Topological Inference via Meshing

Don Sheehy
Theory Lunch

Joint work with
Benoit Hudson, Gary Miller, and Steve Oudot

Computer Scientists want to know the **shape** of data.

Clustering

Principal Component Analysis

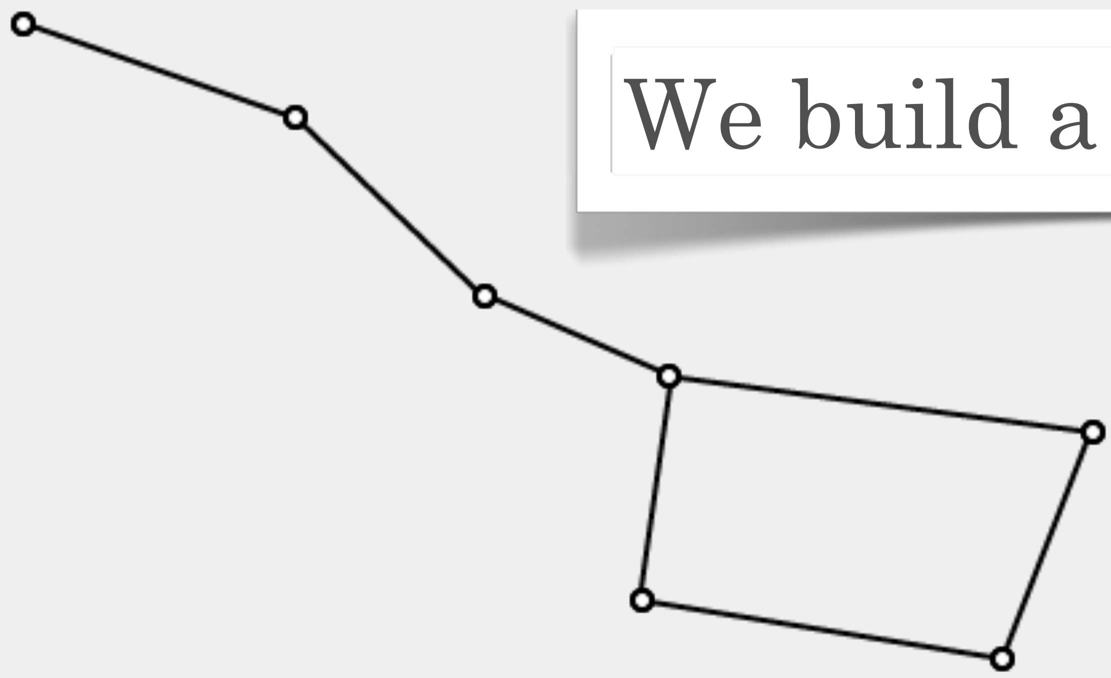
Convex Hull

Mesh Generation

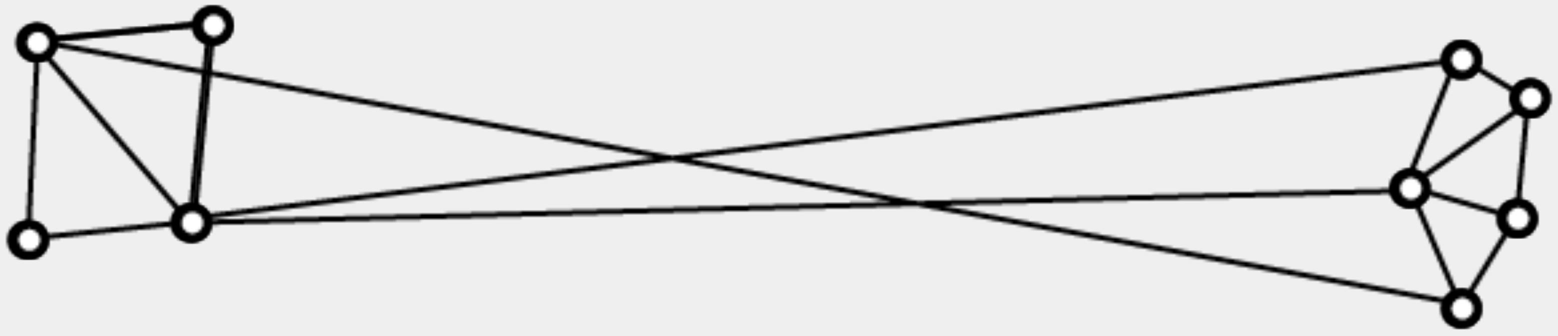
Surface Reconstruction

Point sets **have no shape**...

