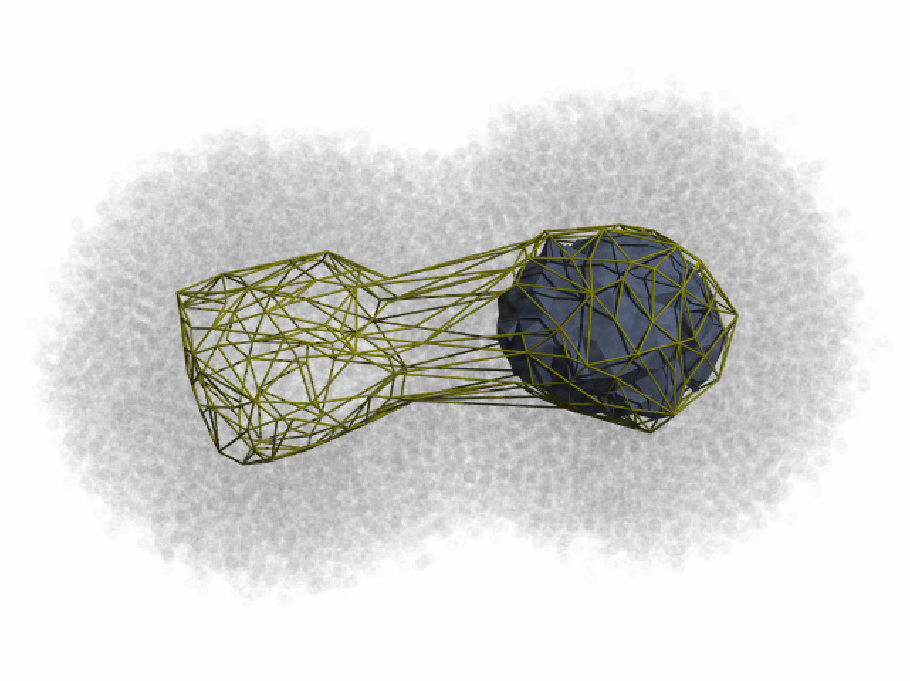
Topological Data Analysis - II



Afra Zomorodian
Department of Computer Science
Dartmouth College
September 4, 2007

Plan

- ☺ Yesterday:
  - Motivation
  - Topology
  - Simplicial Complexes
  - Invariants
  - Homology
  - Algebraic Complexes
- Today
  - Geometric Complexes
  - Persistent Homology
  - The Persistence Algorithm
  - Application to Natural Images

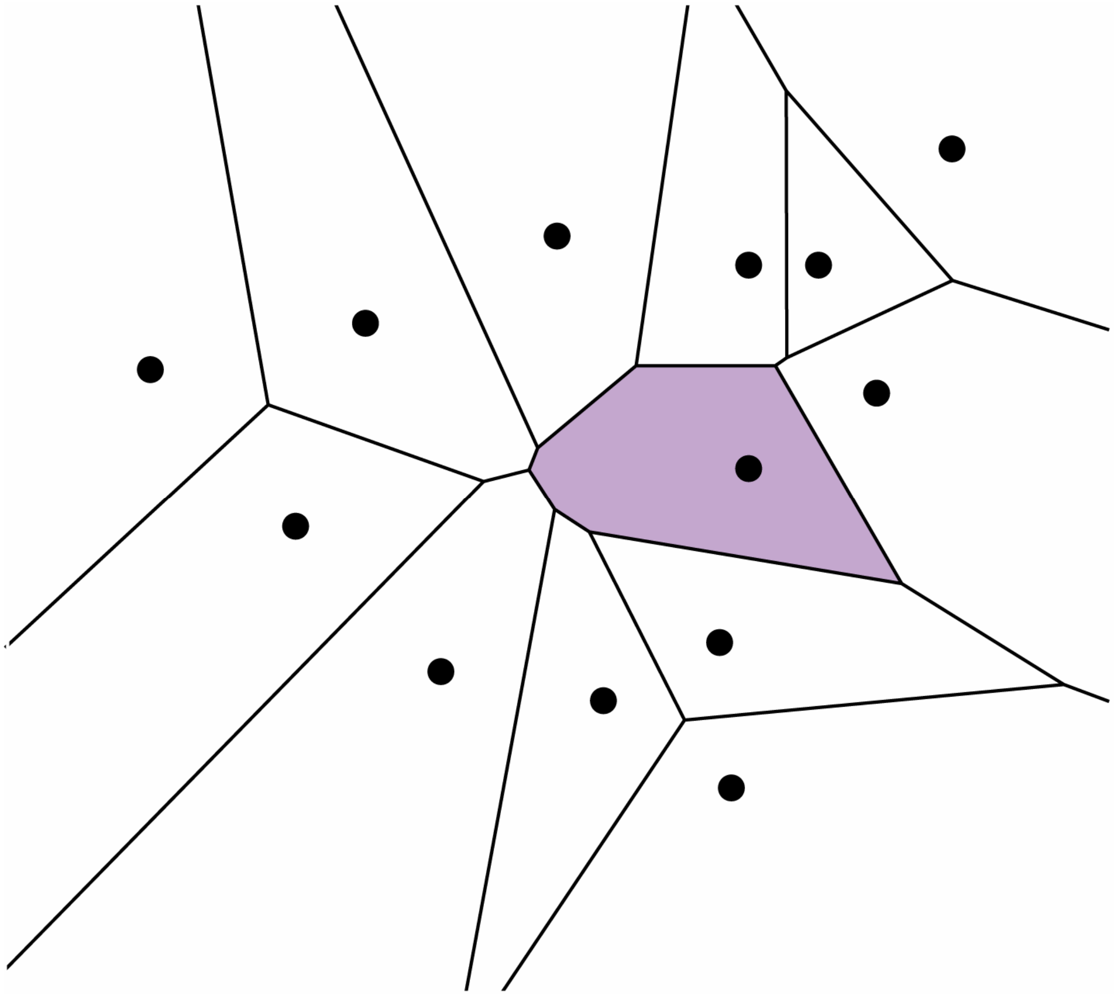
Outline

- Geometric Complexes
  - Voronoi Diagram
  - Delaunay Triangulation
  - Alpha Complex
  - Witness Complex
  - Summary
- Persistent Homology
- The Persistence Algorithm
- Application to Natural Images

Recall

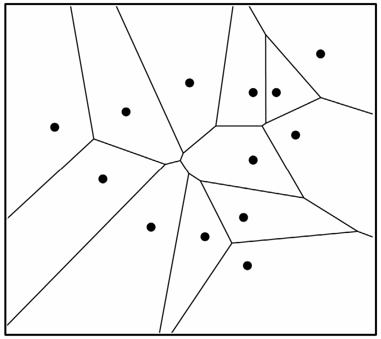
- Procedure
  - Cover points to get approximation of underlying space
  - Take nerve to get combinatorial representation
- Example: ε-balls around points as cover
- Idea: Use geometry of embedding space to generate cover

Voronoi Diagram



Voronoi Diagram

- $p \in M \subseteq \mathbb{R}^2$
- Voronoi cell V(p):
  closest points to p in $\mathbb{R}^2$

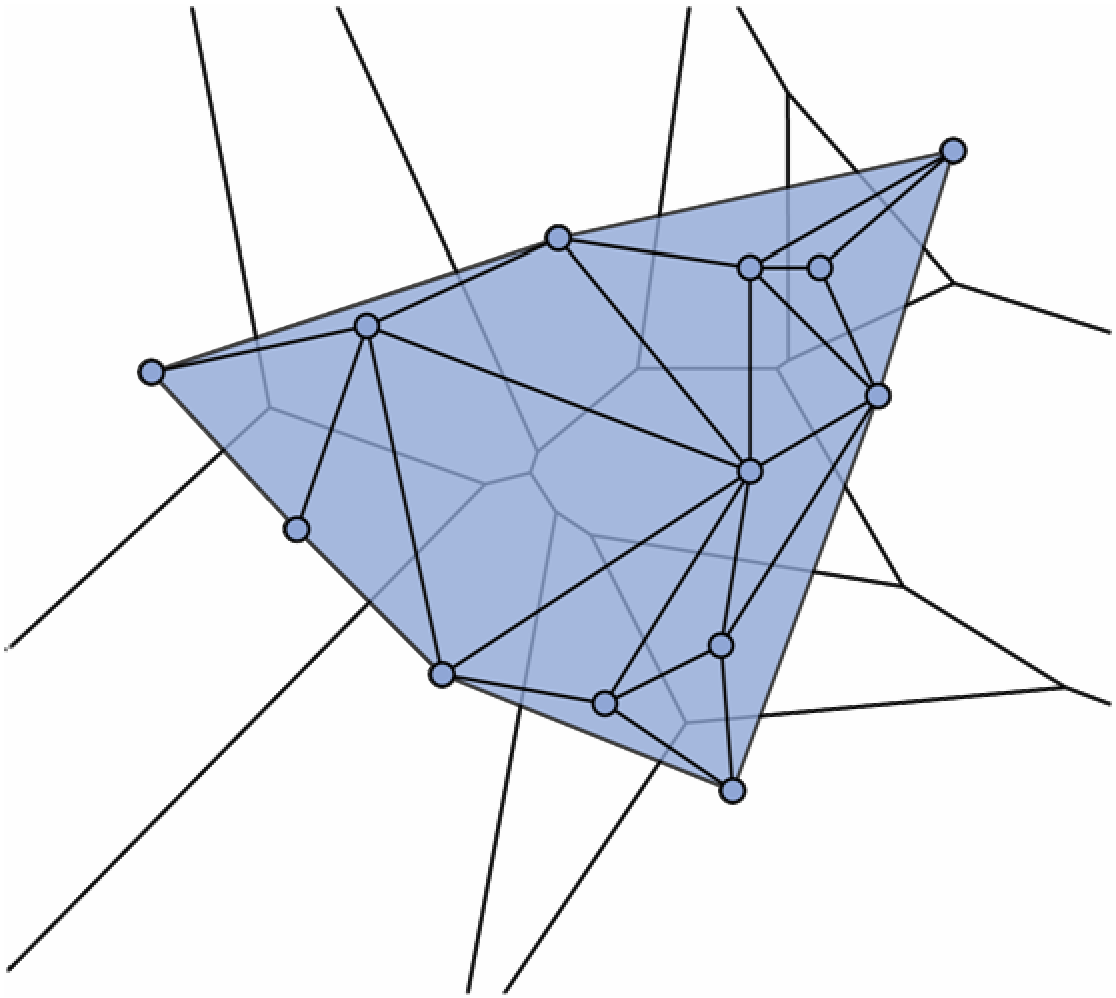


$$V(p) = \left\{ x \in \mathbb{R}^2 \mid \mathrm{d}(x,p) \leq \mathrm{d}(x,y), \forall y \in M \right\}$$

- Voronoi Diagram: Decomposition of $\mathbb{R}^2$ into Voronoi cells
- Voronoi (1868 – 1908)

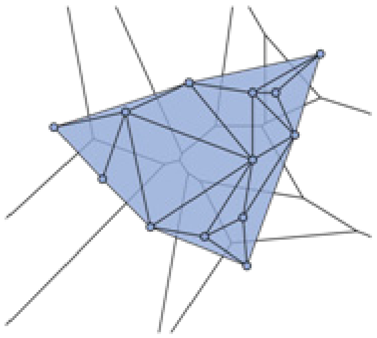
- Idea: Use Voronoi cells as cover!

