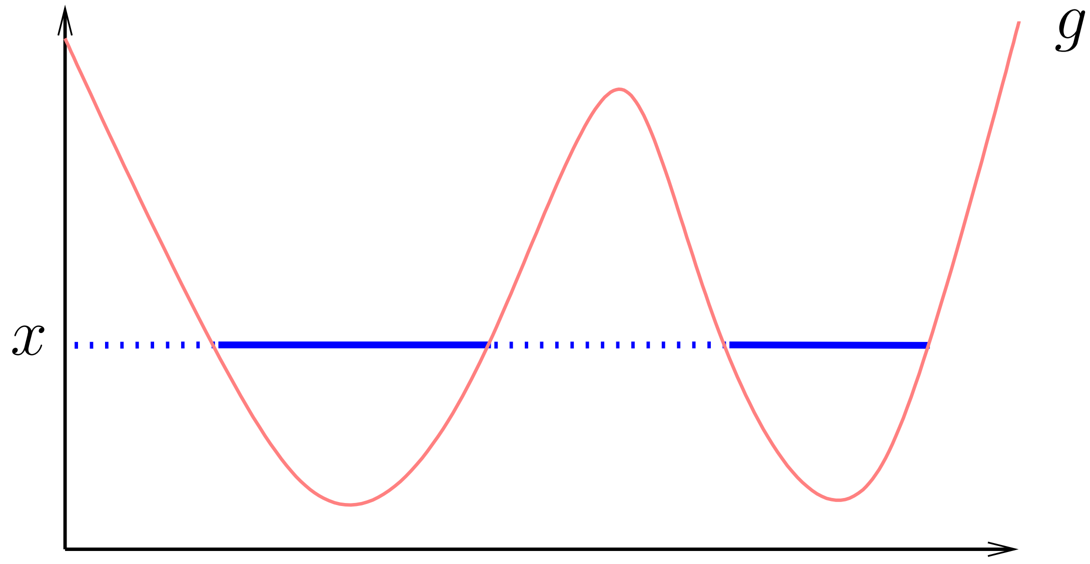
# Persistent homology and geometric inference

David Cohen-Steiner

Workshop on Tensor Valuations in Stochastic Geometry and Imaging

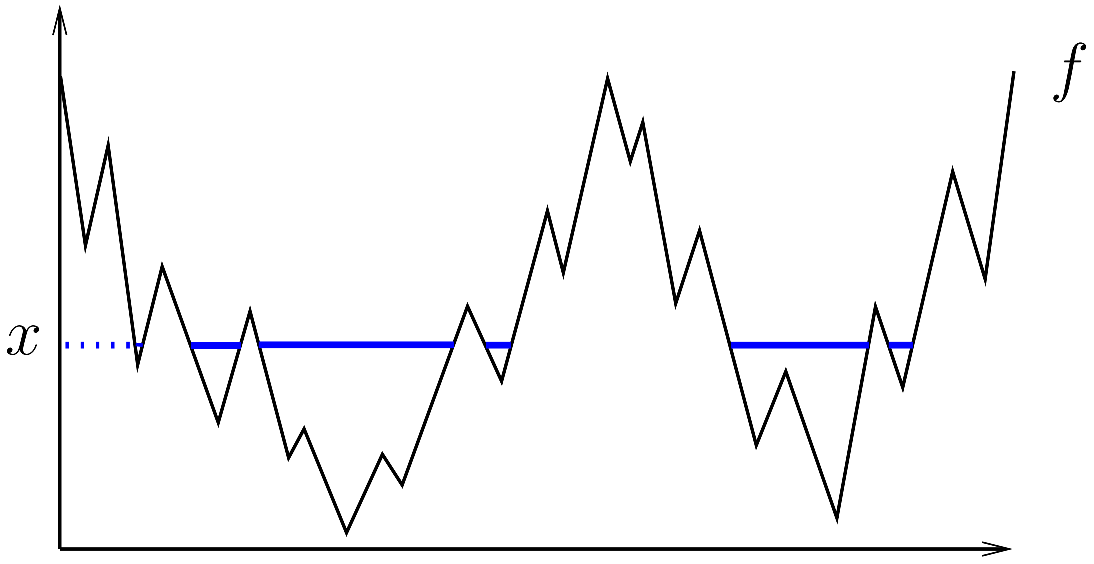
Sandbjerg 2014

# Topological noise



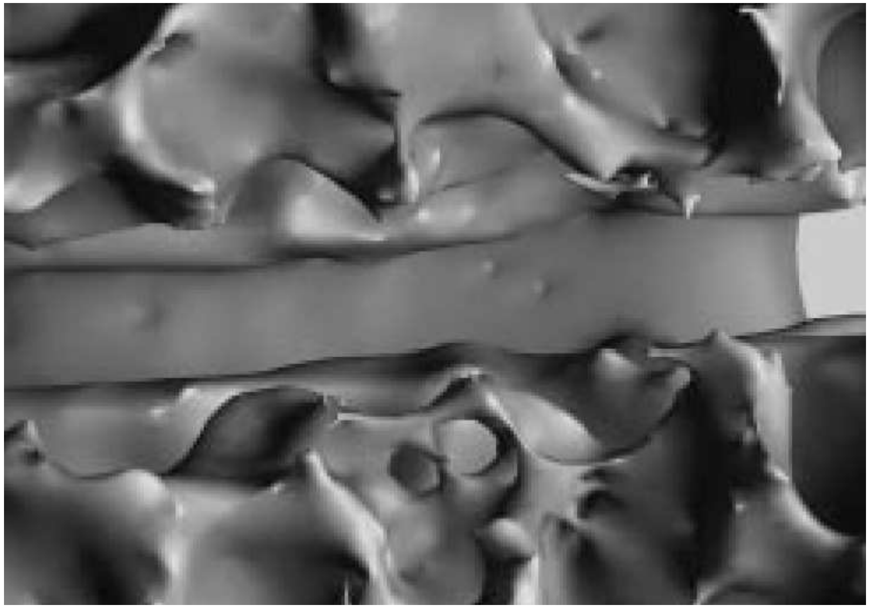

- How many components in $g^{-1}(-\infty, x]$?

# Topological noise




- How many components in $g^{-1}(-\infty, x]$?

# Three-dimensional example

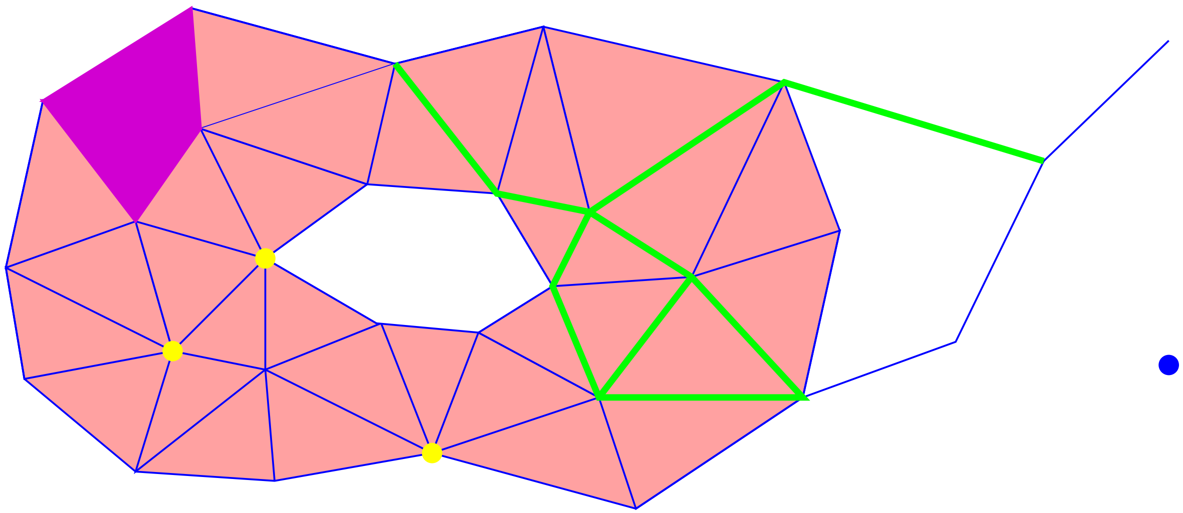


- What is the “actual” topology of this surface?

# Outline

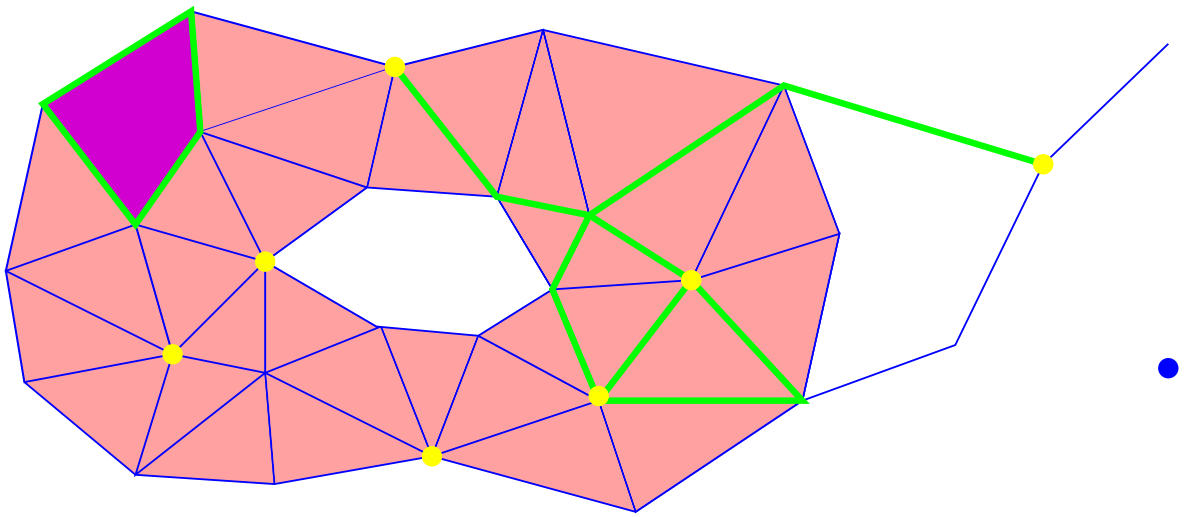
- Homology.
- Persistent homology: definition, structure, algorithm.
- Stability and applications.
- Current work.