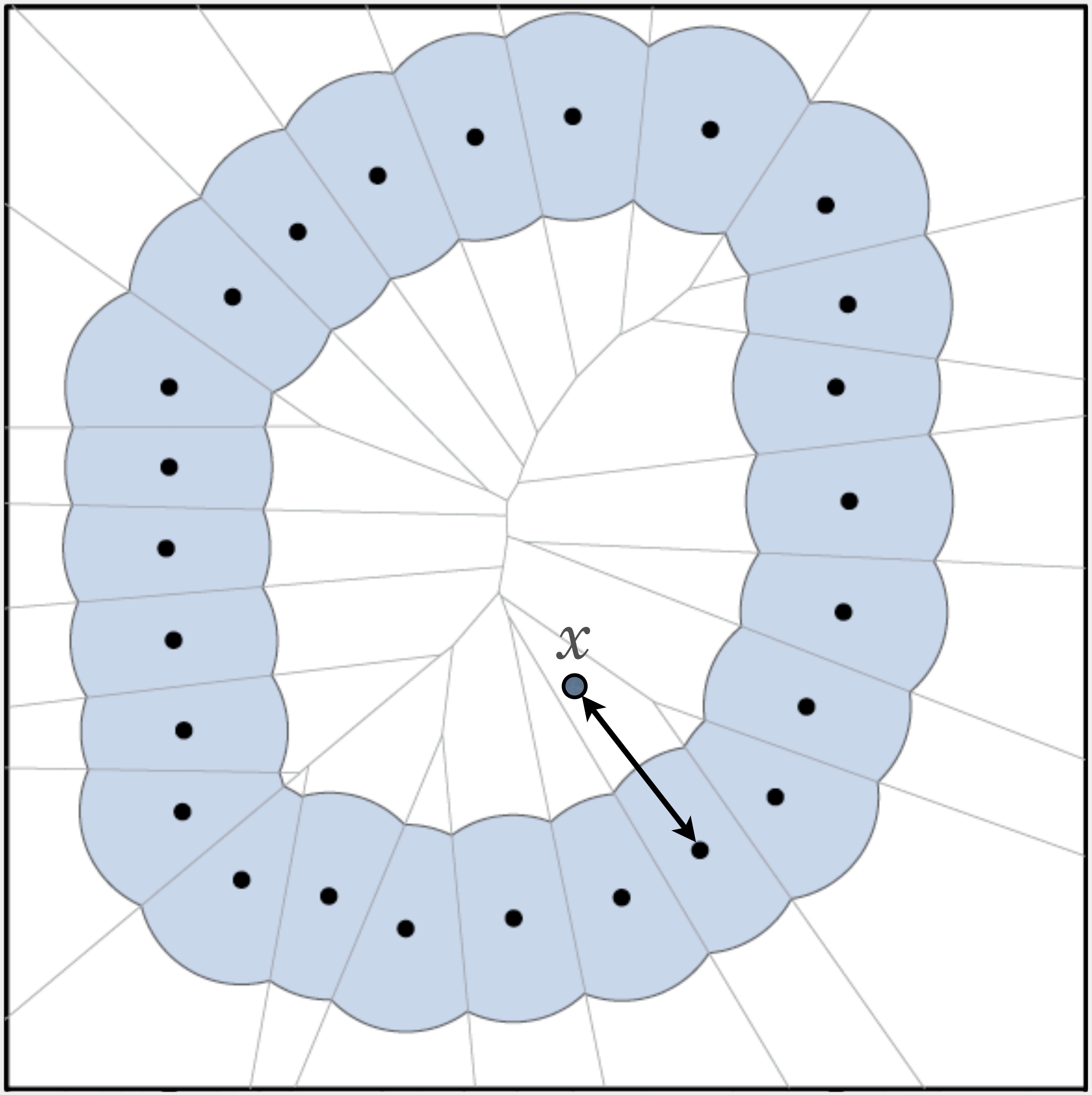
so we have to add it ourselves.



We assume that the **geometry** is **meaningful**.



Distance functions add **shape** to **data**.

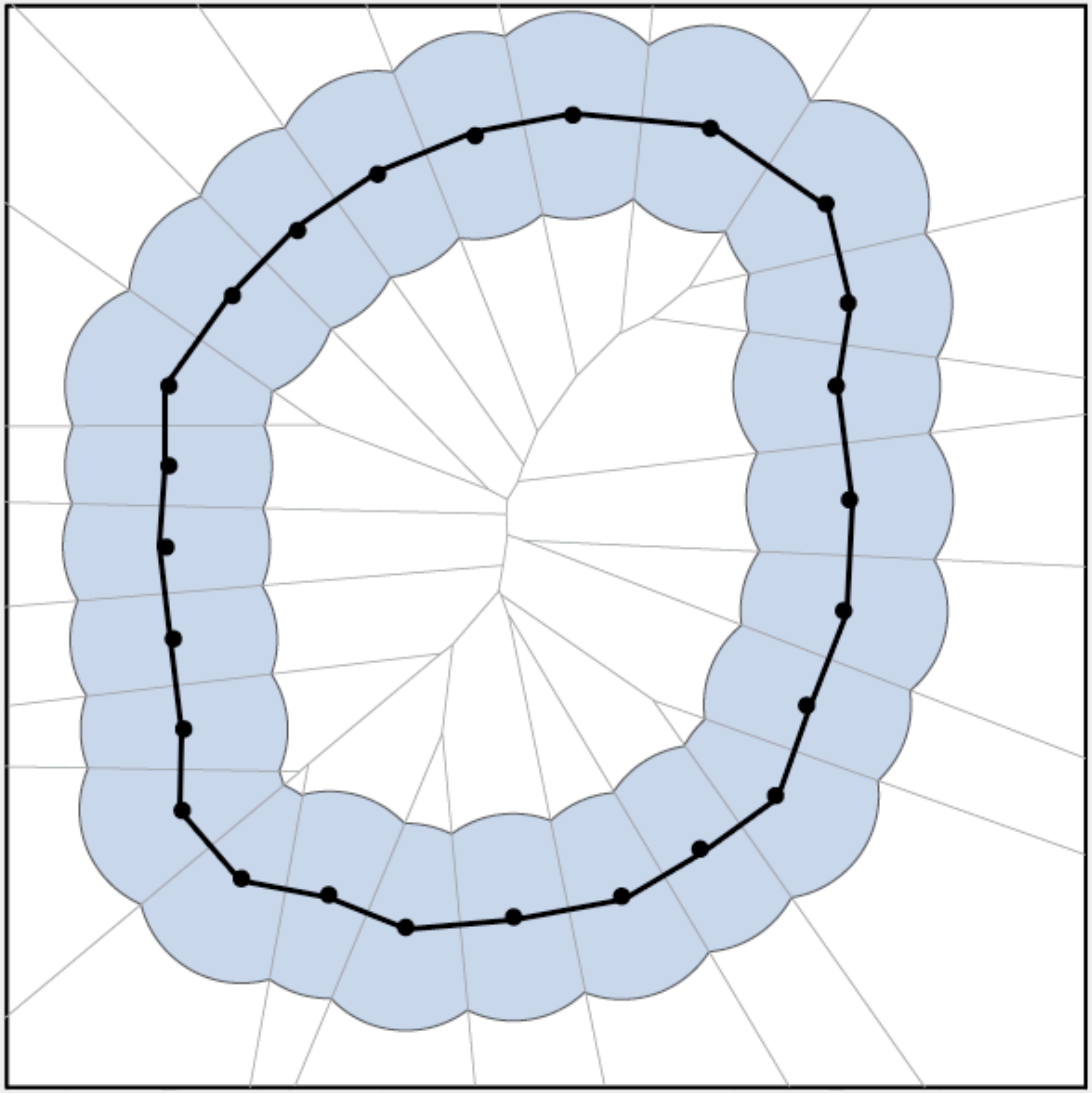


$$d_P(x) = \min_{p \in P} |x - p|$$

$$P^\alpha = d_P^{-1}[0, \alpha]$$

$$= \bigcup_{p \in P} \mathrm{ball}(p, \alpha)$$

If you know the "**scale**" of the data, the situation is often easier.