Delaunay Triangulation

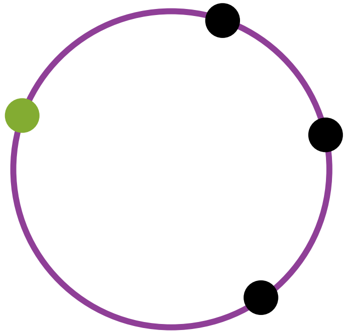


Delaunay Triangulation

- Delaunay Triangulation: nerve of Voronoi cover
- *Computational Geometry*

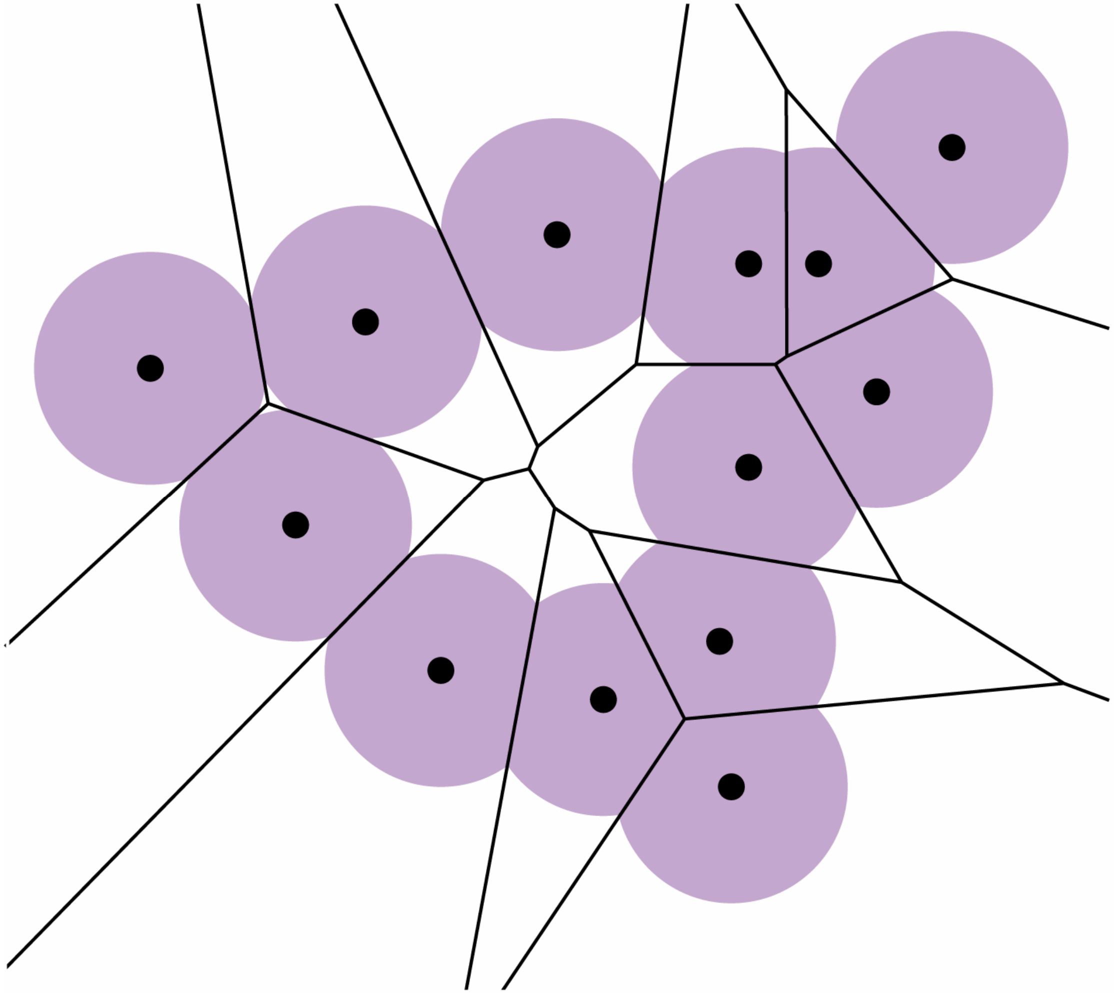


- General position assumption
  - no events with probability 0
  - no k + 1 points on (k – 1)-sphere
  - has to be handled in practice

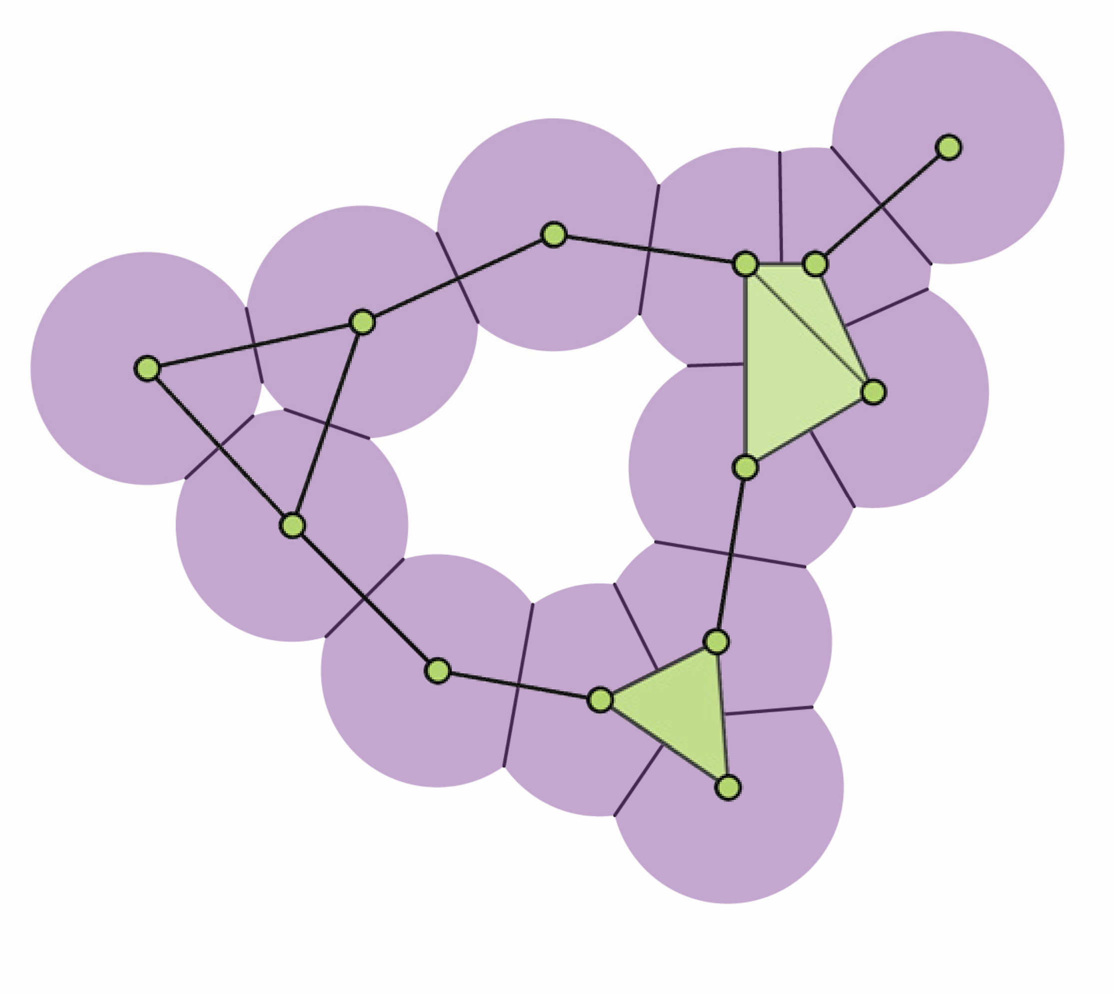


- Fast algorithms for $\mathbb{R}^3$
- Delaunay (1890 – 1980)

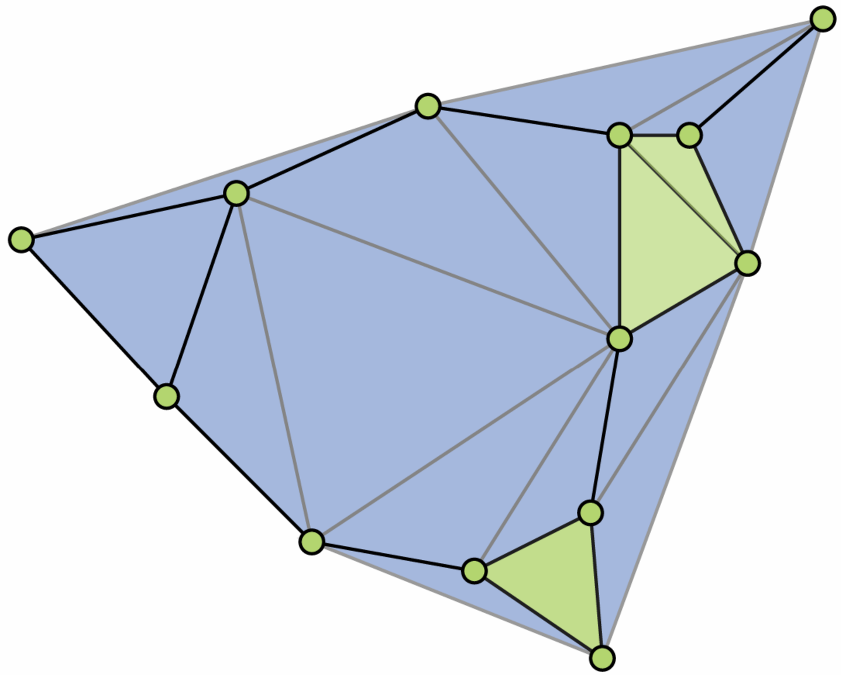
Restricted Voronoi



Alpha Complex

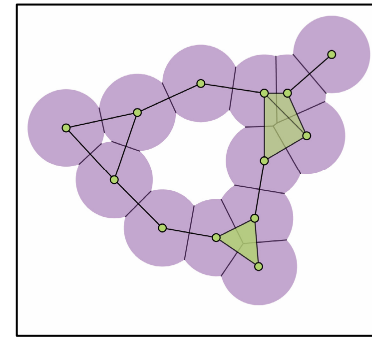


Delaunay Subcomplex



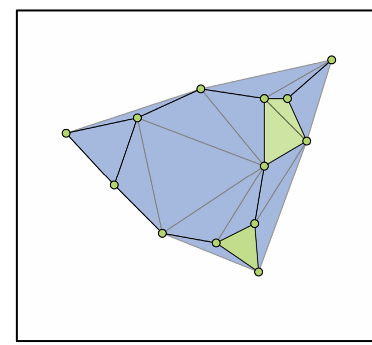
Alpha Complex

- Alpha cell: $A_\varepsilon(p) = B_\varepsilon(p) \cap V(p)$
- Alpha shape: union of alpha cells
- Alpha complex: nerve of alpha shape