# Mod 2 simplicial chains



- $k$-chain of $\mathbb{X}$ = union of $k$-simplices of $\mathbb{X}$ ($k = 0, 1, 2, \ldots$).
- The set of $k$-chains is a $\mathbb{Z}/2$ vector space $C_k(\mathbb{X})$.

# Boundary operator



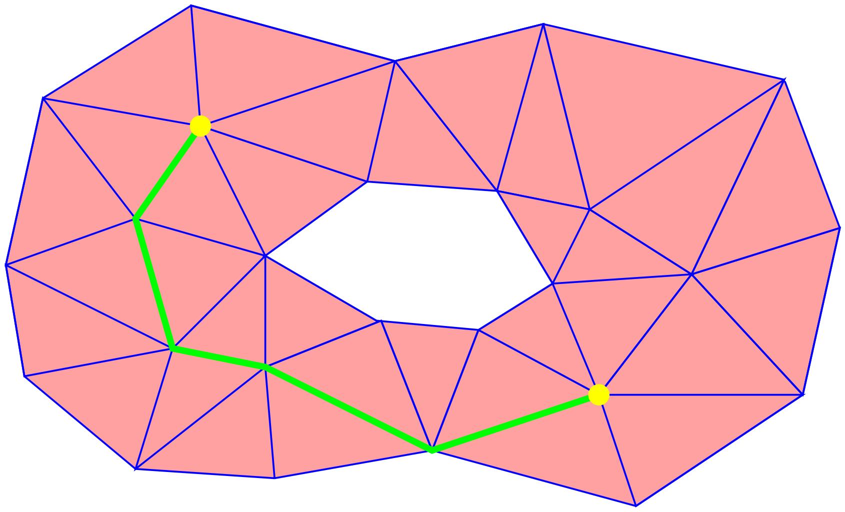
- Let $\partial_k(s)$ be the boundary of a $k$-simplex $s$ in $\mathbb{X}$.
- This defines a linear map $\partial_k : C_k(\mathbb{X}) \to C_{k-1}(\mathbb{X})$.

# Homology groups

$$C_{k+1}(\mathbb{X}) \xrightarrow{\partial_{k+1}} C_k(\mathbb{X}) \xrightarrow{\partial_k} C_{k-1}(\mathbb{X})$$

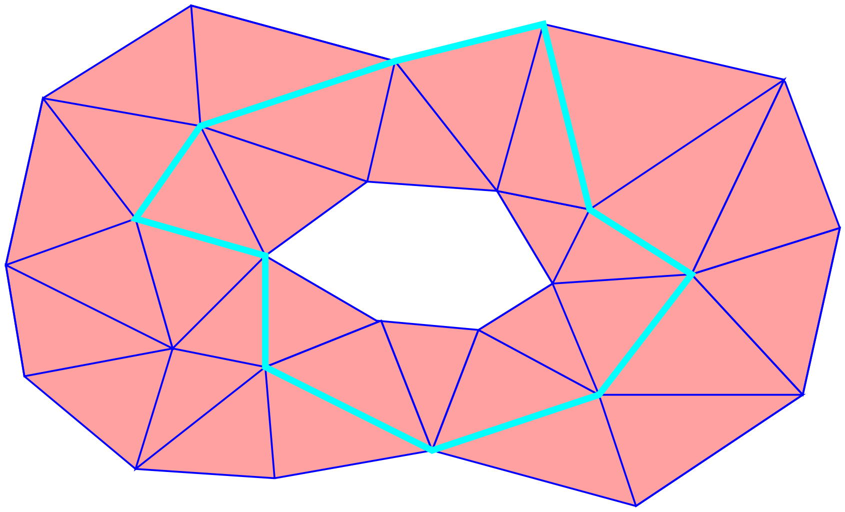
- Cycle space $Z_k(\mathbb{X}) = \ker(\partial_k) \subset C_k(\mathbb{X})$.
- Boundary space $B_k(\mathbb{X}) = \mathrm{im}(\partial_{k+1}) \subset C_k(\mathbb{X})$.
- The boundary of a chain is a cycle : $B_k(\mathbb{X}) \subset Z_k(\mathbb{X})$.
- Let $H_k(\mathbb{X}) = Z_k(\mathbb{X})/B_k(\mathbb{X}) = \ker(\partial_k)/\mathrm{im}(\partial_{k+1})$.

# $H_0(\mathbb{X})$



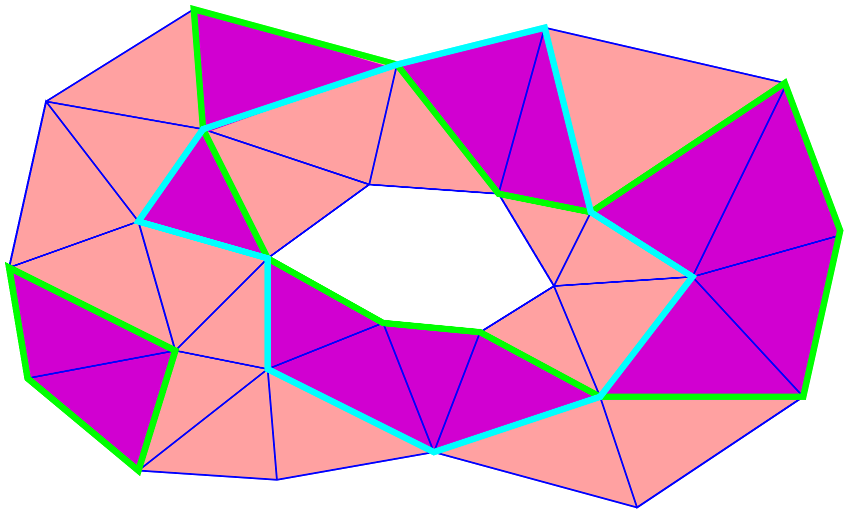
- The dimension $\beta_0(\mathbb{X}) = \dim H_0(\mathbb{X})$ is the number of connected components of $\mathbb{X}$.

# $H_1(\mathbb{X})$



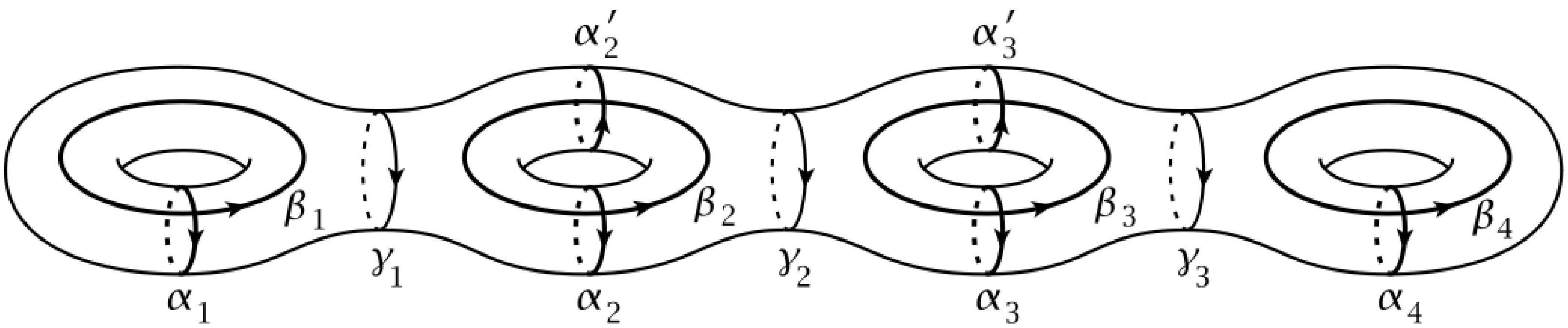
- The dimension $\beta_1(\mathbb{X}) = \dim H_1(\mathbb{X})$ counts the "number of loops" of $\mathbb{X}$.

# $H_1(\mathbb{X})$



- The dimension $\beta_1(\mathbb{X}) = \dim H_1(\mathbb{X})$ counts the "number of loops" of $\mathbb{X}$.

# Closed surfaces




- $\beta_1(\mathbb{X}) = 2$.genus.

picture from Allen Hatcher's algebraic topology book.

# Outline

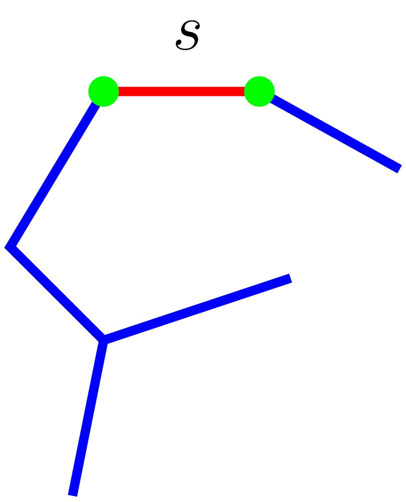
- Homology.
- Persistent homology : definition, structure, algorithm.
- Stability and applications.
- Current work.

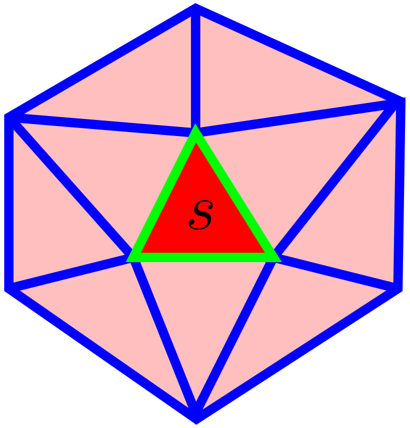
# Filtration

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

- $a_i$ : critical values of $f$.
- $\mathbb{X}^i = f^{-1}(-\infty, a_i]$.
- Filtration of $\mathbb{X}$ : $\emptyset \subset \cdots \subset \mathbb{X}^i \subset \mathbb{X}^{i+1} \subset \cdots \subset \mathbb{X}^n = \mathbb{X}$.
- Simplicial filtration : $\mathbb{X}^i$ is a simplicial complex and $\mathbb{X}^{i+1} = \mathbb{X}^i \cup s$.