There may not be a **right** scale at which to look at the data for two reasons.

1. No **single** scale captures the shape.
2. Interesting features appear at **several** scales.

The Persistence Approach:

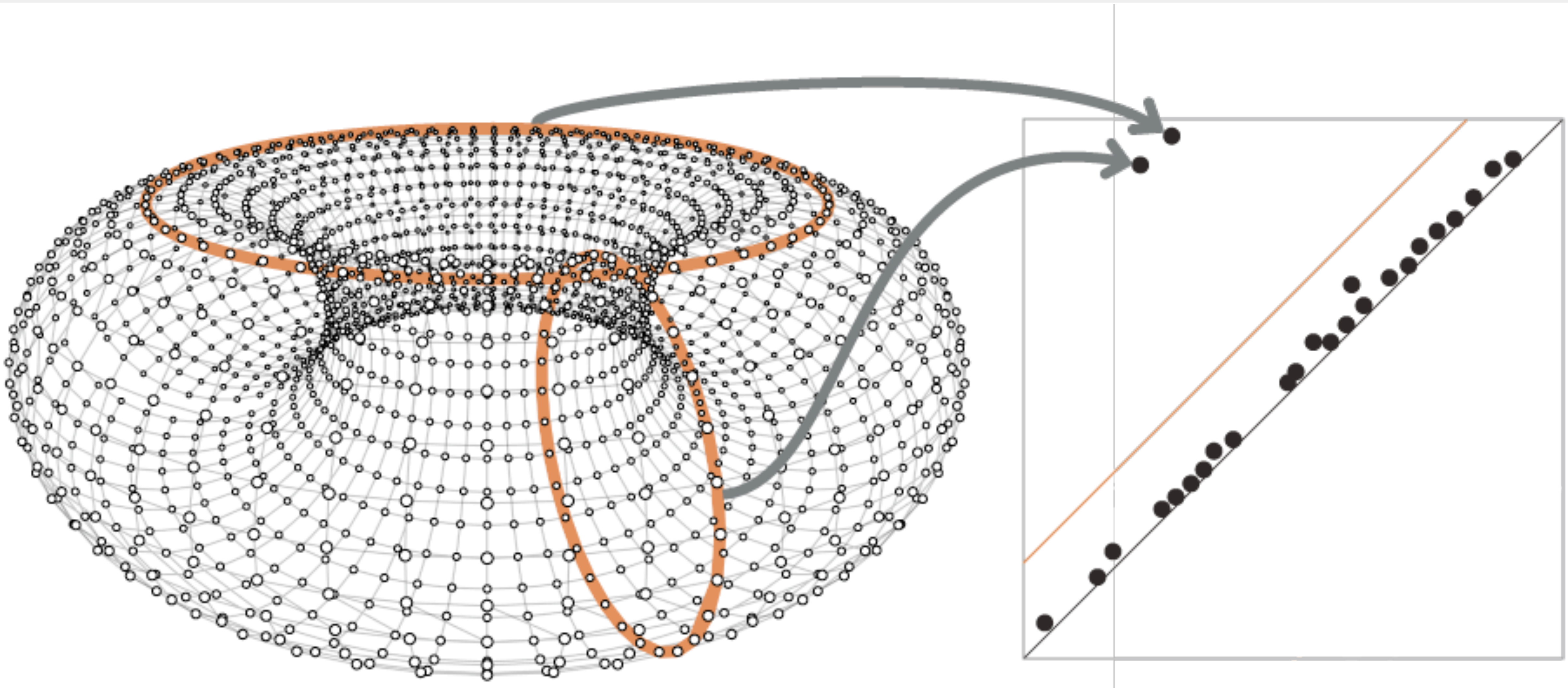
Look at *every* scale and then see what *persists*.

Homology gives a quantitative description of qualitative aspects of shape, i.e. components, holes, voids.

Reduces to linear algebra for simplicial complexes.

0th Homology group is the nullspace of the Laplacian.

Persistent Homology tracks homology across **changes in scale**.



The input to the persistence algorithms is a ***filtration***.

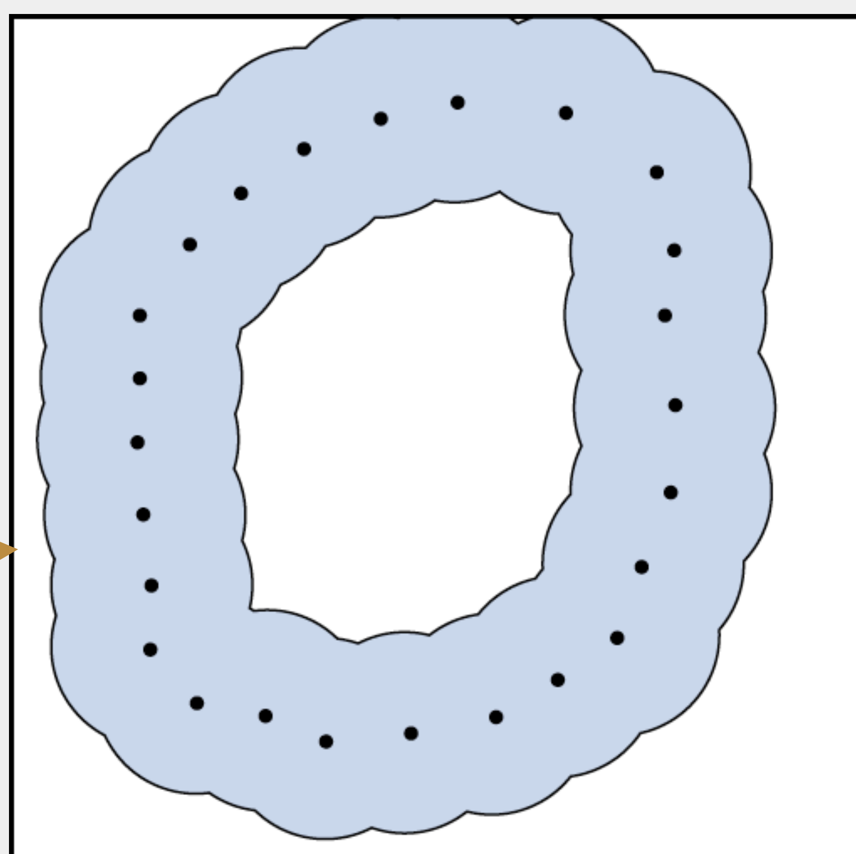
A ***filtration*** is a growing sequence of topological spaces.