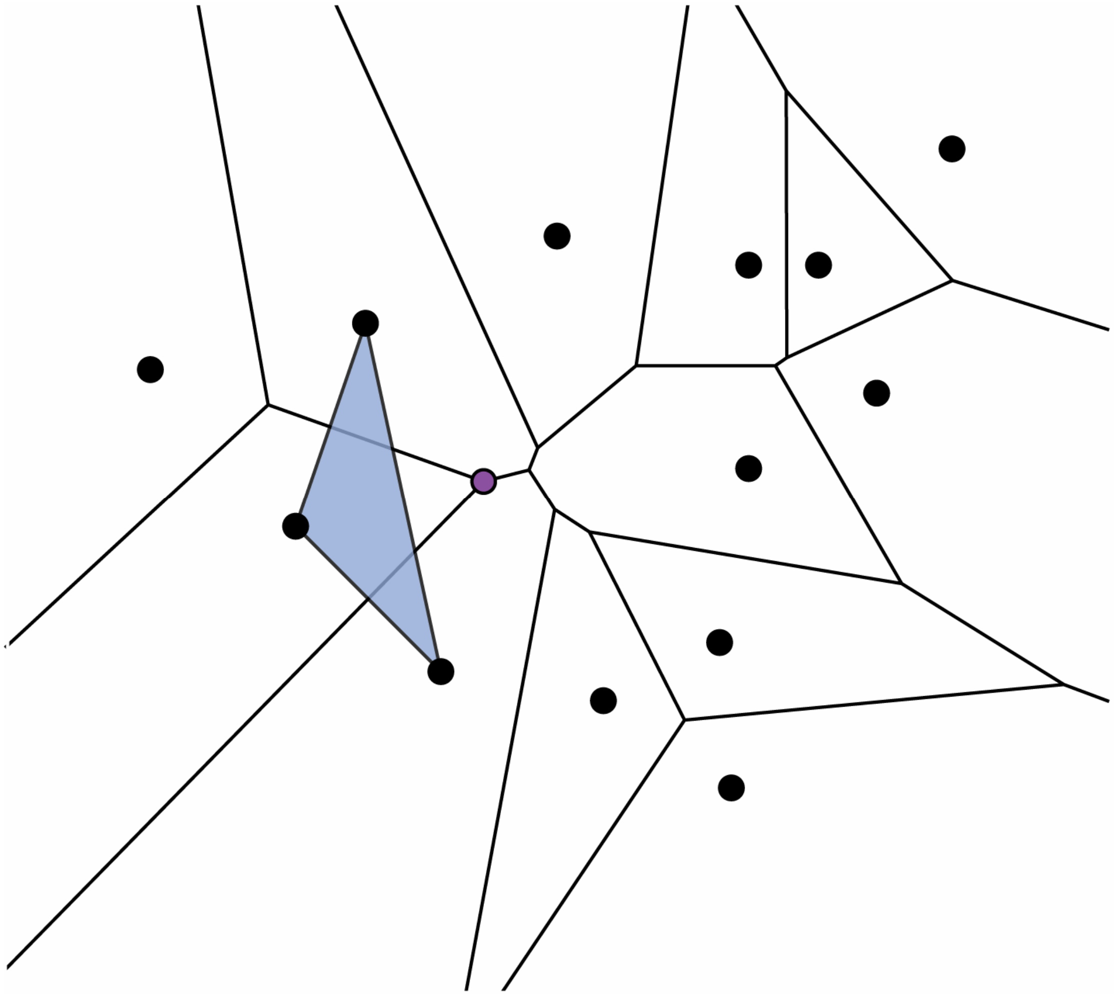
$$A_\epsilon(M) = \left\{ \operatorname{conv} T \mid T \subseteq M, \bigcap_{p \in T} A_\epsilon(p) \neq \emptyset \right\}$$

- Let D be the Delaunay triangulation
  - $A_0 = \emptyset$
  - $A_\varepsilon \subseteq D$
  - $A_\infty = D$



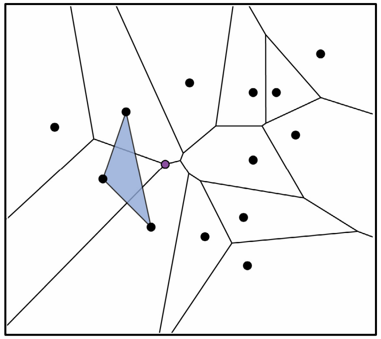
- $A_\varepsilon \simeq C_\varepsilon$
- [Edelsbrunner, Kirkpatrick, and Seidel ’83], et al.

Strong Witness



Strong Witness

- Given: Point set $M \in \mathbb{R}^d$
- Strong witness: $x \in \mathbb{R}^d$
  - x is equidistant from $v_0, \ldots, v_k \in M$
  - x has no closer neighbor in M
  - x witnesses k-simplex $\{v_0, \ldots, v_k\}$



- Idea: Sample for witnesses

- Problem: Prob(strong witness) = 0 for discrete set M

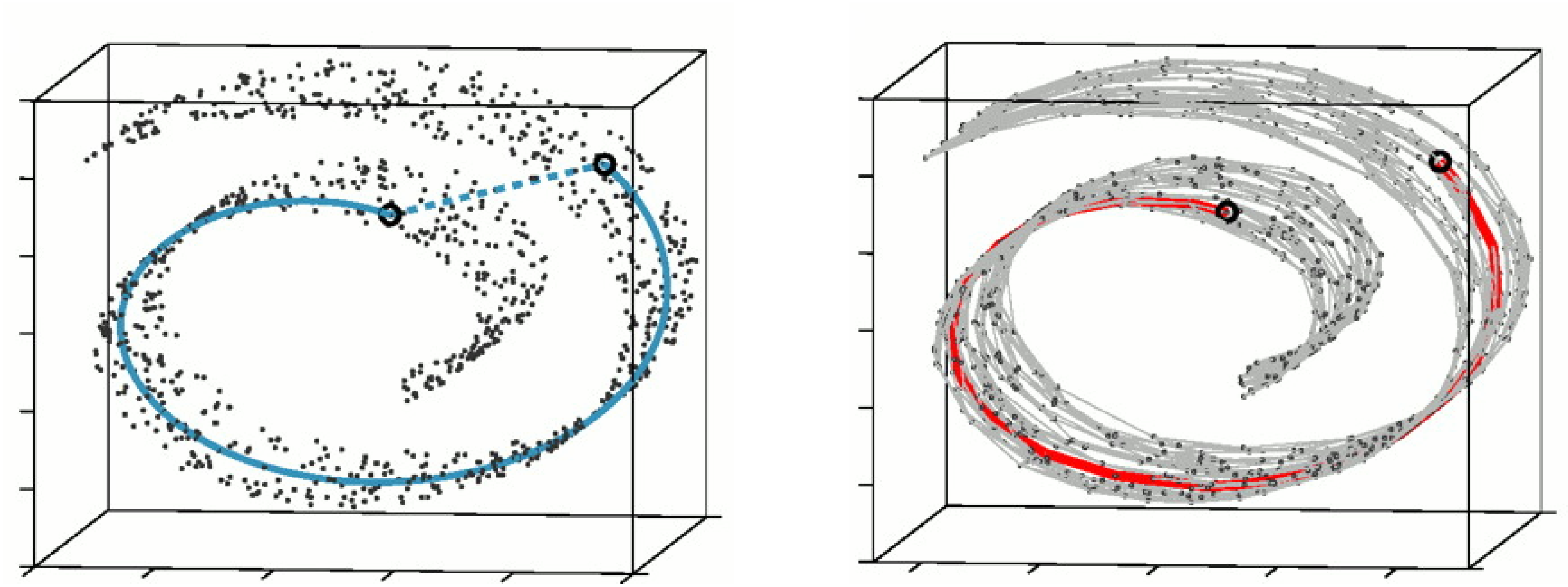
Weak Witness

- Weak witness: $x \in \mathbb{R}^d$
  - $|x - v_i| \leq |x - v|$ for $i = 0, \ldots, k$ and $v \in M \setminus \{v_0, \ldots, v_k\}$
  - x's closest $k + 1$ neighbors are $v_0, \ldots, v_k$
  - x witnesses k-simplex $\{v_0, \ldots, v_k\}$ weakly

- Strong witness $\Rightarrow$ weak witness

- **(Theorem [de Silva])**
  A simplex has a strong witness iff all its faces have weak witnesses.

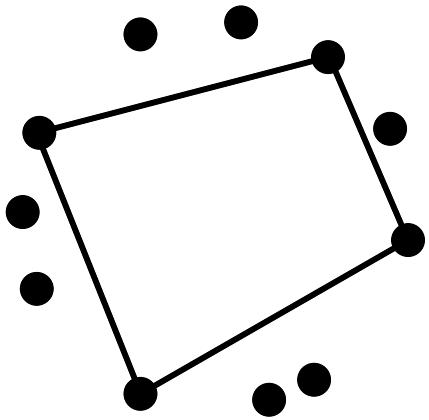
Isomap



- We want to capture the underlying space, not the embedding space
- Idea: Restrict witnesses to given points M

[Tenenbaum et al. ’00]

Witness Complex



- Given: N points M
- Choose: n landmarks L
- M \ L will act as witnesses

- D = n $\times$ N, distance matrix
- Construct ε-graph on L:
  Edge [ab] $\in W_\varepsilon(M)$ iff there exists a witness with max(D(a,i), D(b,i)) $\leq \varepsilon$

- Do Vietoris-Rips Expansion

Complex Summary

Complex	Name	Idea	Scales?	Extends?
Cech	C_ε	Nerve of ε-balls	1K	$\sim$
Vietoris-Rips	V_ε	Pairwise dist < ε	1K	Y
Alpha	A_ε	Nerve of restricted Voronoi	500K	$d \leq 3$
Witness	W_ε	Landmarks and witnesses	1K	Y