# How does $H_*(\mathbb{X}^i)$ change?

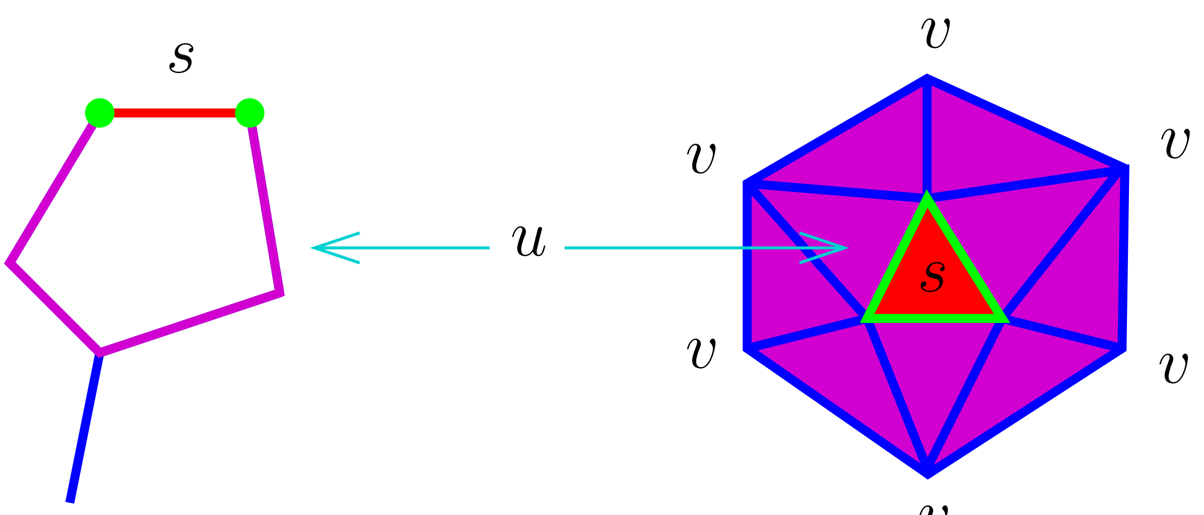



First case : $\partial s \notin B^i_{k-1}$

- $\partial s$ goes into $B^{i+1}_{k-1}$.
- $\partial s \neq 0$ in $H^i_{k-1}$, but $\partial s = 0$ in $H^{i+1}_{k-1}$ : destruction.
- $f^i_{k-1} : H^i_{k-1} \longrightarrow H^{i+1}_{k-1}$ is surjective, with kernel $< \partial s >$.

# How does $H_*(\mathbb{X}^i)$ change?




<u>Second case</u> : $\partial s \in B^i_{k-1}$

- Let $u \in C^i_k$ be such that $ds = du$.
- $\partial(u+s) = 0$, and $u + s \notin B^{i+1}_k$ : creation.
- $f^i_k : H^i_k \longrightarrow H^{i+1}_k$ is injective, $H^{i+1}_k / \mathrm{im}\, f^i_k =< u + s >$.

# Directed system

$$H_k^1 \longrightarrow \ldots \longrightarrow H_k^i \xrightarrow{f_k^i} H_k^{i+1} \longrightarrow \ldots \longrightarrow H_k^n$$

- Describes how the topology of $\mathbb{X}^i$ evolves.
- How can we summarize this information?

# Persistence

[ELZ02]

$$H_k^1 \longrightarrow \ldots \longrightarrow H_k^i \xrightarrow{f_k^i} H_k^{i+1} \longrightarrow \ldots \longrightarrow H_k^n$$

- For $u \in H_k^i$, we define :

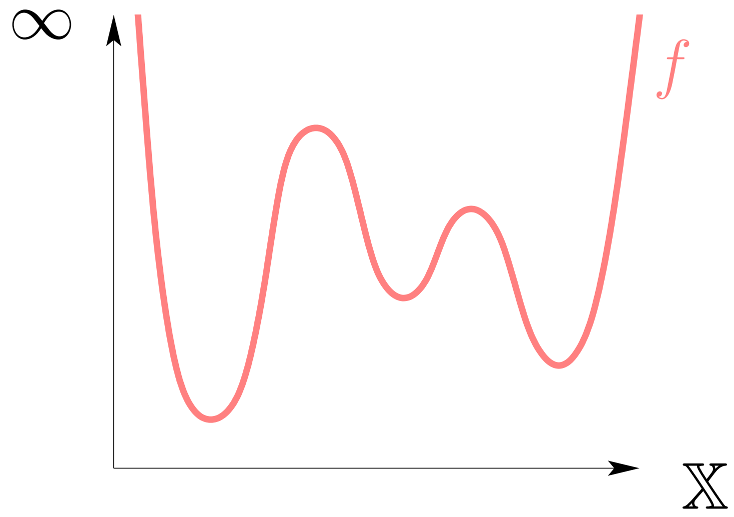
  $$\mathrm{b(u) = \min\{\, j \leq i \,|\, u \in im\,(H_k^j \longrightarrow H_k^i)\}}$$

- For $u \in \ker f_k^i$, we pair $a_{i+1}$ and $a_{\mathrm{b(u)}}$.

  $\rightarrow$ *persistence intervals* $[a_{\mathrm{b(u)}}, a_{i+1}]$

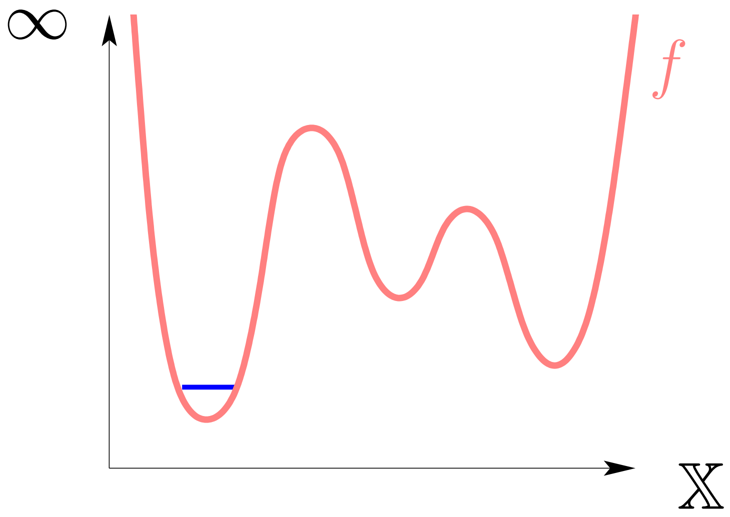
- Each interval represents the life-span of a homology class in the filtration.

# Persistence intervals



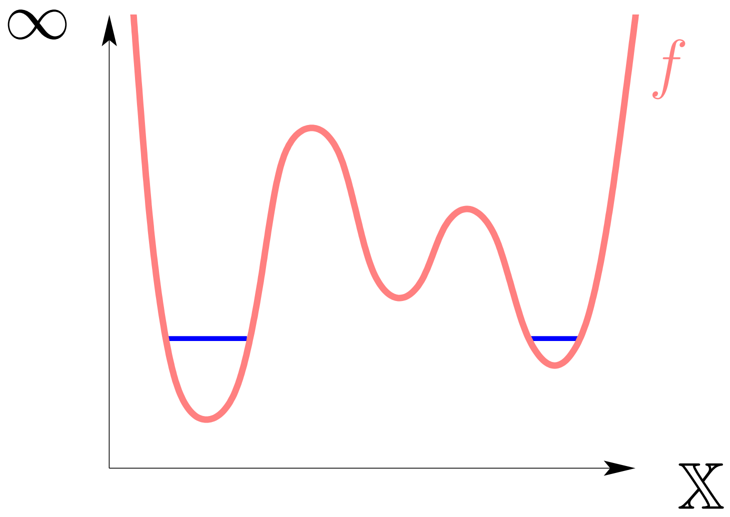

- Track the evolution of the topology of sub-level sets as the threshold increases.
- Pair thresholds that create components with those that destroy them.

# Persistence intervals



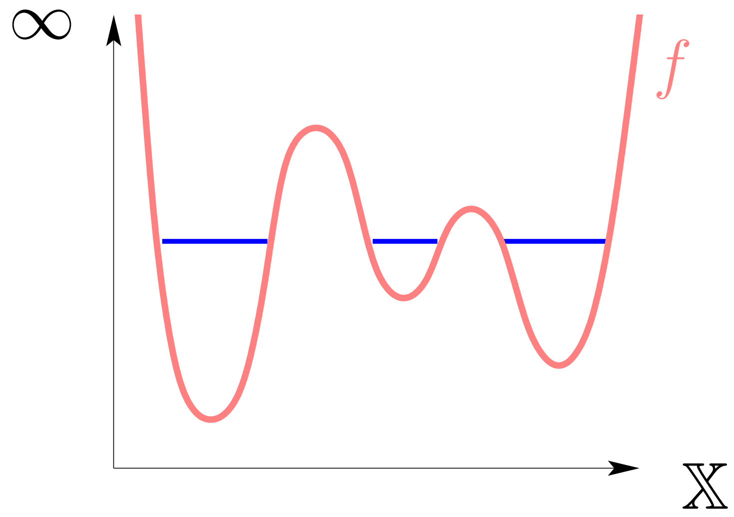

- Track the evolution of the topology of sub-level sets as the threshold increases.
- Pair thresholds that create components with those that destroy them.

# Persistence intervals



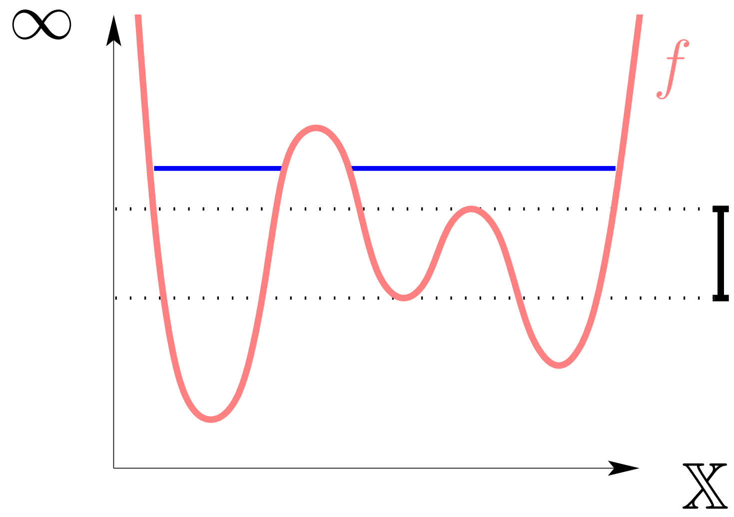

- Track the evolution of the topology of sub-level sets as the threshold increases.
- Pair thresholds that create components with those that destroy them.