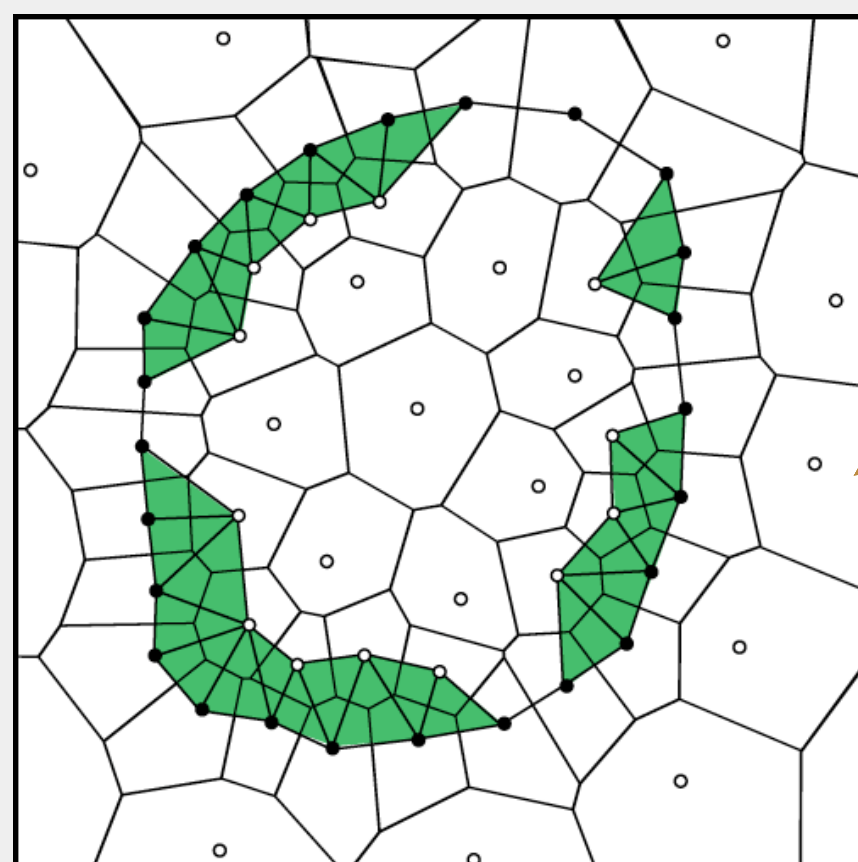
$$\{F_\alpha\}_{\alpha \geq 0} \qquad \forall \alpha \leq \beta, F_\alpha \subseteq F_\beta$$

We care about two different types of filtrations.

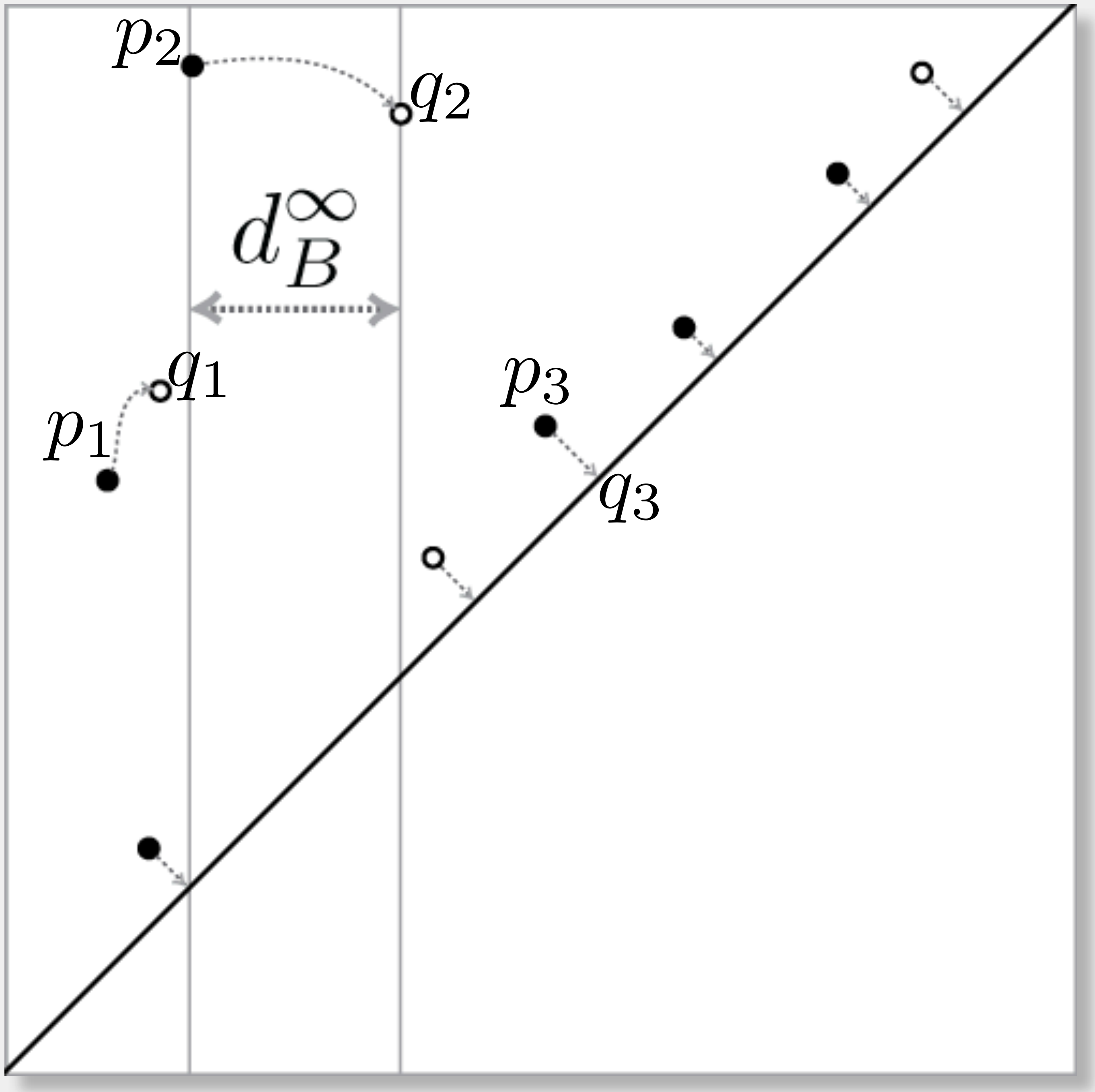
1 Sublevel sets of well-behaved functions.