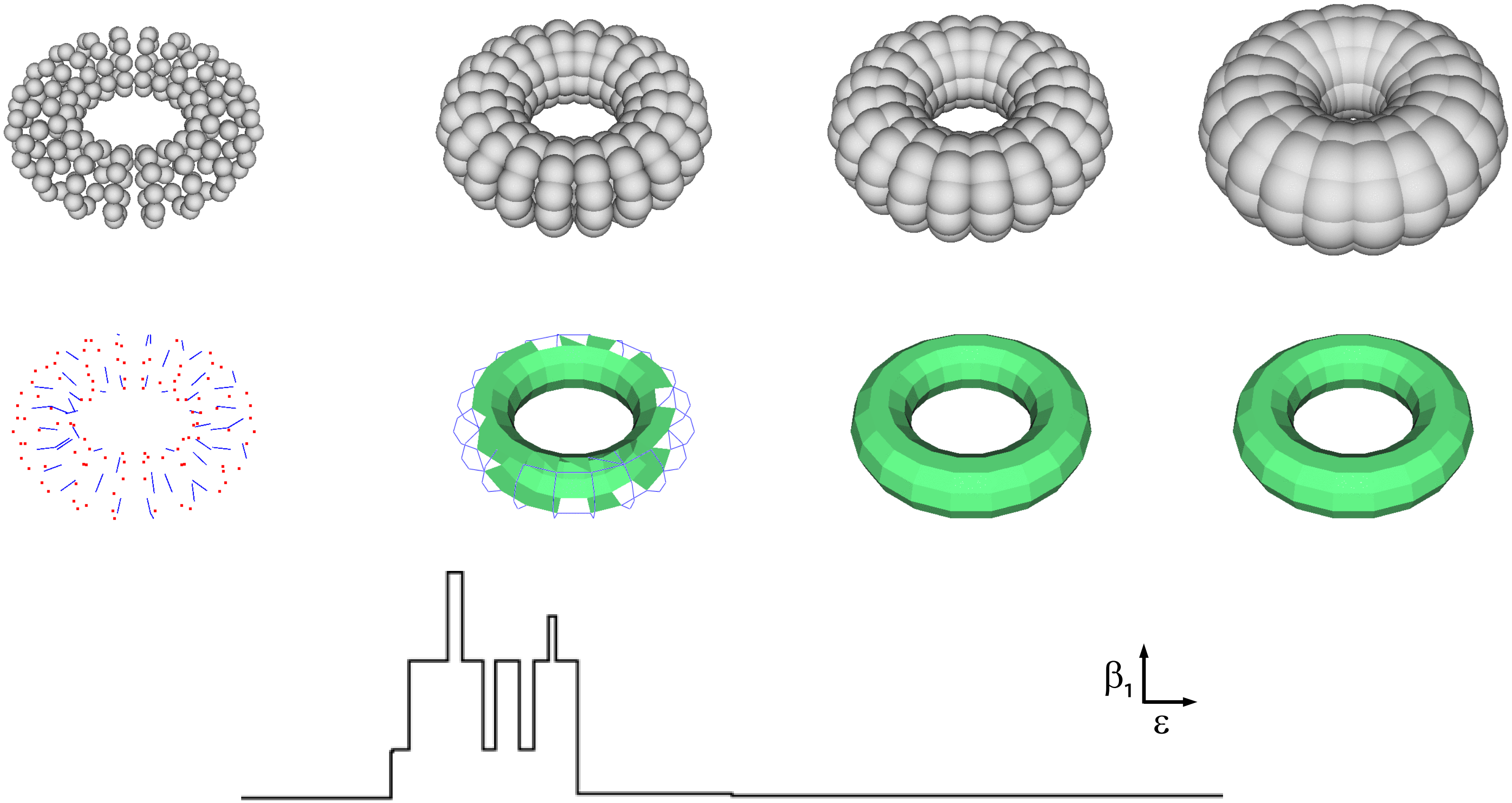
- Conformal Alpha – no global scale parameter
- Flow – stable manifolds of distance function
- Cubical – rasterize, usually interpretation of images

Outline

☺ Geometric Complexes

- Persistent Homology
  - Filtrations
  - Algebraic Result
  - Simple Examples
- The Persistence Algorithm
- Application to Natural Images

The Question of Scale



Combinatorial Topology

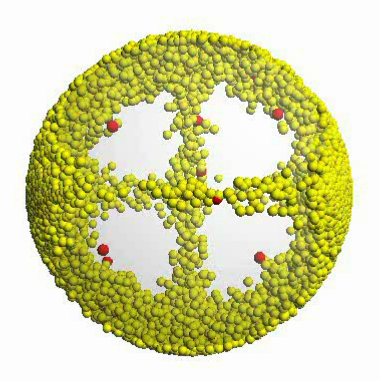
Filtration

- A filtration of a space *X* is a nested sequence of subspaces:

$$\emptyset = X^0 \subseteq X^1 \subseteq \cdots \subseteq X^l \subseteq \cdots \subseteq X^m = X$$

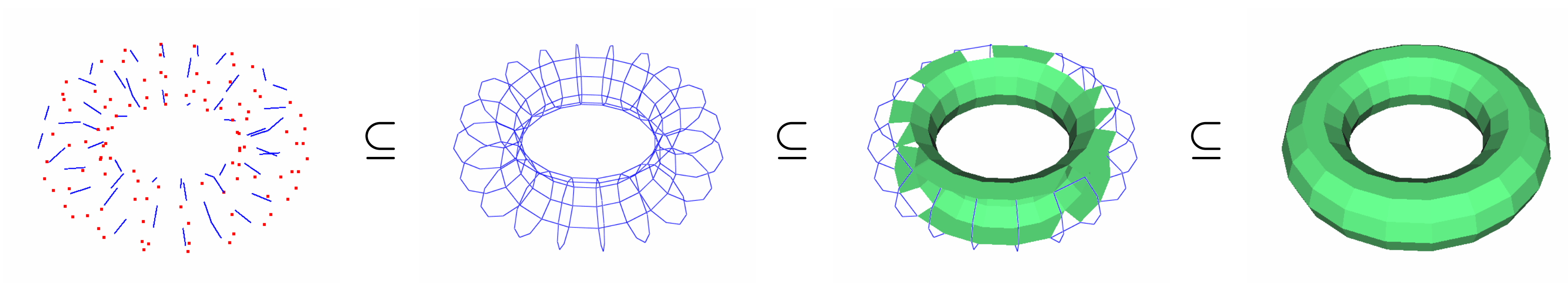
- $C_\varepsilon \subseteq C_{\varepsilon'}$ if $\varepsilon \leq \varepsilon'$ (Also true for V_ε, A_ε, and W_ε)

- Simplices are always added, never removed
- Implies partial order on simplices
- Full order: sequence of simplices
- K_i = union of first i simplices in sequence



Witness Complex

Inductive Systems



$$K^{250} \xrightarrow{\;i\;} K^{500} \longrightarrow K^{994} \longrightarrow K^{1452}$$

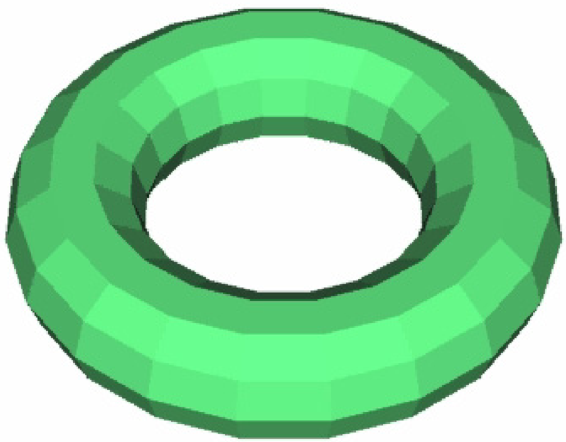
Functoriality $\downarrow$

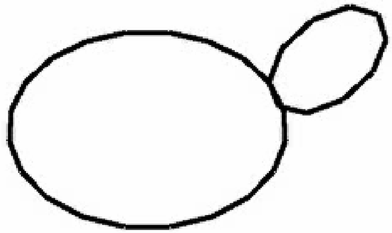
$$H_k(K^{250}) \longrightarrow H_k(K^{500}) \longrightarrow H_k(K^{994}) \longrightarrow H_k(K^{1452})$$

Idea: Follow basis elements from birth to death

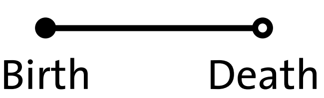
Problem: Need a compatible basis!

Persistent Homology





- Persistence barcode: multiset of intervals



Algebraic Result

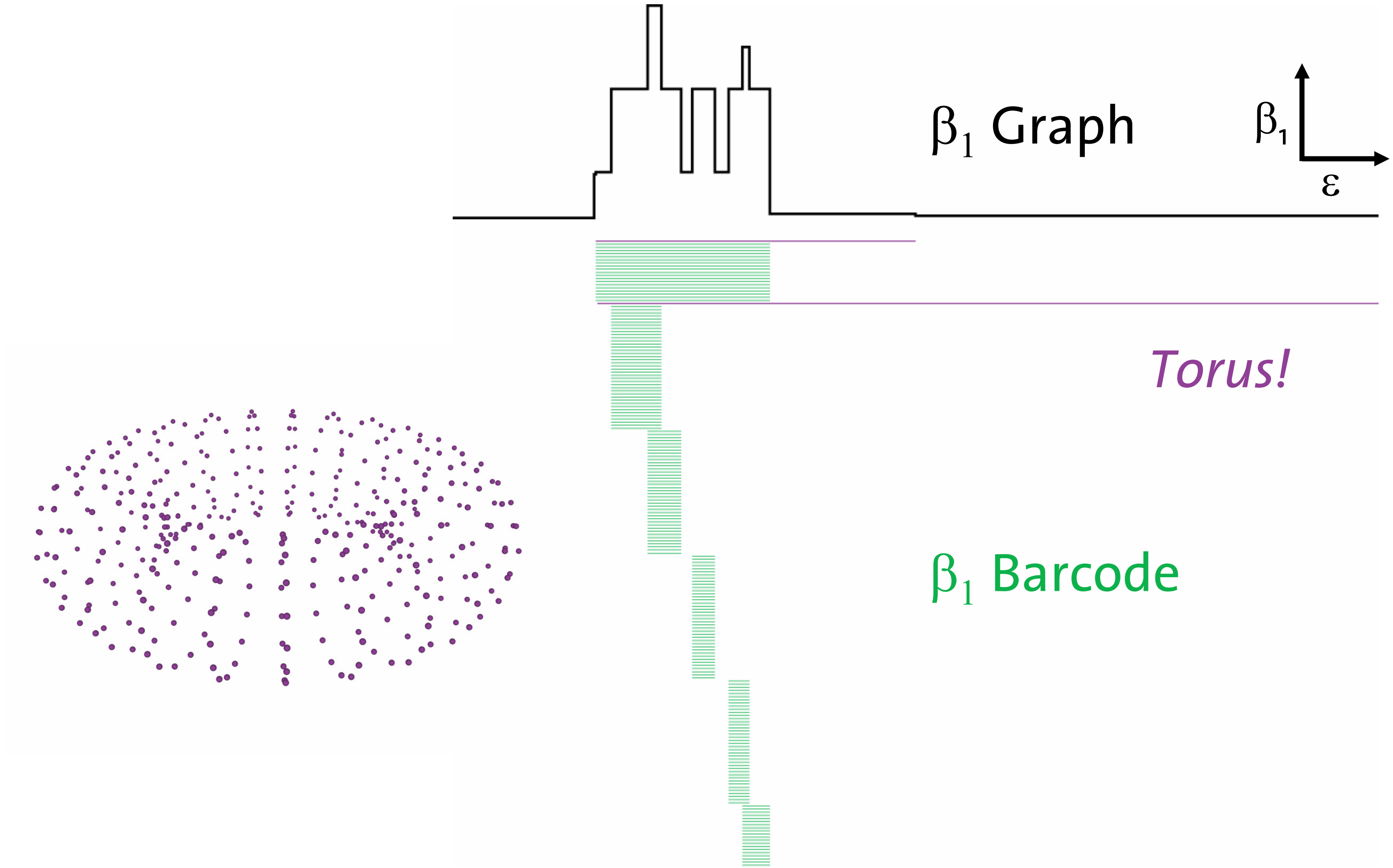
1. Correspondence
   - Input: Filtration
   - Structure of homology: graded *k*[*t*]-module
2. Classification
   - *k*, a field ⇒ *k*[*t*] is a PID
   - Structure theorem for graded PIDs
3. Parameterization
   - *n* half-infinite
   - *m* finite
   - Barcode: multiset of *n*+*m* intervals (birth, death)
   - *Complete discrete invariant!*

$$\left(\bigoplus_{i=1}^{n} \Sigma^{\alpha_i} k[t]\right) \oplus \left(\bigoplus_{j=1}^{m} \Sigma^{\gamma_j} k[t]/(t^{n_j})\right)$$

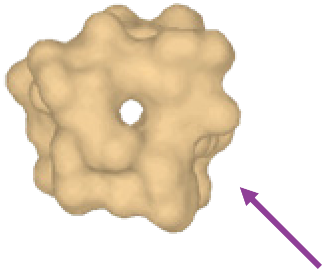
$$\Sigma^{\alpha_i} k[t] \mapsto (\alpha^i, \infty)$$