# Persistence intervals



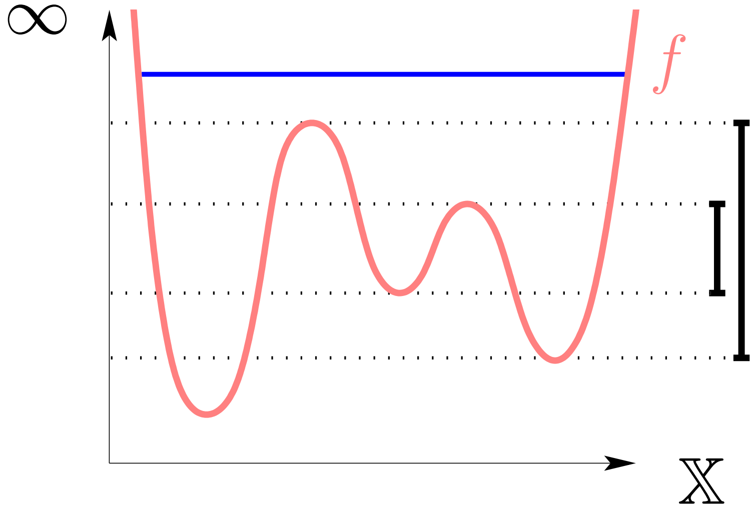

- Track the evolution of the topology of sub-level sets as the threshold increases.
- Pair thresholds that create components with those that destroy them.

# Persistence intervals



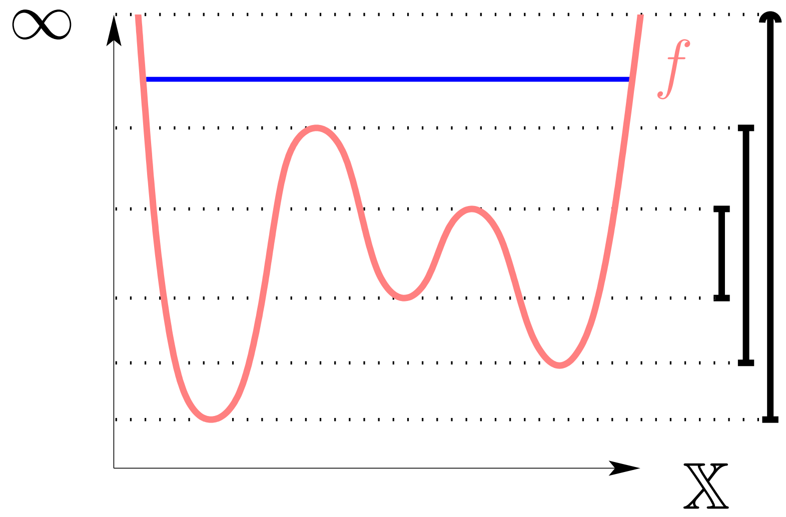

- Track the evolution of the topology of sub-level sets as the threshold increases.
- Pair thresholds that create components with those that destroy them.

# Persistence intervals




- Track the evolution of the topology of sub-level sets as the threshold increases.
- Pair thresholds that create components with those that destroy them.

# Persistence intervals




- Track the evolution of the topology of sub-level sets as the threshold increases.
- Pair thresholds that create components with those that destroy them.

# Incremental step

- $\partial\bar{\sigma}_{i+1}$ was created before $j \leq i$ iff

$$\partial\sigma_{i+1} \in Z(K_j) + B(K_i)$$

- Use the $i$ first columns of $D$ to push the lowest $1$ of $D_{i+1}$ as high as possible.
- Gauss pivoting on the boundary operator

# Algorithm

**Persistence algorithm**

Sort the simplices by increasing function values.

Build the mod 2 incidence matrix.

**while** two columns have their last 1 on the same row

**do**

add the leftmost to the rightmost.

**end while**

**return** $\{(value(s_i), value(s_{last(i)}))\}$

# Quiver representations

- Interval module :

$$0 \longrightarrow \ldots \longrightarrow 0 \longrightarrow \overset{i}{\mathbb{Z}/2} \longrightarrow \ldots \longrightarrow \overset{j}{\mathbb{Z}/2} \longrightarrow 0 \longrightarrow \ldots \longrightarrow 0$$

- A directed system of vector spaces can be written uniquely as the direct sum of interval modules.
- Allows to define persistence intervals in the non simplicial case.

# Persistent Betti numbers

- Define

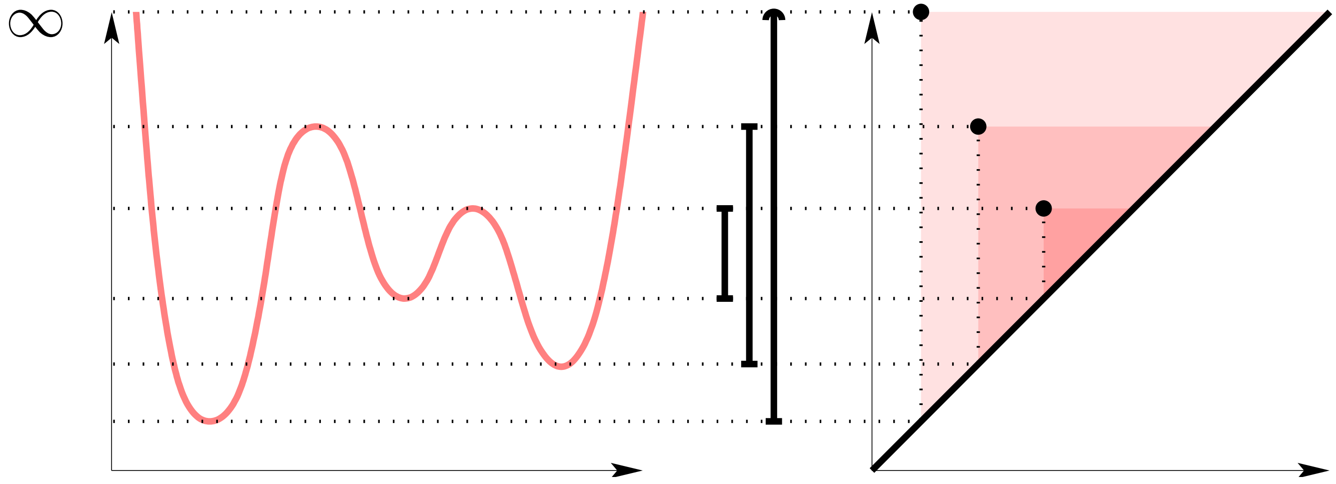
$$\beta_k^{i,j}(f) = \mathsf{rk}(H_k(K_i) \to H_k(K_j))$$

- $\beta_k^{i,j}(f)$ is the number of persistence intervals that contain $[i, j]$ (*"k-triangle lemma"*).
- Persistence intervals are given by $-\frac{\partial^2 \beta_k^{i,j}(f)}{\partial_i \partial_j}$.

# Outline

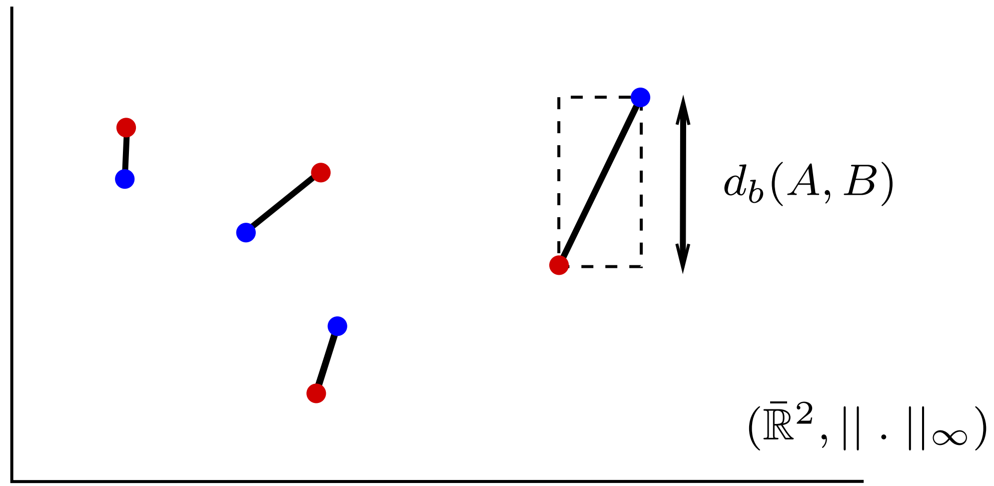
- Homology.
- Persistent homology : definition, structure, algorithm.
- Stability and applications.
- Current work.

# Persistence diagrams




- Persistence intervals become points in the plane.
- The diagonal is included.

