$). A cover similar to the one in Example 1 can use

$$I_{i,j}^{\varepsilon} = (l-\varepsilon, l+1+\varepsilon) \times (j-\varepsilon, j+1+\varepsilon).$$

Example 4 Let $\mathbb{X} = (-m, m) \subseteq \mathbb{R}$ and $f : \mathbb{X} \to \mathbb{R}$ be given as $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$. Then we have $Z = f(x)$ given by $\left[0, \frac{1}{\sigma\sqrt{2\pi}}\right] = [0, \bar{f}]$. We consider the pullback cover elements, i.e., $f^{-1}(\mathcal{U}_\alpha)$ for $\alpha = 1, 2, 3, 4$.



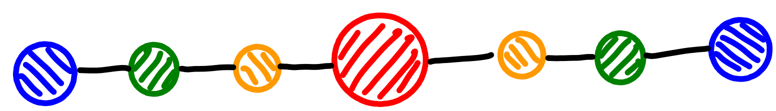
$f^{-1}(\mathcal{U}_1)$ is 1 component.

$f^{-1}(\mathcal{U}_2)$ is 2 components.

$f^{-1}(\mathcal{U}_3)$ is 2 components.

$f^{-1}(\mathcal{U}_4)$ is 2 components.

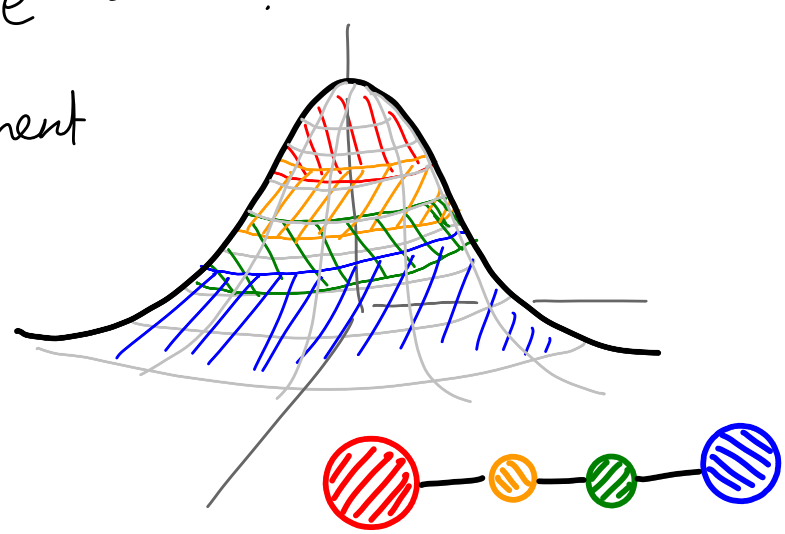
Here is the mapper. Sizes of nodes are proportional to relative sizes of the $f^{-1}(\mathcal{U}_\alpha)$ sets.



Example 5 Let $\mathbb{X} = \mathbb{R}^2$, and we apply the same covering of Z as in Example 4 with

$$f(x,y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x^2+y^2)}{2\sigma^2}}.$$

Here, $f^{-1}(\mathcal{U}_\alpha)$ is a single component for all $\alpha = 1, 2, 3, 4$. The corresponding mapper is shown, and has fewer nodes and edges than in Example 4.



Implementation

Find the range of values of a function of interest, Z.
Cover Z by choosing two parameters:

* interval length ℓ, and
* percentage overlap, p.

We get $\mathcal{C}$, the (set of intervals in the) cover.

<u>Example 6</u>: Let $Z = [0,3]$, $\ell = 1$, $p = 20\%$. We get

$$\mathcal{C} = \{[0,1], [0.8, 1.8], [1.6, 2.6], [2.4, 3]\}.$$

could go to 3.4;

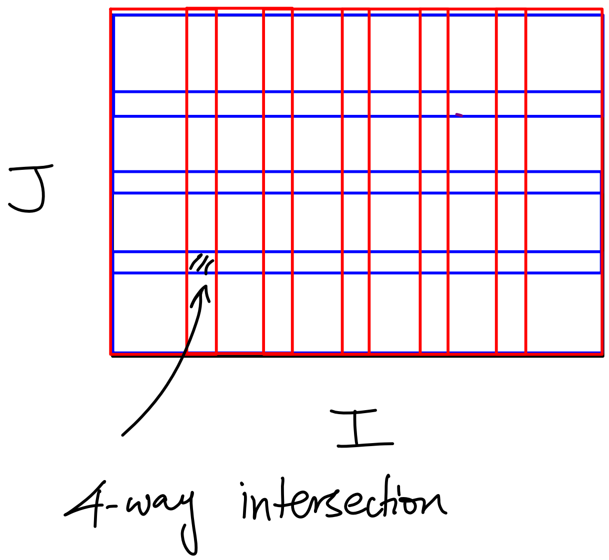
$p = 20\%$ is honored.

If the last interval is set as $[2.4, 3.4]$, then the length choice is also honored. But it is not too critical for the last interval.

For each interval $I_j \in \mathcal{C}$, find $X_j = \{\bar{x} \mid f(\bar{x}) \in I_j\}$, the set of points that form the domain of I_j. The X_j's form a cover of $\mathbb{X}$. For each X_j, find clusters $\{X_{jk}\}$. We could use a subset of original dimensions (even a single one), or all dimensions to do this clustering.

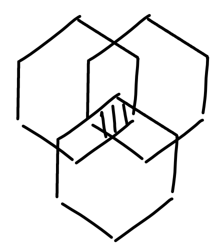
We represent each cluster by a node, and draw an edge between X_{jk} and $X_{\ell m}$ when $X_{jk} \cap X_{\ell m} \neq \emptyset$, a triangle when three clusters have non-empty intersection, and so on.

With two functions, your intervals are now rectangles (by default). But other shapes, e.g., hexagons, could be used for these intervals.



With rectangular intervals, one would get 4-way intersections and even higher order intersections if the overlap is larger. Hence we could get tetrahedra, or even higher dimensional simplices.

But with hexagonal cover elements, and appropriate overlap percentage(s), we could ensure we got at most 3-way intersections. Hence we get at most triangles in the mapper.



Check out the mapper implementations available in scikit-tda and giotto-tda.