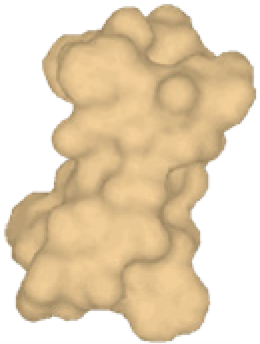
$$\Sigma^{\gamma_j} k[t]/(t^{n_j}) \mapsto (\gamma_j, \gamma_j + n_j)$$

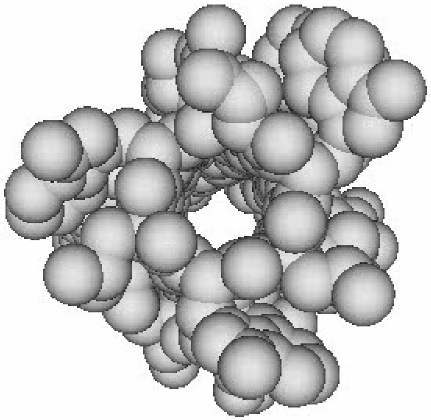
Deconstructing the Graph (2D)

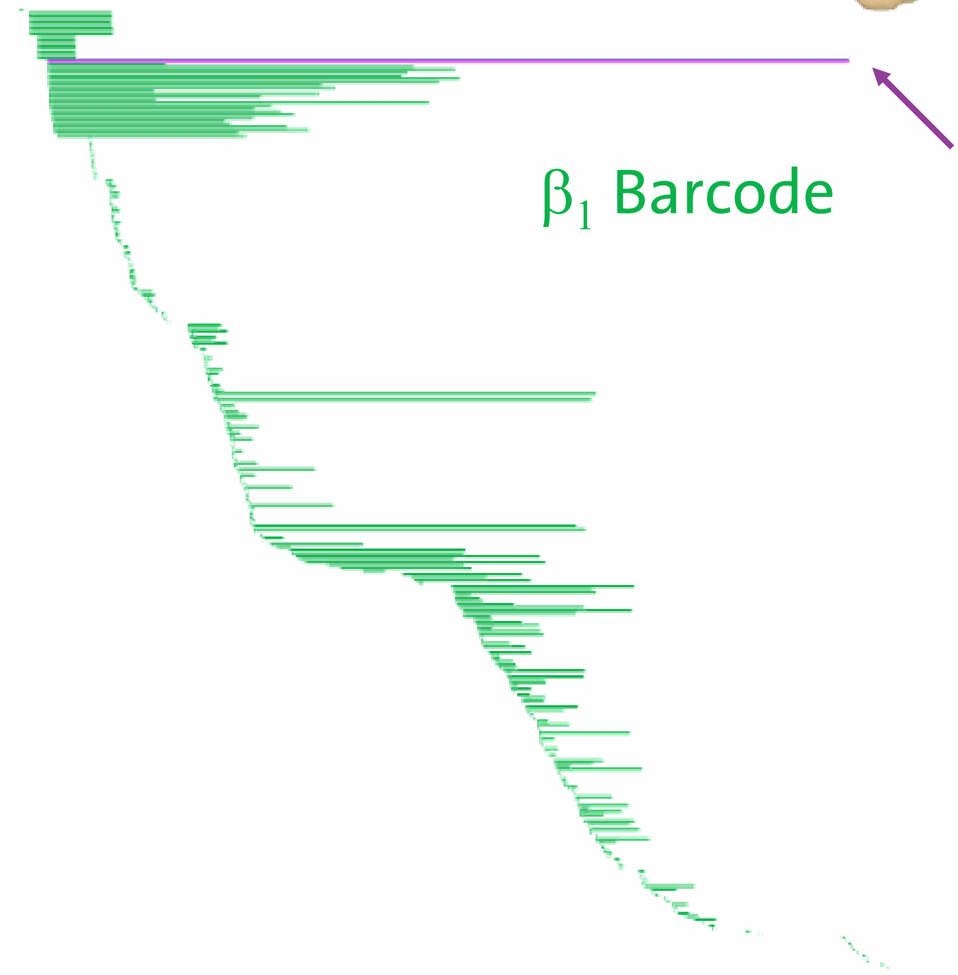


Discovering 3D Structure









Outline

- ☺ Geometric Complexes
- ☺ Persistent Homology
- The Persistence Algorithm
  - Adding a Simplex
  - Example Filtration
- Application to Natural Images

Adding a Simplex

- Given: Filtered complex K
- $K_i = K_{i-1} \cup \sigma$, where σ is a k-simplex

- Let $c = \partial\sigma$. c is a $(k-1)$-chain.
- **(Lemma)** c is a cycle.
- Proof: $\partial c = \partial\partial\sigma = 0$.

- **(Lemma)** c is in K_{i-1}.
- Proof: K_i is a simplicial complex.

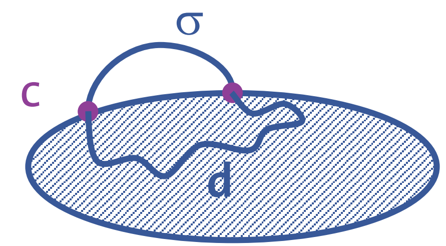
Gaussian Elimination

- σ is a k-simplex
- $c = \partial\sigma$ is a $(k-1)$-cycle in K_{i-1}
- Two cases: c is a boundary or not in K_{i-1}

- M_k is matrix for ∂_k
- c is a boundary iff
  - it is in range(M_k)
  - we can write it in terms of a basis for M_k
- Gaussian elimination maintains a basis for range(M_k)
- Filtration and persistence imply ordering on pivots

Case 1: c is a boundary in K_{i-1}

- If c is a boundary, then
  $\exists d \in C_{k+1}(K_{i-1})$, such that $c = \partial d$

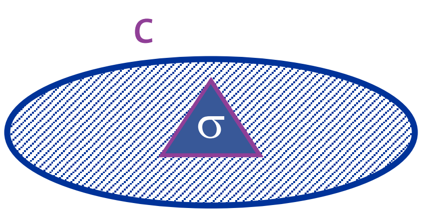


- **(Lemma)** $\sigma + d$ is a k-cycle in K_i.
- Proof: $\partial(\sigma + d) = \partial\sigma + \partial d = c + \partial d = 0$.

- σ creates a new k-cycle class
- σ is a creator

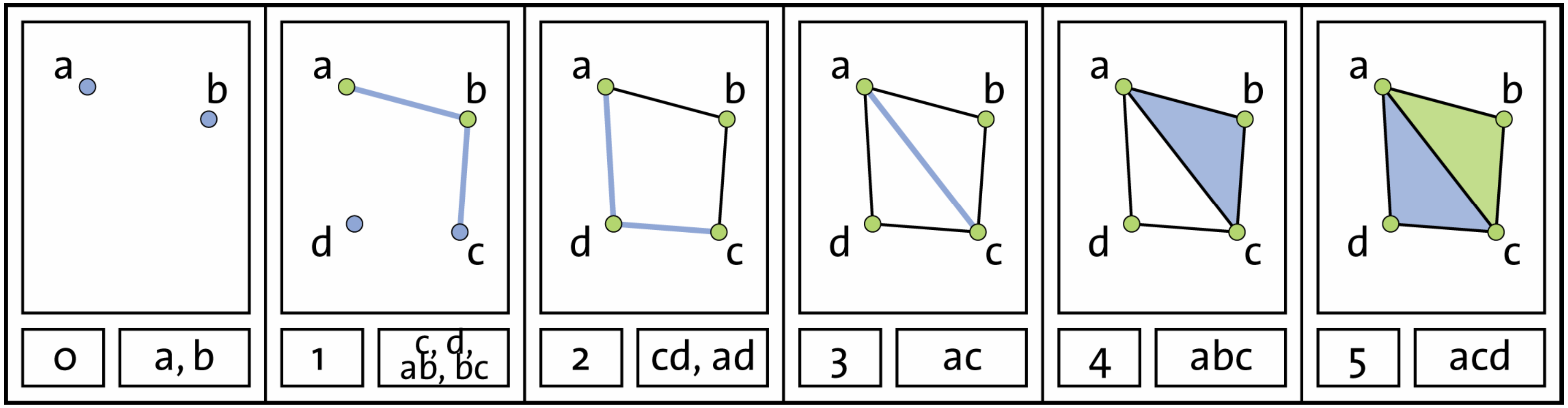
Case 2: c is not a boundary in K_{i-1}

- **(Lemma)** c becomes a boundary in K_i.
- Proof: $c = \partial\sigma$.



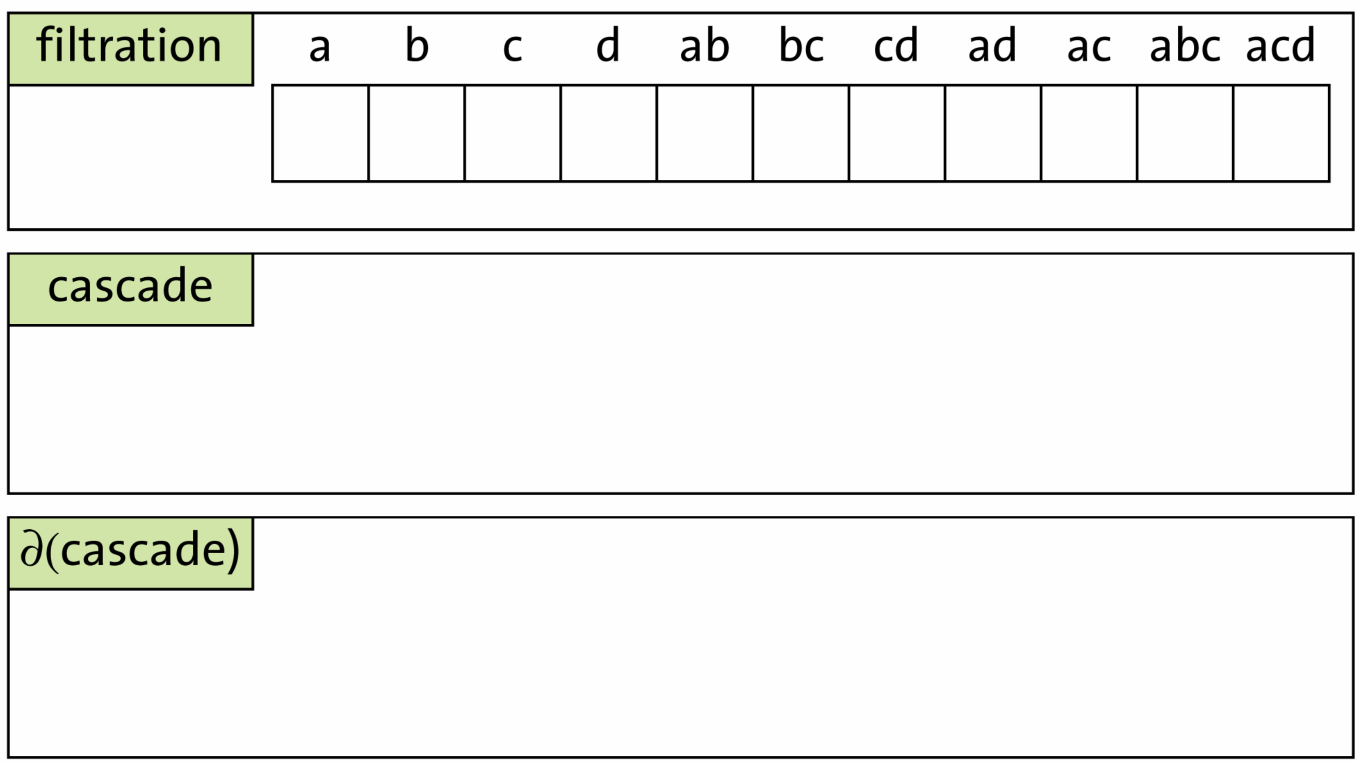
- In K_{i-1}
  - c is a cycle
  - c is not a boundary
  - c is in a non-boundary homology class
- In K_i: c is a boundary, so its homology class is trivial.
- σ destroys a $(k-1)$-dimensional class
- σ is a destroyer
- Suppose τ created that class that σ destroyed
- We pair (τ, σ) to get the lifetime interval