# Metric on diagrams

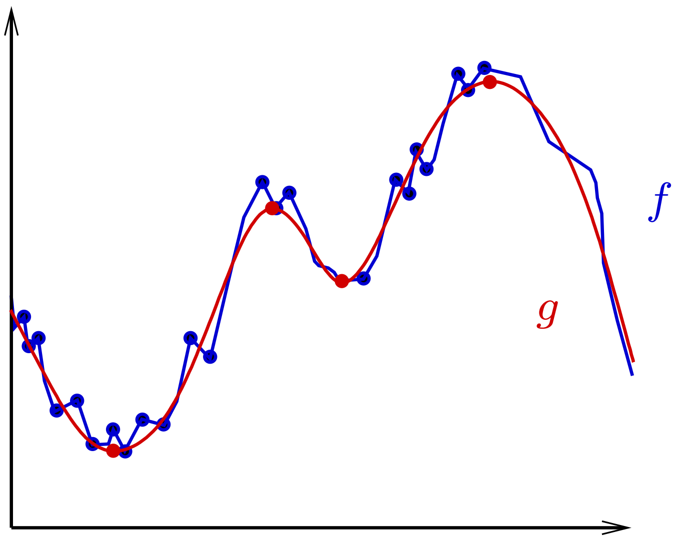


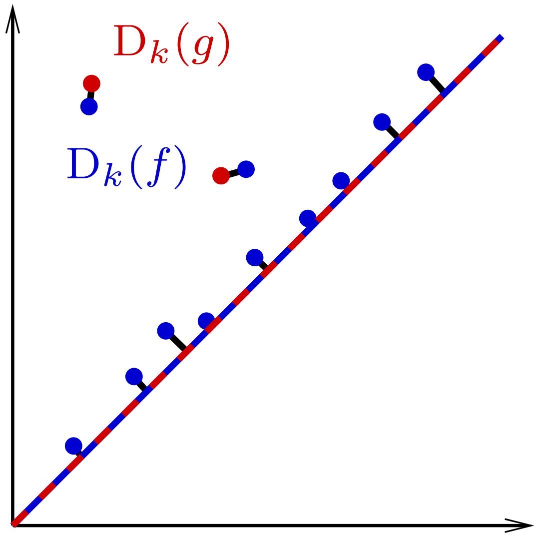

**Definition.** The *bottleneck distance* between multisets $A$ and $B$ is:

$$d_b(A, B) = \inf_{\gamma} \sup_{a} \|a - \gamma(a)\|_\infty$$

over all $a \in A$ and all bijections $\gamma : A \to B$.

# Stability

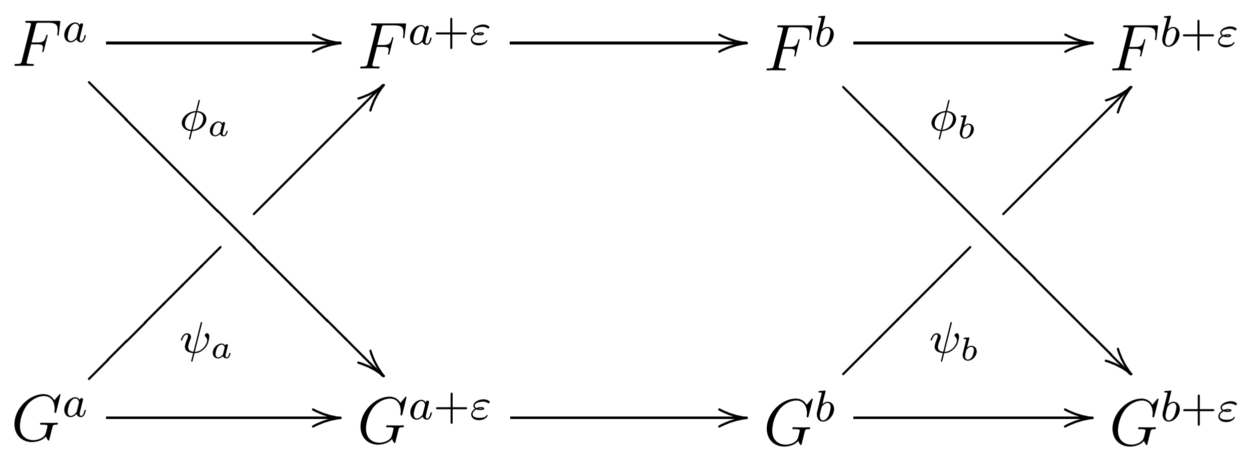



**Theorem.** *[CSEH04].For two* tame *functions $f$ and $g$ on a finitely triangulable space:*

$$d_b(\mathrm{D}_k(f), \mathrm{D}_k(g)) \leq \|f - g\|_\infty$$

# Interleaving

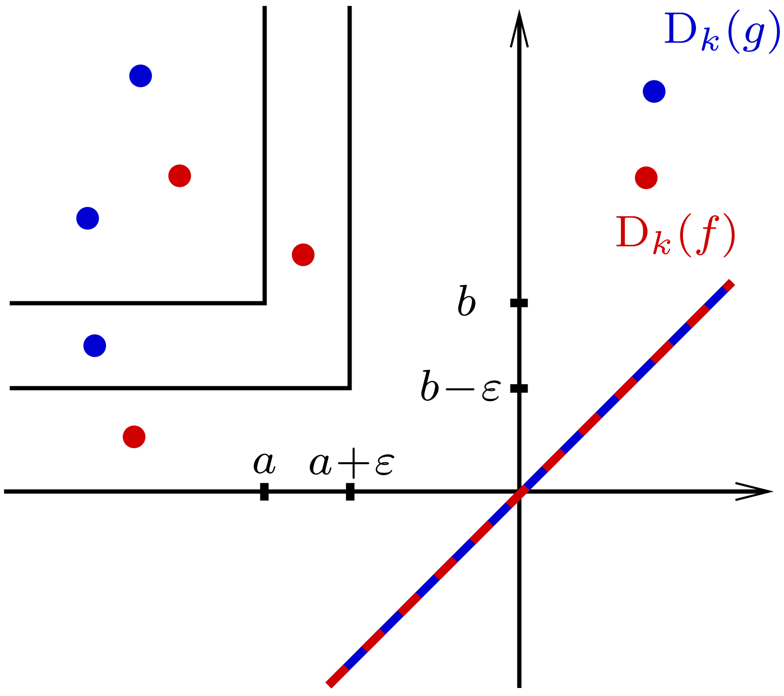



where:

- $\varepsilon \geq \|f - g\|_\infty$
- $F^x = H_k(f^{-1}(-\infty, x])$
- $G^x = H_k(g^{-1}(-\infty, x])$

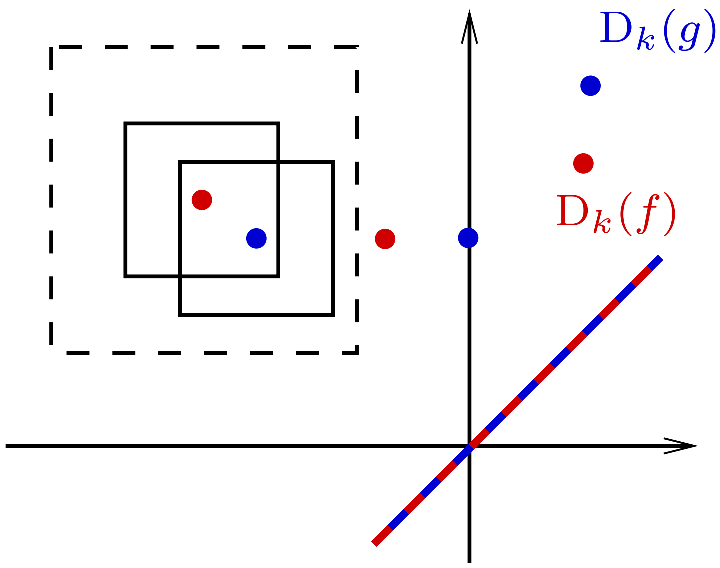
# Quadrant Lemma

$$\begin{array}{ccc} F^{a-\varepsilon} & \longrightarrow & F^{b+\varepsilon} \\ \phi_{a-\varepsilon} \downarrow & & \uparrow \psi_b \\ G^a & \longrightarrow & G^b \end{array}$$



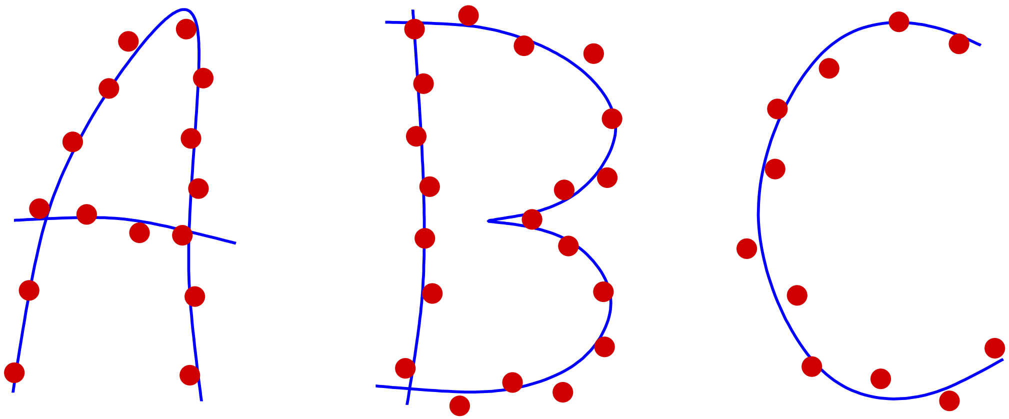

- $\beta_k^{a,b}(g) \geq \beta_k^{a-\varepsilon,b+\varepsilon}(f)$

# Sketch of proof



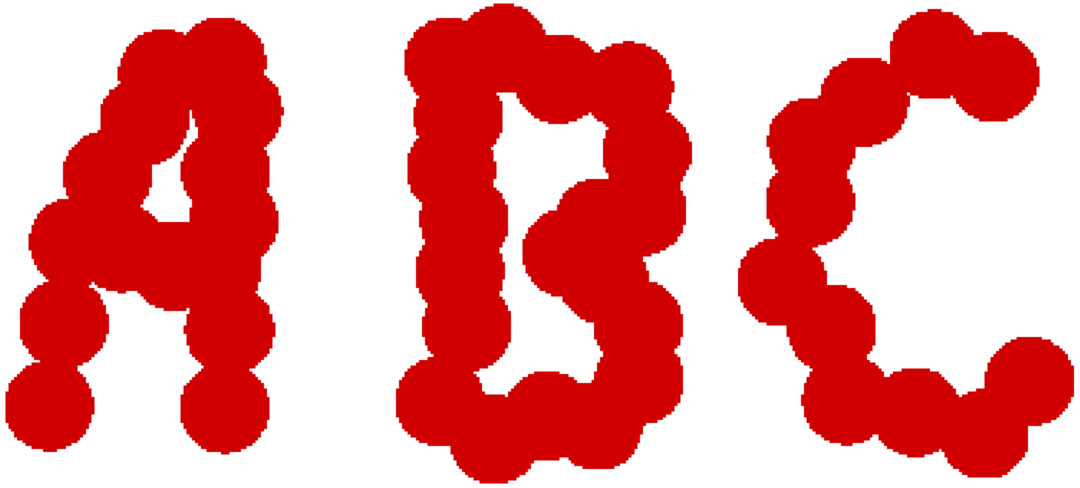

- The quadrant lemma extends to boxes.
- Bound on Hausdorff distance: $d_H(\mathrm{D}_k(f), \mathrm{D}_k(g)) \leq \|f - g\|_\infty$
- Bound on bottleneck distance for sufficiently close functions
  *i.e.* $\|f - g\|_\infty < \frac{1}{2}$ (minimum distance between two points of $\mathrm{D}_k(f)$))
- Linearly interpolating between $f$ and $g$ concludes the proof.