2 Filtered simplicial complexes.



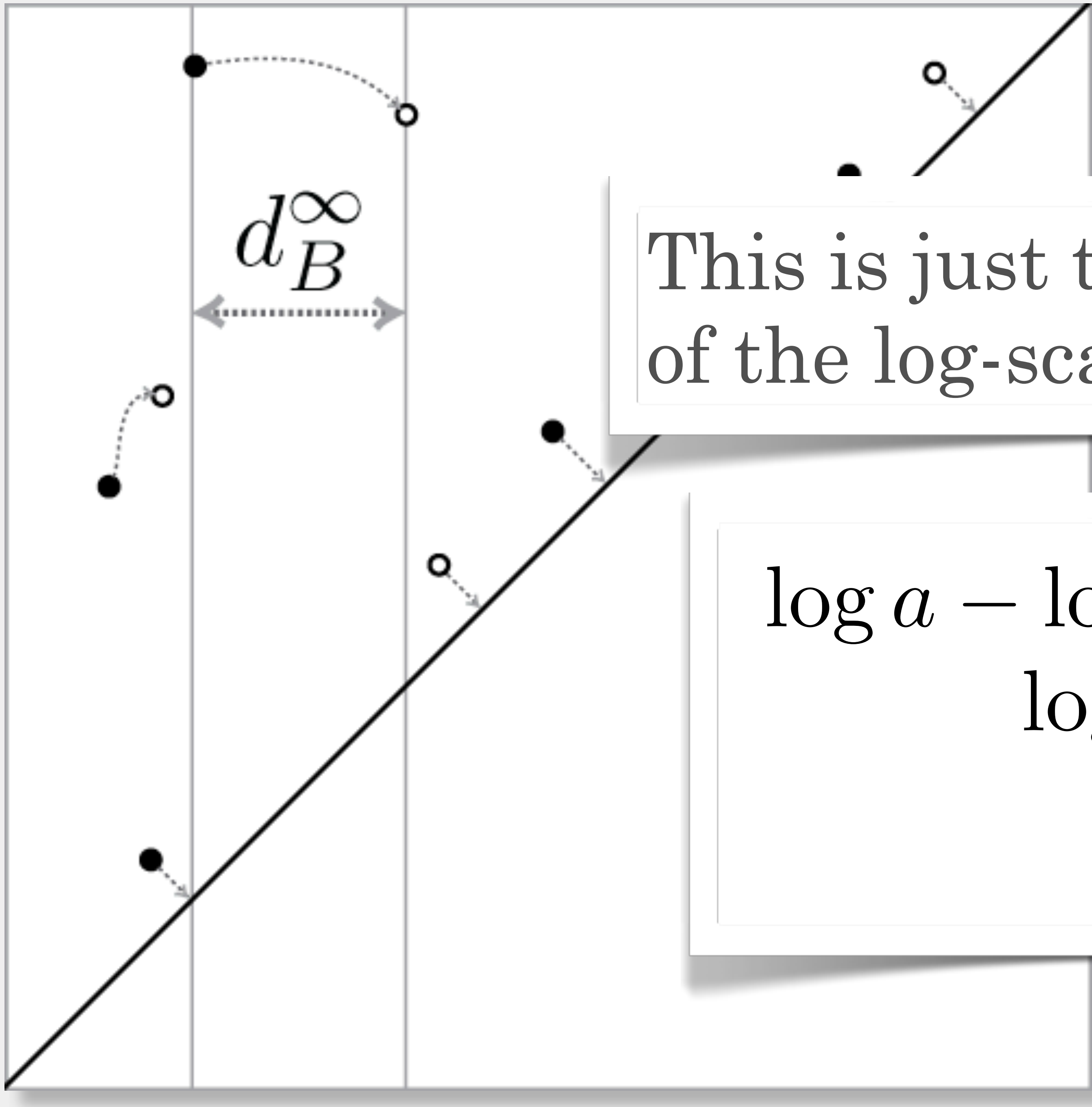


The distance between two diagrams is the bottleneck of a matching.



$$d_B^\infty = \max_i |p_i - q_i|_\infty$$

Approximate persistence diagrams have features that are born and die within a constant factor of the birth and death times of their corresponding features.



This is just the bottleneck distance of the log-scale diagrams.

$$\begin{aligned} \log a - \log b &< \varepsilon \\ \log \frac{a}{b} &< \varepsilon \\ \frac{a}{b} &< 1 + \varepsilon \end{aligned}$$

There are two phases, one is geometric the other is topological.

Geometry

Build a filtration, i.e. a filtered complex.

We'll focus on this side.

Topology
(linear algebra)

Compute the persistence diagram

(Run the Persistence Algorithm).

Running time is polynomial in the **size** of the complex.

Idea 1: Use the Delaunay Triangulation

Good: It works, (alpha-complex filtration).

Bad: It can have size $\mathbf{n^{O(d)}}$.

Idea 2: Connect up everything close.