Example



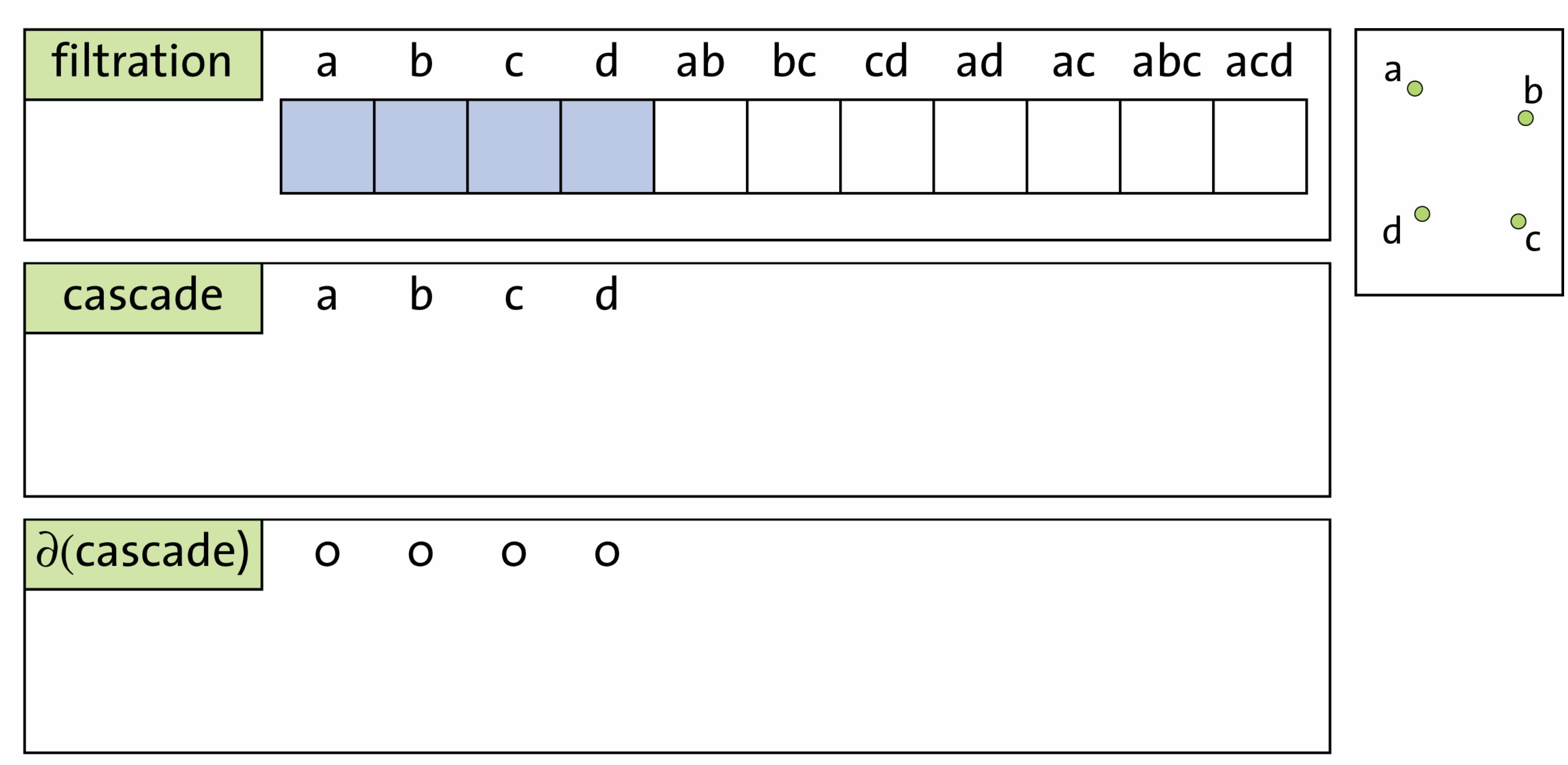
Filtration

- Initially, cascade = σ_i



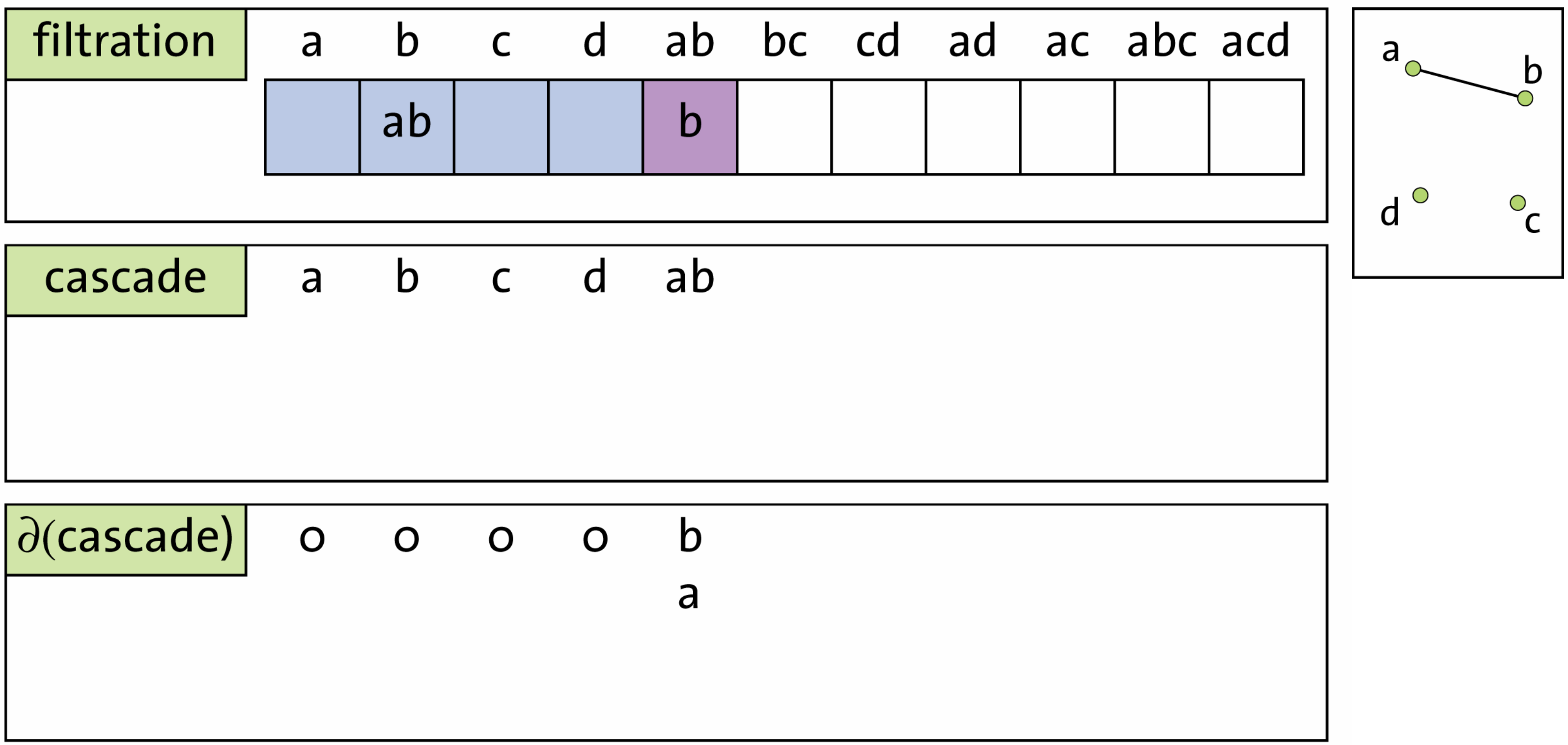
Vertices a, b, c, d

- $\partial\sigma = 0$ for all vertices σ



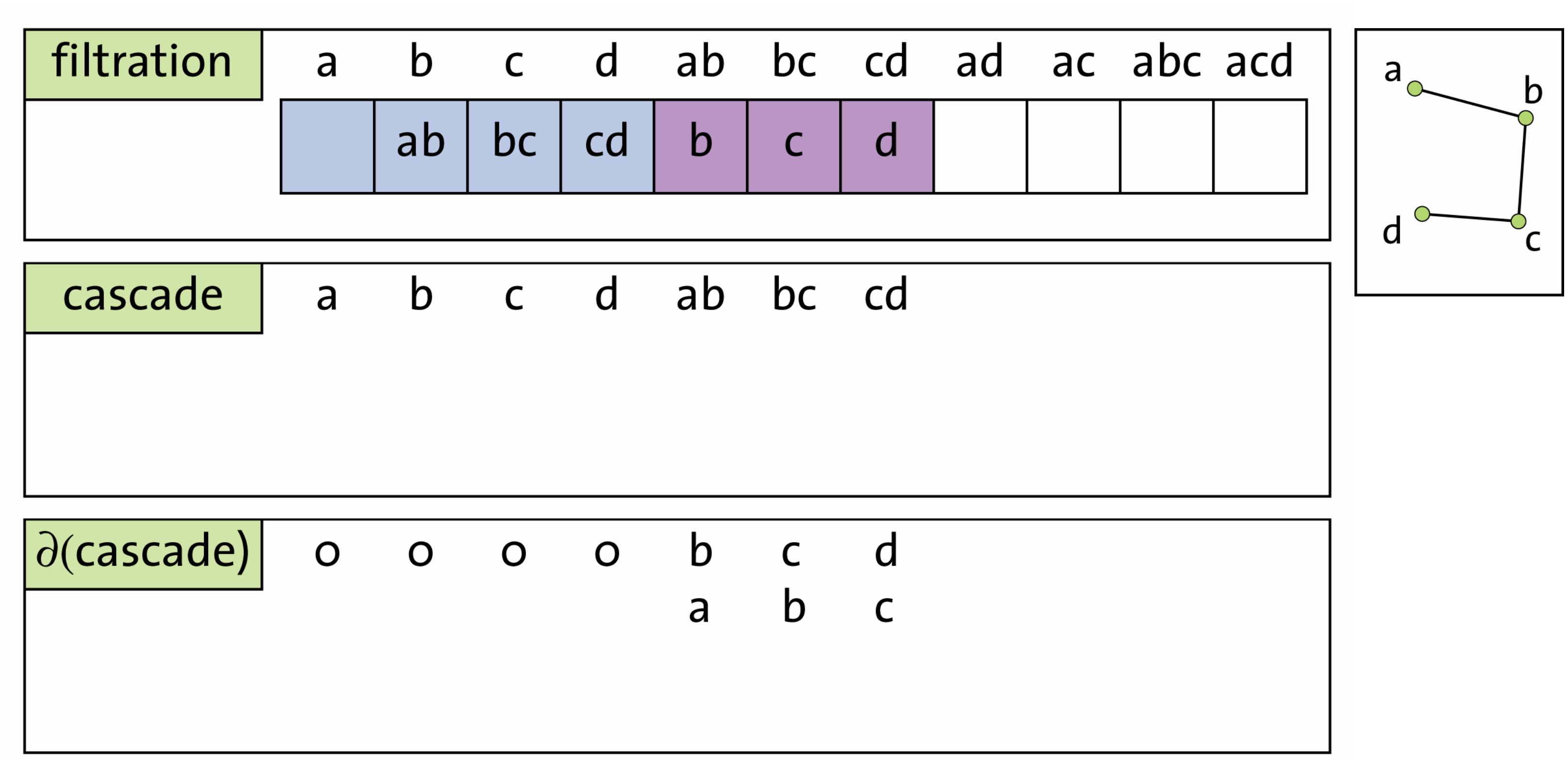
ab

- We sort ∂ab = **b** + a by youngest
- Since b is unpaired, pair with ab



bc, cd

- $\partial bc = \mathbf{c} + b$
- $\partial cd = \mathbf{d} + c$



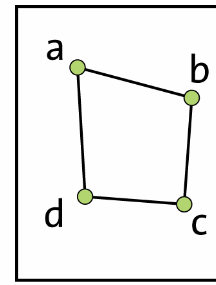
ad

- $\partial ad = (d + a) \sim (d + a) + (d + c) = c + a$
  $\sim (c + a) + (c + b) = b + a \sim (b + a) + (b + a) = 0$

filtration	a	b	c	d	ab	bc	cd	ad	ac	abc	acd
		ab	bc	cd	b	c	d				

cascade	a	b	c	d	ab	bc	cd	ad
								cd
								bc
								ab

∂(cascade)	0	0	0	0	b	c	d	0
					a	b	c	



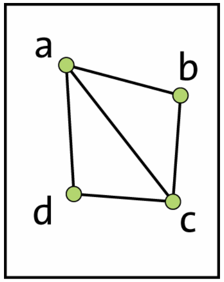
ac

- $\partial ac = (c + a) \sim (c + a) + (c + b) = b + a$
  $\sim (b + a) + (b + a) = 0$

filtration	a	b	c	d	ab	bc	cd	ad	ac	abc	acd
		ab	bc	cd	b	c	d				

cascade	a	b	c	d	ab	bc	cd	ad	ac
								cd	bc
								bc	ab
								ab	

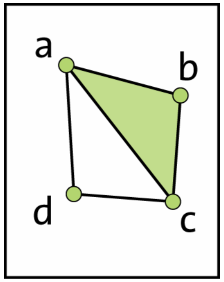
∂(cascade)	0	0	0	0	b	c	d	0	0