# Betti numbers from samples



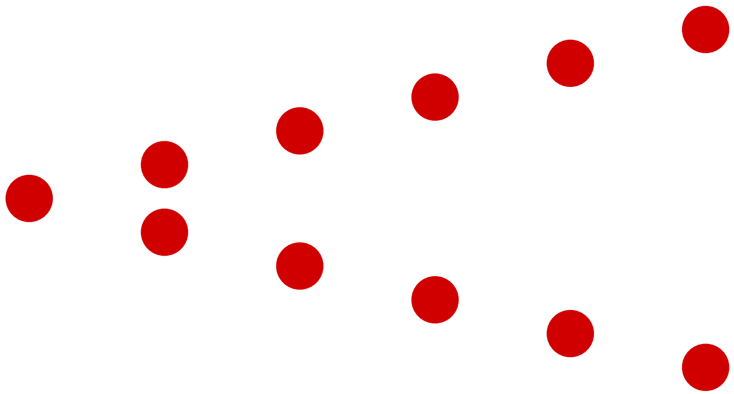
- Build a simplicial approximation of the unknown shape and compute its Betti numbers.
- Use offsets/alpha-shapes.

# Reconstruction by offset

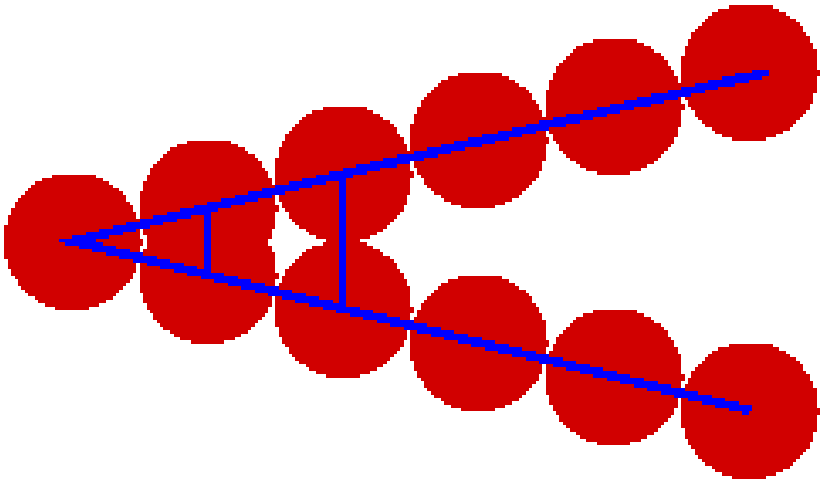


- Works for a large class of shapes in $\mathbb{R}^n$ [CCSL06].
- But might require many data points.