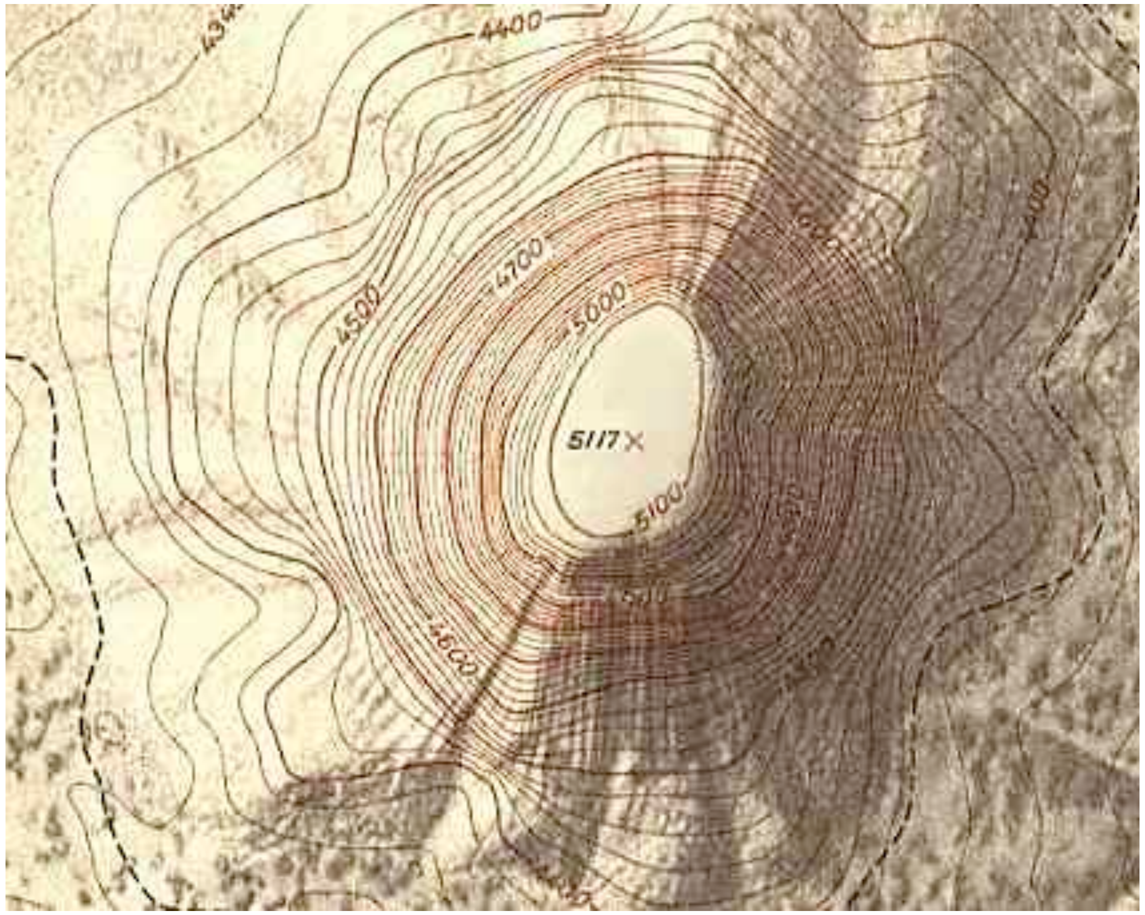
Čech Filtration: Add a k-simplex for every $k+1$ points that have a smallest enclosing ball of radius at most α .

Rips Filtration: Add a k-simplex for every $k+1$ points that have all pairwise distances at most α .

Still $\mathbf{n^d}$, but we can quit early.

Topology is not Topography

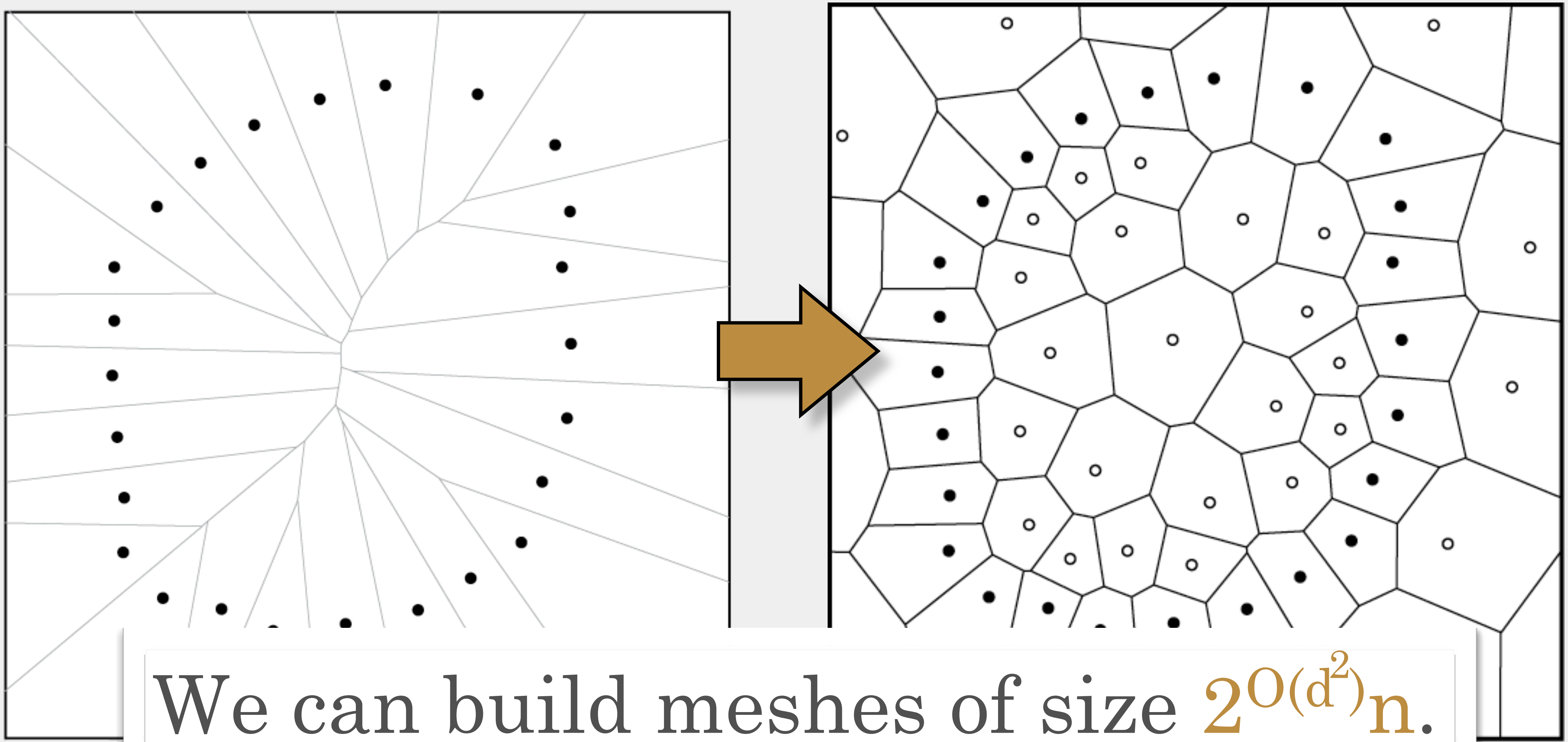
Sublevel sets



(But in our case, there are some similarities)

Nobel Peace Prize!