					a	b	c		



abc

- $\partial abc =$ **ac** + bc + ab

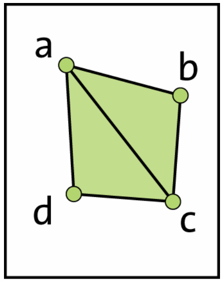
filtration	a	b	c	d	ab	bc	cd	ad	ac	abc	acd
		ab	bc	cd	b	c	d		abc	ac	
cascade	a	b	c	d	ab	bc	cd	ad cd bc ab	ac bc ab	abc	
∂(cascade)	o	o	o	o	b a	c b	d c	o	o	ac bc ab	



acd

- $\partial acd = ac + ad + cd \sim$
  $(ac + ad + cd) + (ac + bc + ab) = \mathbf{ad} + cd + bc + ab$

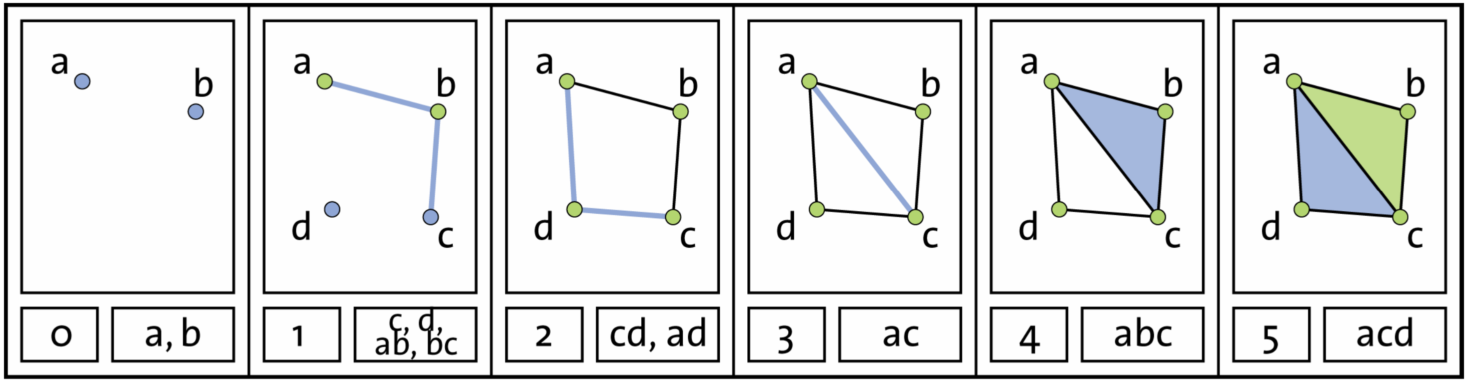
filtration	a	b	c	d	ab	bc	cd	ad	ac	abc	acd
		ab	bc	cd	b	c	d	acd	abc	ac	ac



cascade	a	b	c	d	ab	bc	cd	ad	ac	abc	abc
								cd	bc		
								bc	ab		
								ab			

∂(cascade)	0	0	0	0	b	c	d	0	0	ac	ad
					a	b	c			bc	cd
										ab	bc
											ab

Barcode



- β_0: a is unpaired $\Rightarrow [0, \infty)$
- β_0: (b, ab) $\Rightarrow [0, 1)$
- β_0: (c, bc) $\Rightarrow \emptyset$
- β_0: (d, cd) $\Rightarrow [1, 2)$
- β_1: (ad, acd) $\Rightarrow [2, 5)$
- β_1: (ac, abc) $\Rightarrow [4, 5)$

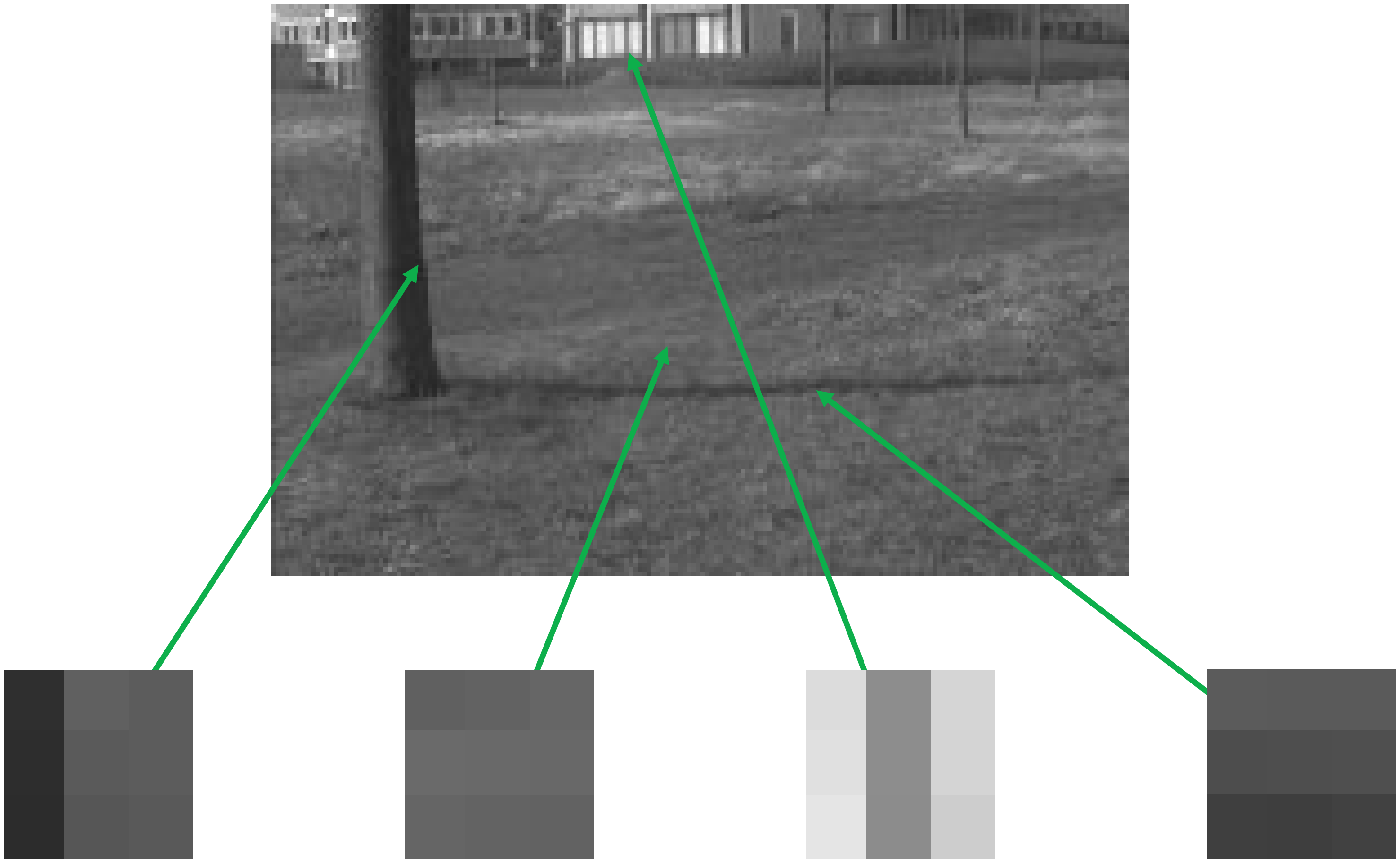
Outline

- ☺ Geometric Complexes
- ☺ Persistent Homology
- ☺ The Persistence Algorithm
- Application to Natural Images