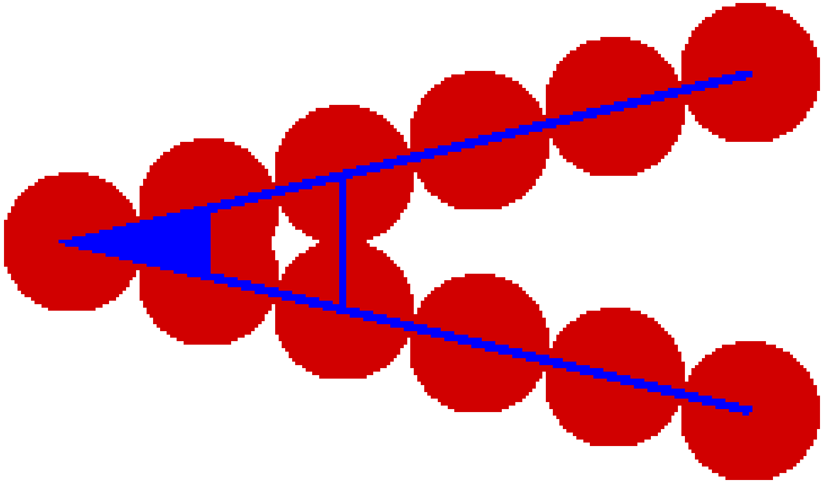
# “Sharp” angle problem



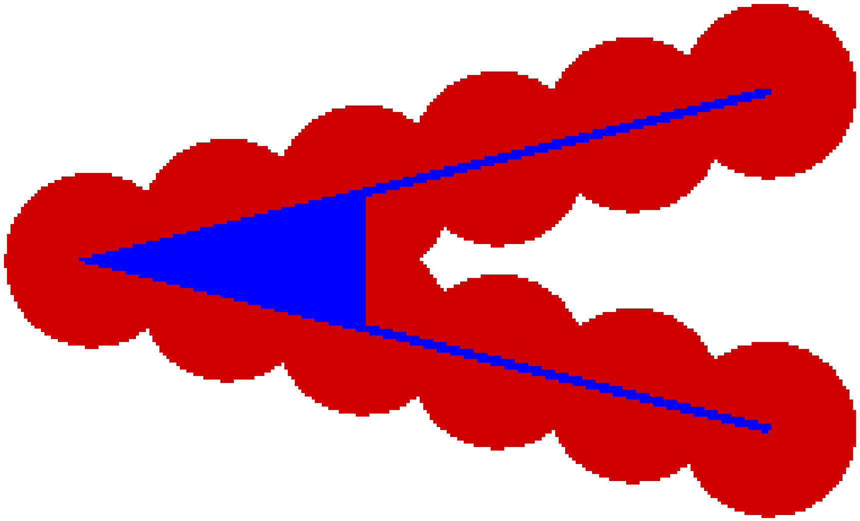
# “Sharp” angle problem



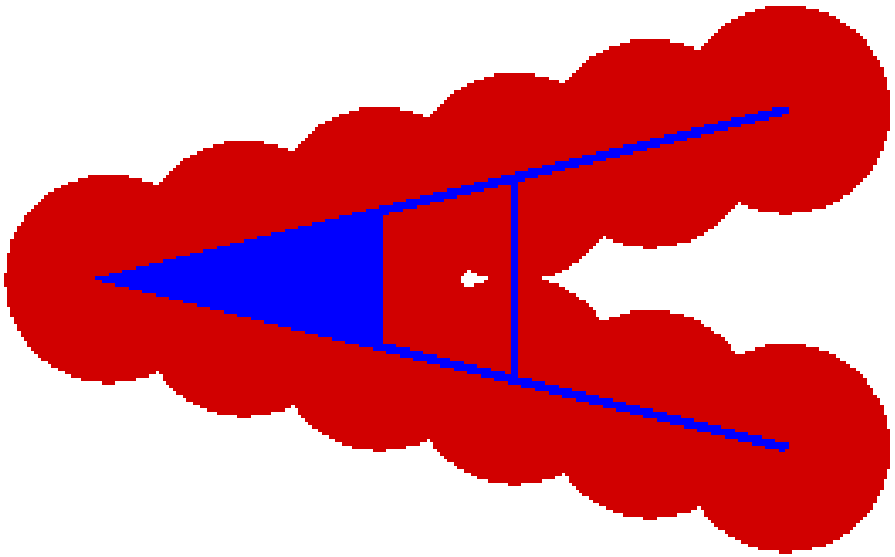
# “Sharp” angle problem



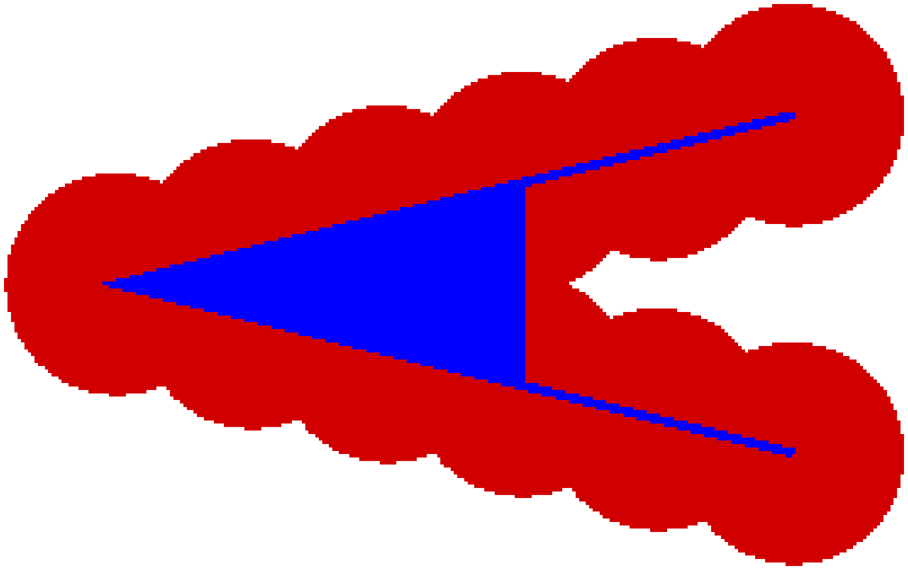
# "Sharp" angle problem



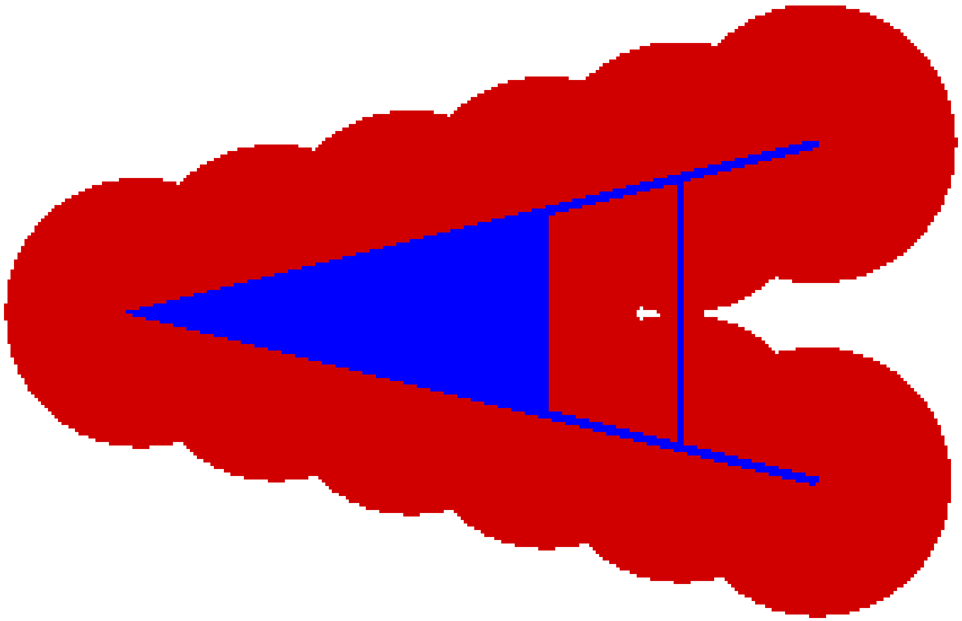
# “Sharp” angle problem



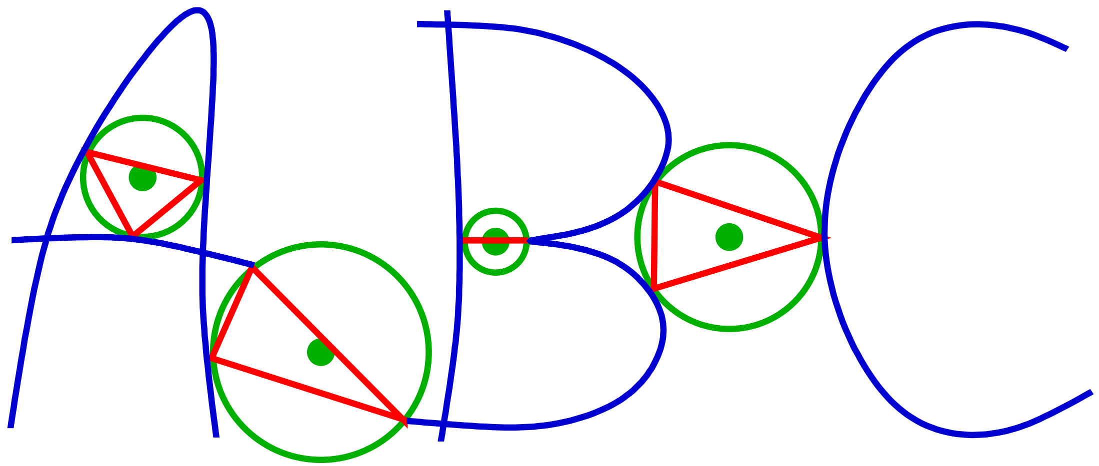
# “Sharp” angle problem



# "Sharp" angle problem

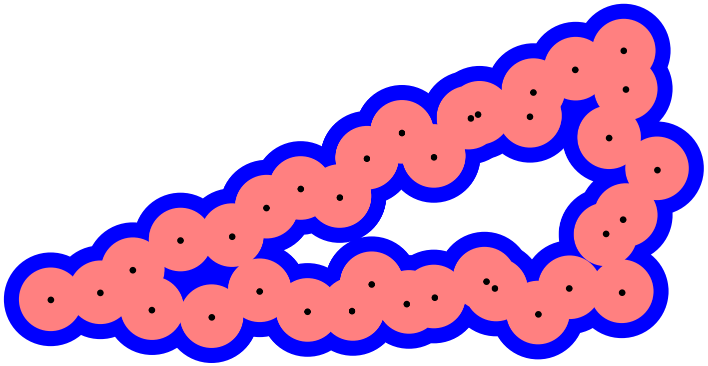


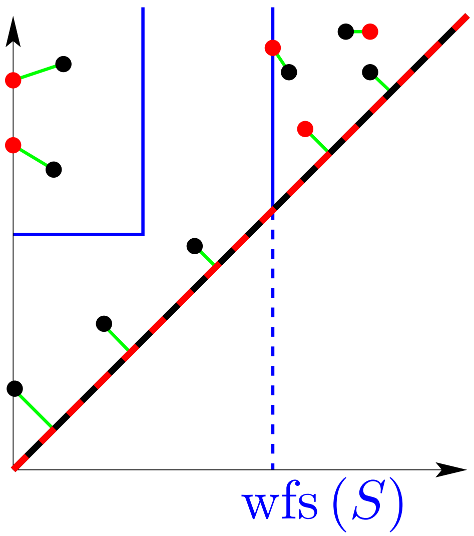
# Weak feature size



- $\mathrm{wfs}\,(S) = $ inf $\{$ positive critical value of dist$_S \}$
- $\mathrm{wfs}\,(S) > 0$ if $S \subset \mathbb{R}^n$ is subanalytic [Fu95].

# Betti numbers from samples

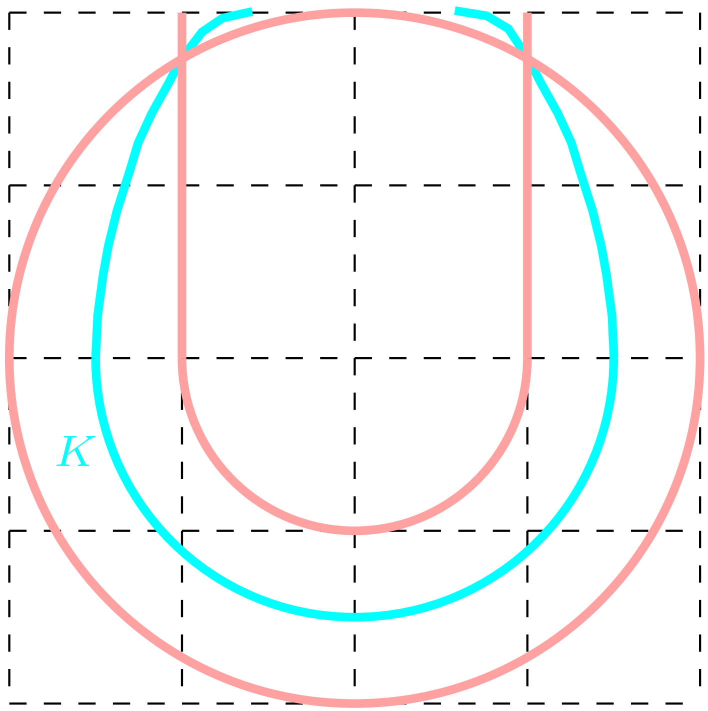





**Theorem.** [CSEH/CL04]. Let $S$ and $P$ be closed subsets of $\mathbb{R}^n$. If $l$ is such that $d_H(S, P) < l < \mathrm{wfs}\,(S)/4$:

$$\beta_k(S) = \beta_k^{l,3l}(\mathsf{dist}_P)$$

# Optimality

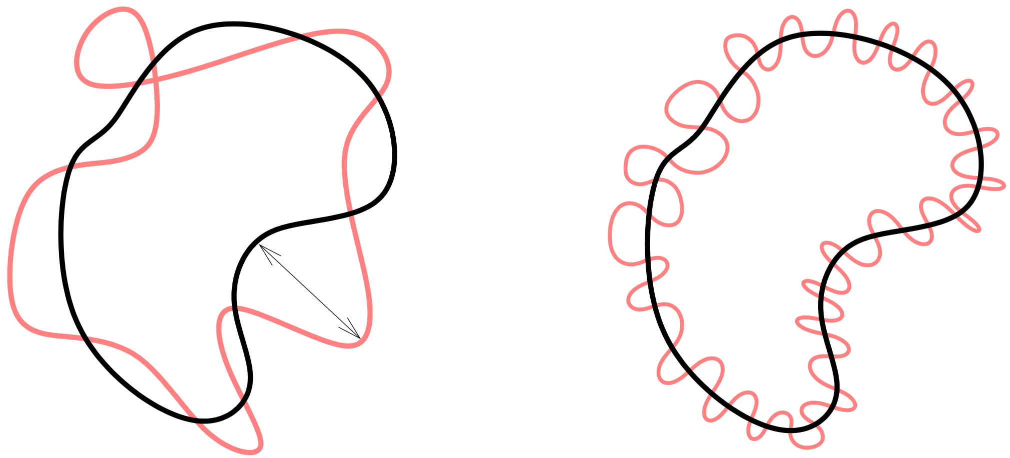



- $\text{wfs}\,(O) = 2$, $\text{wfs}\,(U) = +\infty$
- $d_H(K, O) = d_H(K, U) = 1/2$

# Comments

- Persistent Betti numbers easily computable from the Delaunay triangulation of the sample points.
- You do not get any simplicial complex with the correct Betti numbers.
- Case of high dimensional ambiant space: witness complexes [CdS03]

# Problem for curves

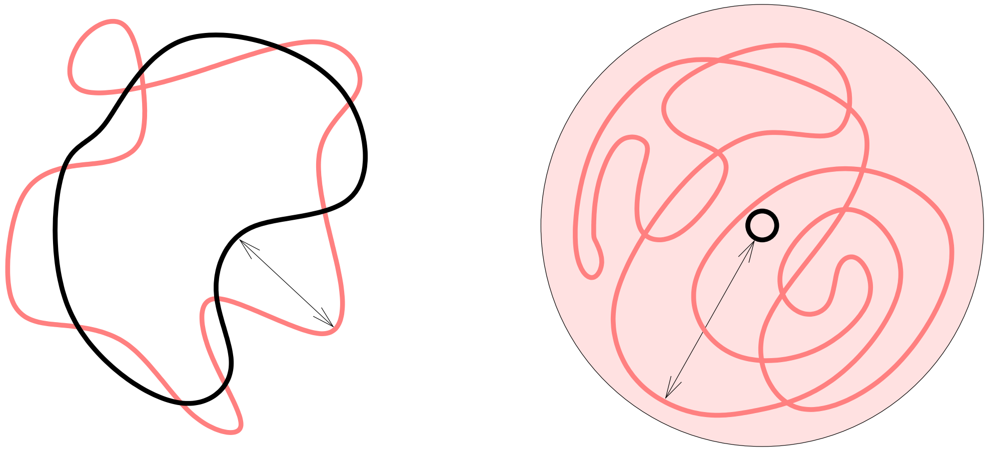


- If two curves are close, does it imply that their lengths are close?
- Fréchet distance between $C_1$ and $C_2$:

$$d_F(C_1, C_2) = \inf_{\phi_1, \phi_2} \sup_s d(\phi_1(s), \phi_2(s))$$

where $\phi_i$ ranges over all parameterizations of $C_i$.