Our Idea: Build a **quality mesh**.



We can build meshes of size $2^{O(d^2)}n$.

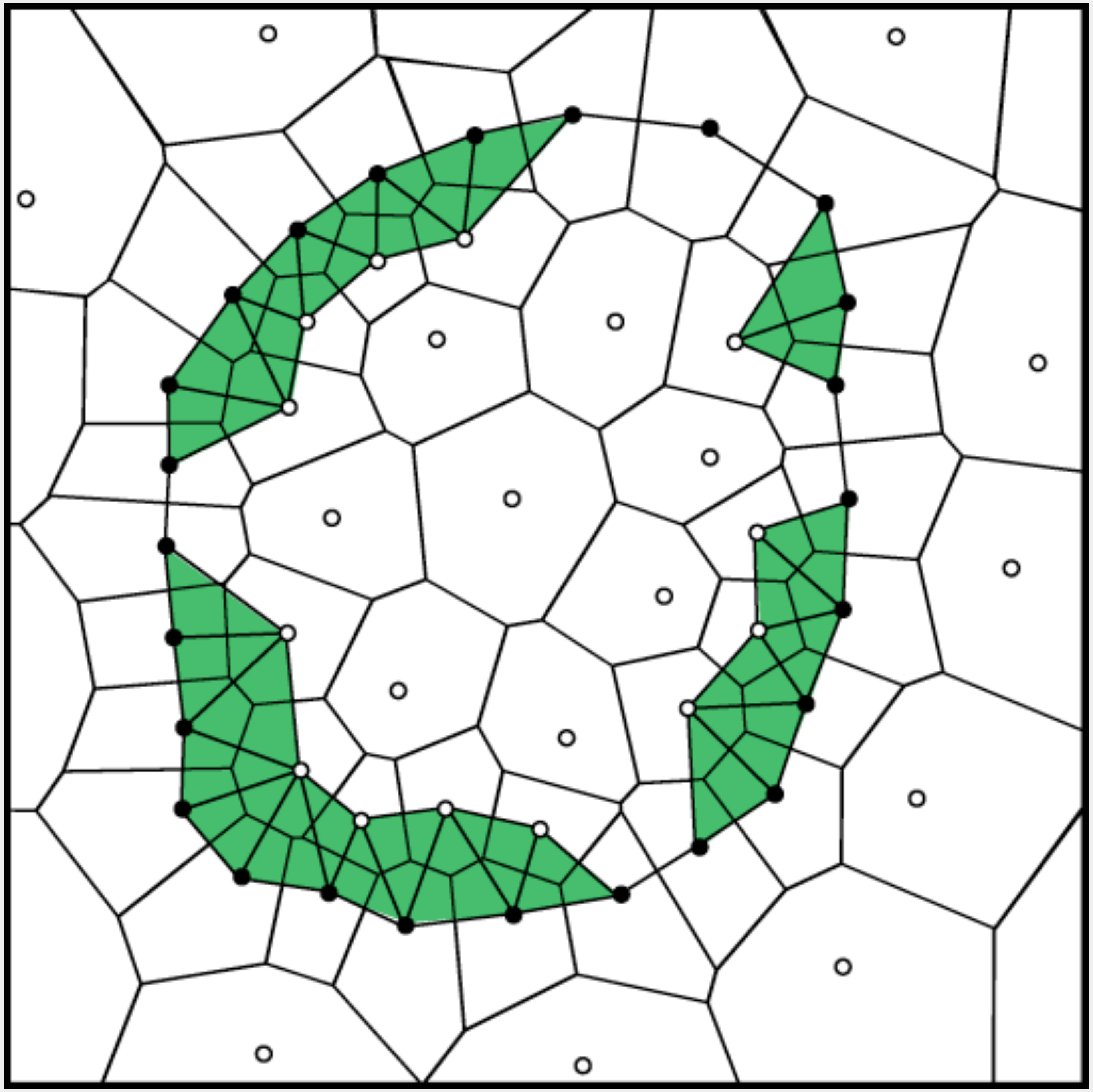
The α-mesh filtration

1. Build a mesh M.

2. Assign birth times to vertices based on distance to P (special case points very close to P).

3. For each simplex s of Del(M), let birth(s) be the min birth time of its vertices.

4. Feed this filtered complex to the persistence algorithm.



Approximation via interleaving.

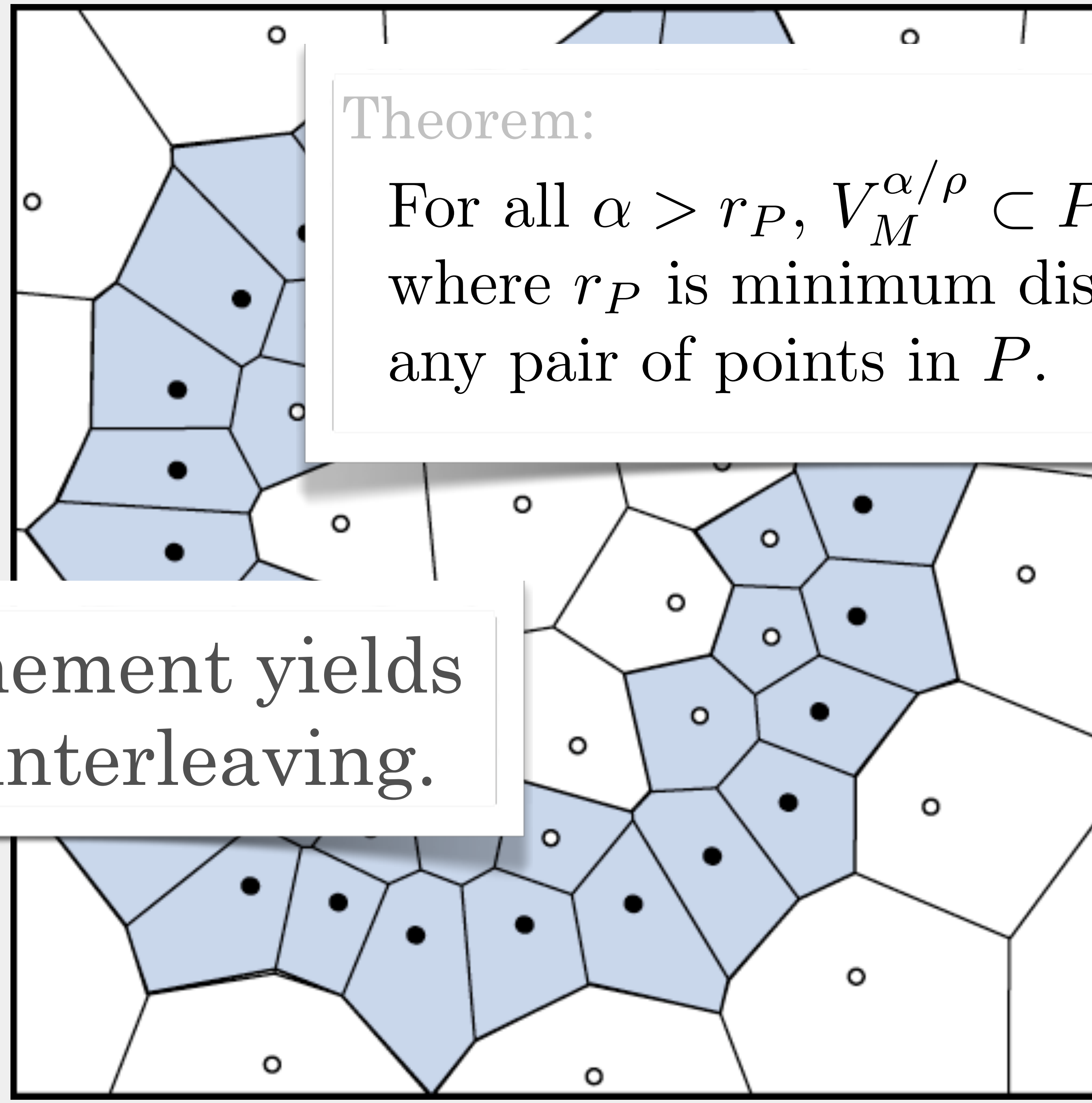
Definition:

Two filtrations, $\{P_\alpha\}$ and $\{Q_\alpha\}$ are ε-*interleaved* if $P_{\alpha-\varepsilon} \subseteq Q_\alpha \subseteq P_{\alpha+\varepsilon}$ for all α.

Theorem [Chazal et al, '09]:

If $\{P_\alpha\}$ and $\{Q_\alpha\}$ are ε-interleaved then their persistence diagrams are ε-close in the bottleneck distance.

The Voronoi filtration interleaves with the offset filtration.



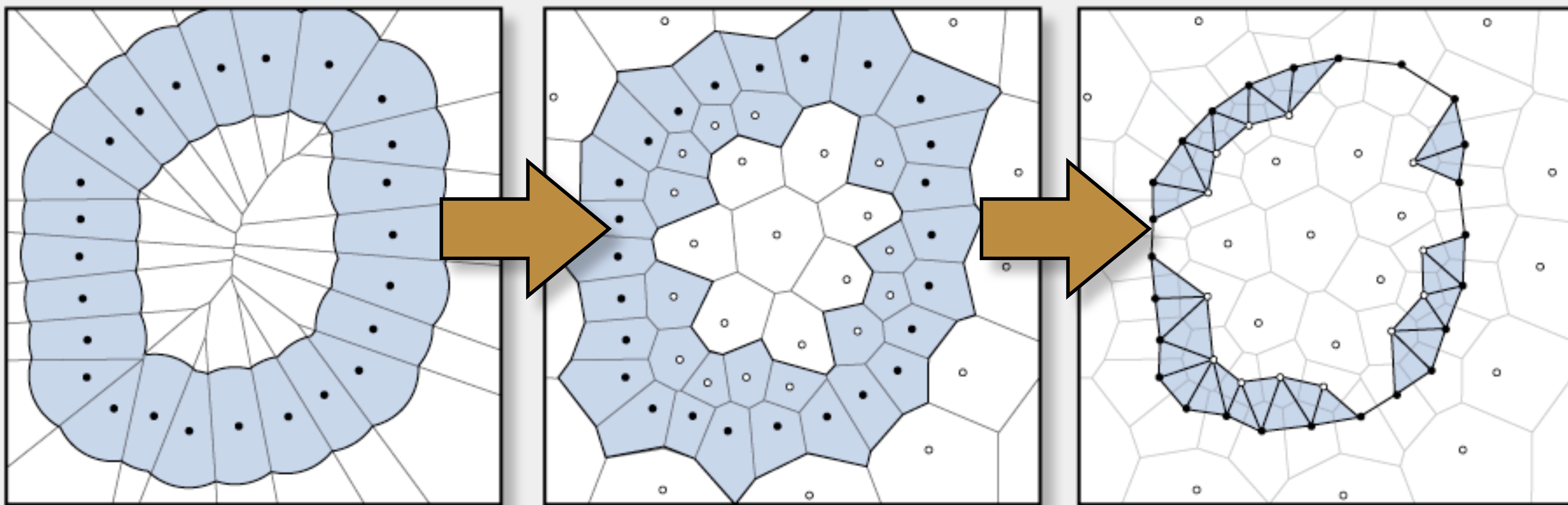
Theorem:

For all $\alpha > r_P$, $V_M^{\alpha/\rho} \subset P^\alpha \subset V_M^{\alpha\rho}$, where r_P is minimum distance between any pair of points in P.

Finer refinement yields a tighter interleaving.

Geometric Approximation

Topological Approximation



If it's so easy, why didn't anyone think of this before?

Theorem [Hudson, Miller, Phillips, '06]:

A quality mesh of a point set can be constructed in $O(n \log \Delta)$ time, where Δ is the spread.

Theorem [Miller, Phillips, Sheehy, '08]:

A quality mesh of a *well-paced* point set has size $O(n)$.

The Results

1. Build a mesh M.
   Over-refine it.
   Use linear-size meshing.

2. Assign birth times to vertices based on distance to P (special case points very close to P).**

3. For each simplex s of Del(M), let birth(s) be the min birth time of its vertices.

4. Feed this filtered complex to the persistence algorithm.

	Approximation ratio	Complex Size
Previous Work	1	$n^{O(d)}$
Simple mesh filtration	ρ	$2^{O(d^2)} n \log \Delta$
Over-refine the mesh	$1 + \varepsilon$	$\varepsilon^{-O(d^2)} n \log \Delta$
Linear-Size Meshing	$1 + \varepsilon + 3\theta$	$(\varepsilon\theta)^{-O(d^2)} n$

Thank you.