Natural Images

J. H. van Hateren, Neurobiophysics, U. Groningen

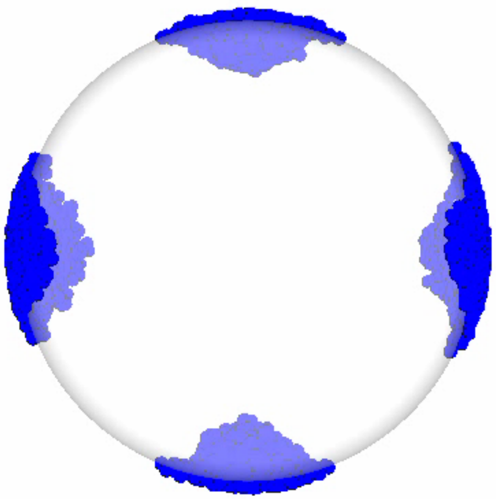
Local Structure: 3 x 3 Patches



$(0.81, 0.62, 0.64, 0.82, 0.65, 0.64, 0.83, 0.66, 0.65) \in \mathbb{R}^9$

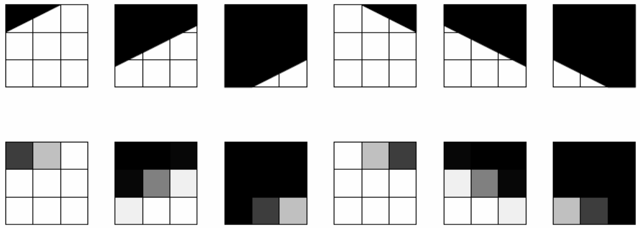
Mumford Dataset

- David Mumford (Brown)
  - 3×3 patches ($\mathbb{R}^9$)
  - Subtract mean intensity ($\mathbb{R}^8$)
  - Remove low contrast patches
  - Rescale to unit length ($\mathbb{S}^7$)
- 2.5 million points on $\mathbb{S}^7$
- What is its structure?
- Examine dense areas

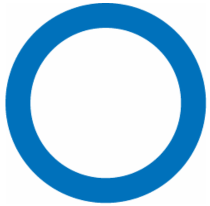


Space of Idealized Lines

- Lines in natural images
- Rasterized in 3×3 patches



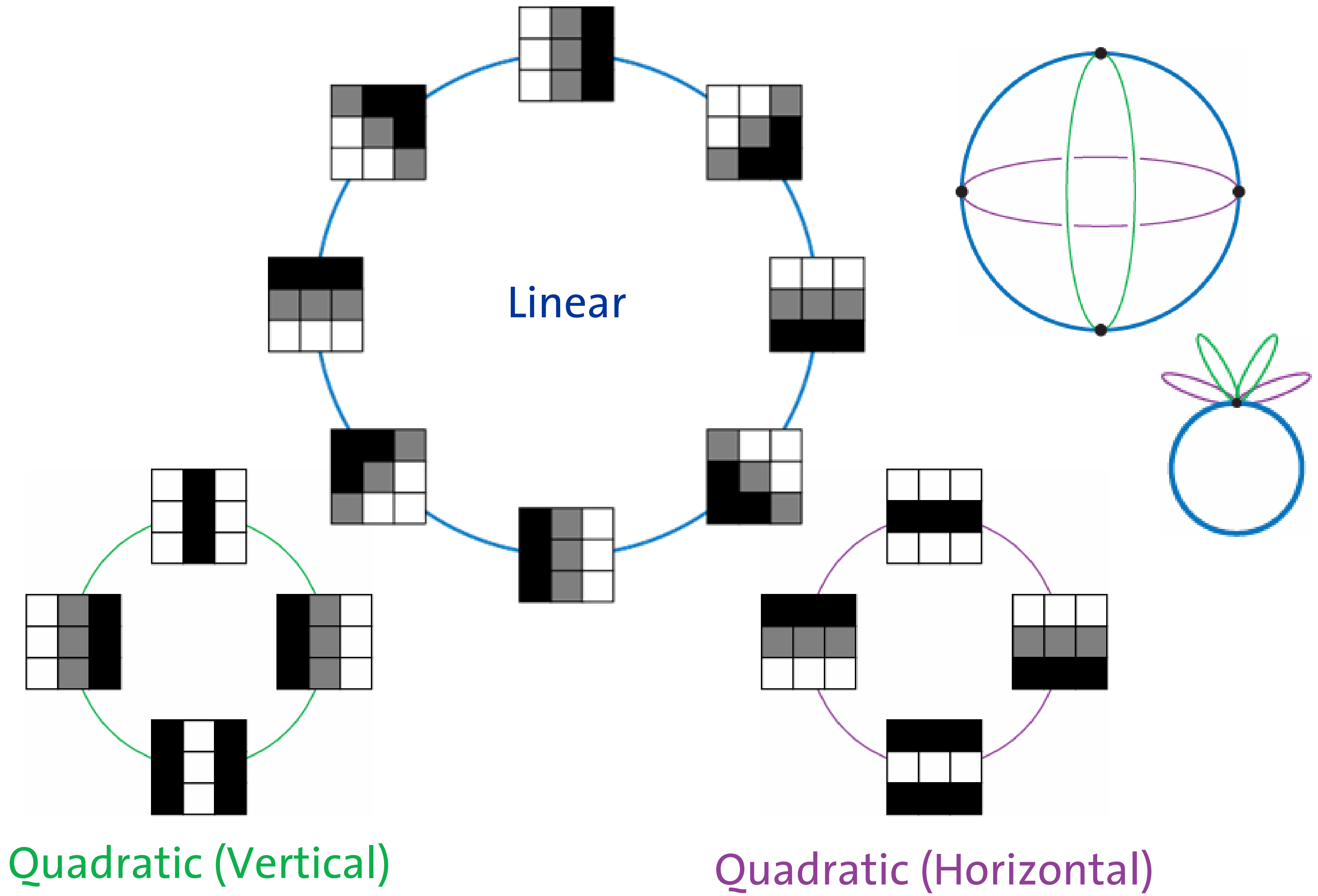
- Parameterization
  - Distance to center: $\mathbb{I}$
  - Angle: $\mathbb{S}^1$
  - Space is annulus: $\mathbb{I} \times \mathbb{S}^1$



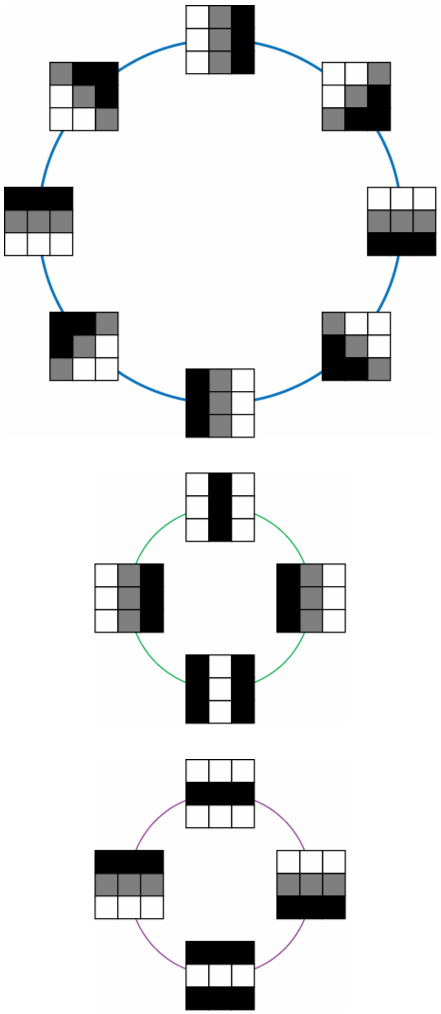
$\mathbb{I} \times \mathbb{S}^1 \simeq \mathbb{S}^1$

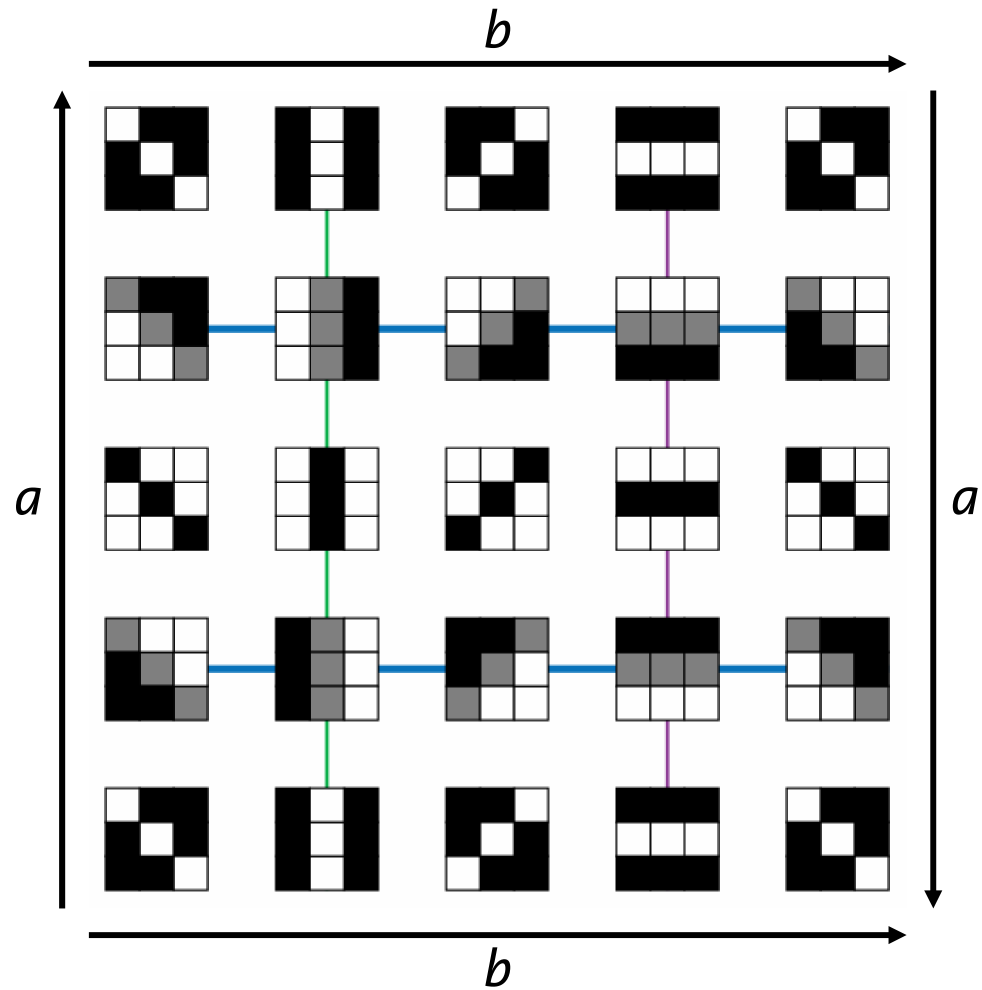
Demo

Graph Structure

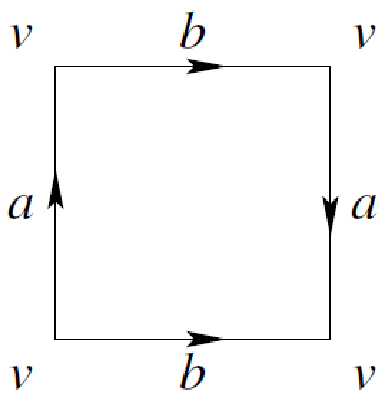


2D Structure





The Klein Bottle



- *Can we design a compression algorithm that uses the Klein bottle?*

Software

- PLEX: comptop.stanford.edu/programs/plex
  - Cech
  - Vietoris-Rips
  - Witness
  - Persistence
- CGAL: www.cgal.org
  - Alpha
  - Persistence (?)
- CHomP: chomp.rutgers.edu
- Alpha Shapes: biogeometry.duke.edu/software/alphashapes
- GGobi: www.ggobi.org

Conclusion

- We are flooded by point set *data* and need to find structure in them
- *Topology* studies connectivity of spaces
- *Topological analysis* may be viewed as generalization of clustering
- To analyze point sets, we require a *combinatorial representation* approximating the original space
- *Homology* focuses on the structure of cycles
- *Persistent homology* analyzes the relationship of structures at multiple scales