# Result



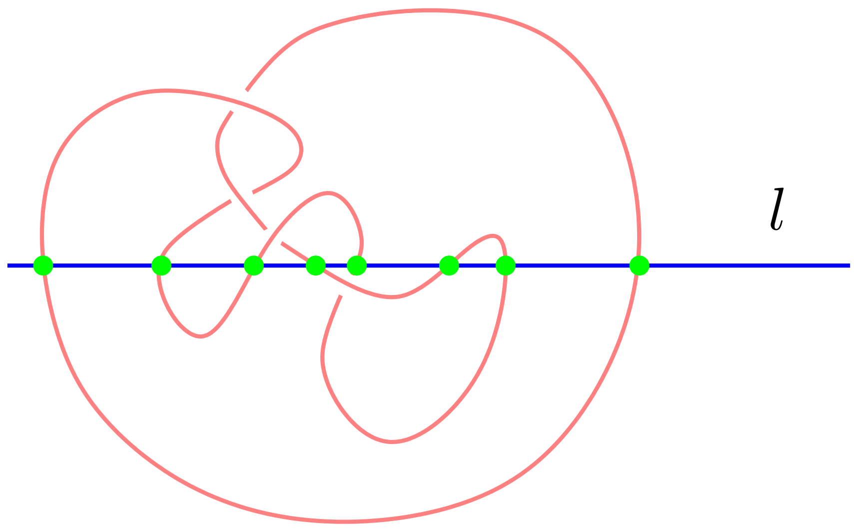
**Theorem.** *Let $C_1$ and $C_2$ be two closed curves in $\mathbb{R}^n$.*

*Let $L_i$ be the length of $C_i$, and $K_i$ be the integral of its curvature.*

*One has:*

$$|L_1 - L_2| \leq \frac{2\mathrm{vol}(\mathbb{S}^{n-1})}{\mathrm{vol}(\mathbb{S}^n)}[K_1 + K_2 - 2\pi]\, d_F(C_1, C_2)$$

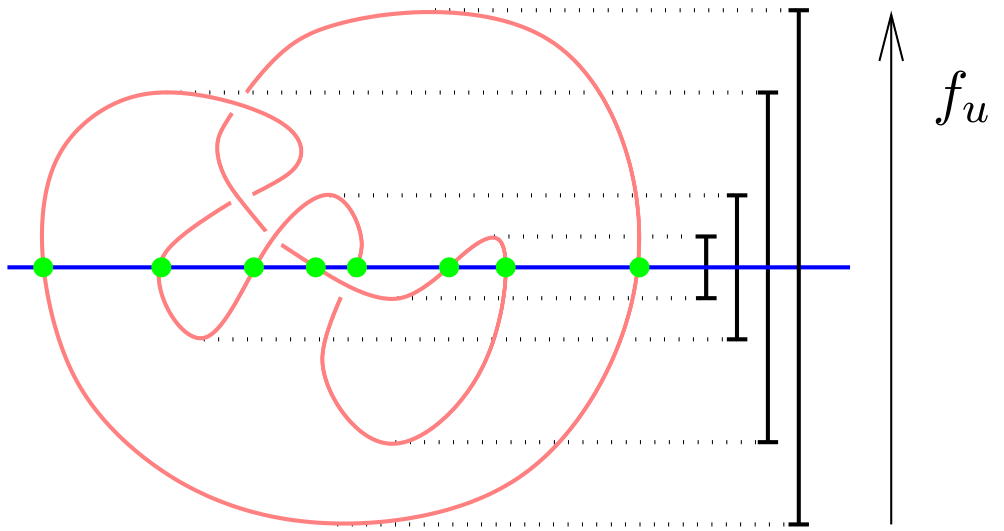
# Proof




- Crofton formula:

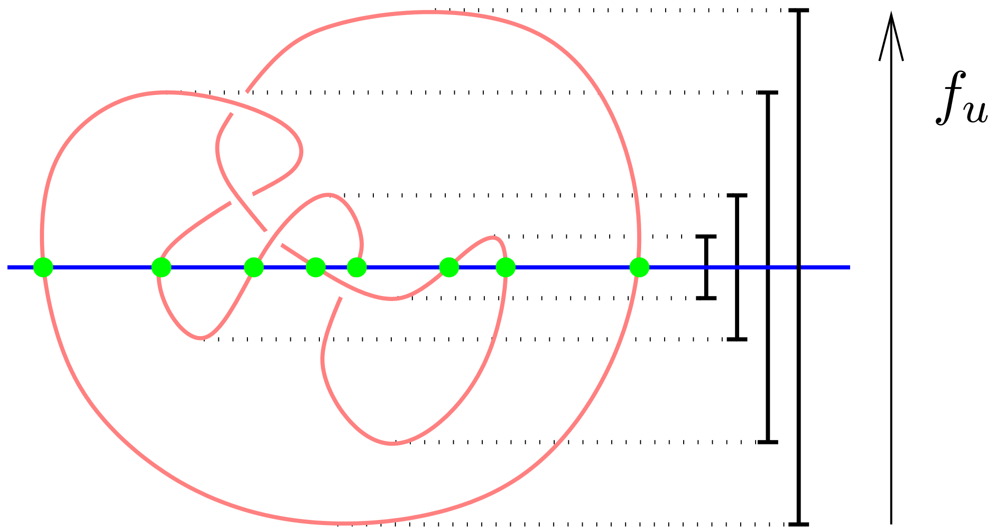
$$L(C) = \frac{\pi}{\mathrm{vol}(\mathbb{S}^n)} \int_{\text{hyperplane } l \subset \mathbb{R}^n} \sharp(l \cap C)$$

# Proof




- Let $f^u : C \to \mathbb{R}$ be the height function in the direction $u$.
- If $l$ has normal vector $u$, then $\sharp(l \cap C)$ is twice the number of "persistence intervals" of $f_u$ stabbed by $l$.

# Proof

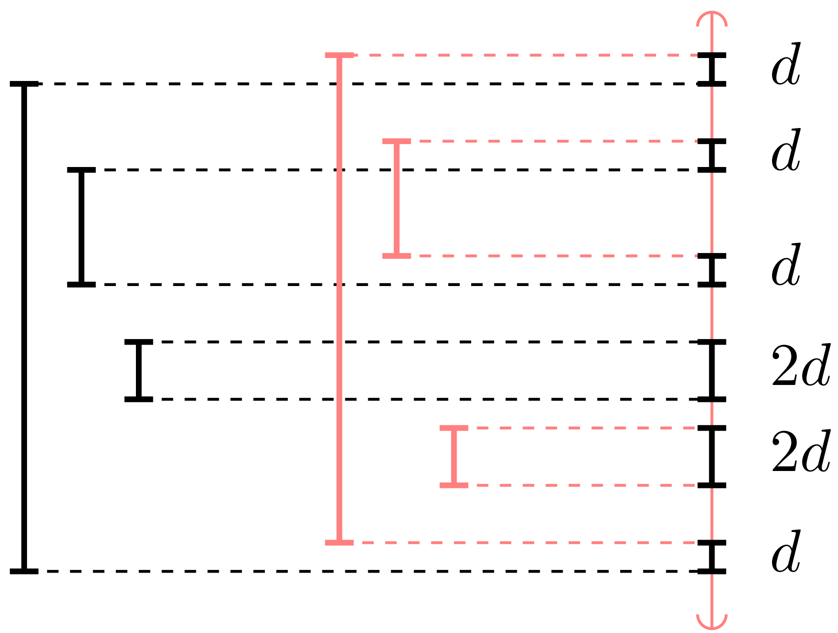



- Hence :

$$\int_{l \text{ hyperplane with normal } u} \sharp(l \cap C)$$

is twice the total length of the persistence intervals of $f^u$.

# Proof




- Stability theorem : the bounds of the persistence intervals of $f^u$ move by at most $d_F(C_1, C_2) = d$.
- Hence the total length of these intervals changes by at most $d(n_1^u + n_2^u - 2)$, where $n^u$ is the number of critical points of $f^u$.

# Proof

- By integrating over all directions :

$$|L_1 - L_2| \leq 2d \, \frac{\pi}{\mathrm{vol}(\mathbb{S}^n)} \int_{u \in \mathbb{S}^{n-1}} n_1^u + n_2^u - 2 \, du$$

- The integral of the number of critical points $n_i^u$ over $u \in \mathbb{S}^{n-1}$ is the integral of the curvature of $C_i$ divided par $\pi/\mathrm{vol}(\mathbb{S}^{n-1})$

# Surfaces

**Theorem.** *Let $S_1 = \partial V_1$ and $S_2 = \partial V_2$ be two closed surfaces in $\mathbb{R}^3$ with the same genus $g$. Let $H_i$ be the total mean curvature of $S_i$, and $K_i$ be its total absolute Gauss curvature. One has:*

$$|H_1 - H_2| \leq [K_1 + K_2 - 4\pi(1+g)]\, d_F(V_1, V_2)$$

- Holds for piecewise-linear surfaces, for which simple formula exist: accurate total mean curvature estimation from a mesh.
- Closeness between normals to the surfaces is not required.

# Outline

- Homology.
- Persistent homology : definition, structure, algorithm.
- Stability and applications.
- Current work.