TDA: Lecture 2
Complexes and filtrations

Wesley Hamilton

University of North Carolina - Chapel Hill

wham@live.unc.edu

02/05/20

1. Complexes
   - Simplicial complexes
   - Cubical complexes
2. Filtrations
3. From data to filtrations
   - Vietoris-Rips filtrations
   - Sub/superlevel set filtrations
   - Alpha shapes

Complexes

- **Complexes** are the combinatorial building blocks used in TDA.
- The two types of complexes we'll focus on are **simplicial** complexes and **cubical** complexes.
- Simplicial complexes are easier to "construct" with, and more common in algebraic topology.
- Cubical complexes are better adapted to image/pixelated/voxel data.

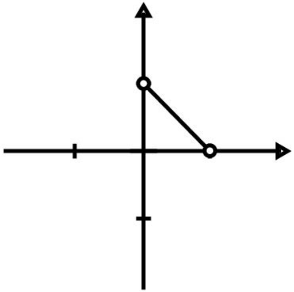
Simplices

A **geometric** n**-simplex** σ is the set

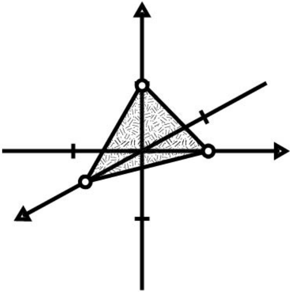
$$\sigma = \left\{ \sum_{i=0}^{n} a_i e_i \colon \sum_i a_i = 1, a_i \geq 0 \right\},$$

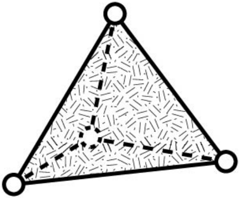
where the e_i are standard basis vectors for $\mathbb{R}^n$.

Examples of simplices, dim 0 and 1



Examples of simplices, dim 2 and 3





Simplices and barycentric coordinates

Given an n-simplex σ, we can specify points within σ using **barycentric coordinates**:

$$(a_0, \ldots, a_1) \mapsto \sum_{i=0}^{n} a_i e_i \in \sigma.$$

Faces and facets

The **face maps** ∂_j, defined on n-simplexes, are the restriction of a simplex σ's barycentric coordinates to $a_j = 0$, i.e.

$$\partial_j \sigma = \{\sum_{i=0}^{n} a_i e_i \colon \sum_i a_i = 1, a_i \geq 0, a_j = 0\}.$$

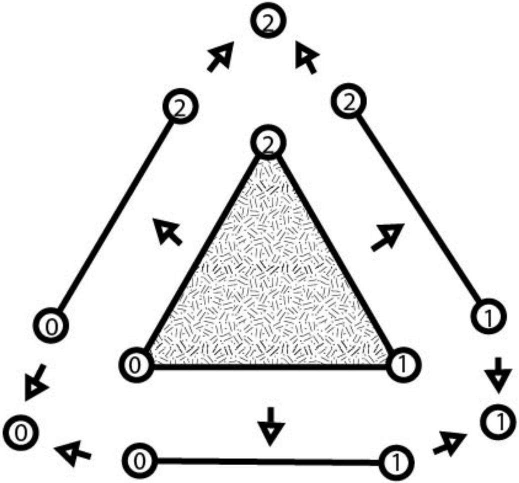
Faces and facets

The **face maps** ∂_j, defined on n-simplexes, are the restriction of a simplex σ's barycentric coordinates to $a_j = 0$, i.e.

$$\partial_j \sigma = \{\sum_{i=0}^{n} a_i e_i \colon \sum_i a_i = 1, a_i \geq 0, a_j = 0\}.$$

The jth face of a simplex σ is the set $\partial_j \sigma$.

Examples of faces

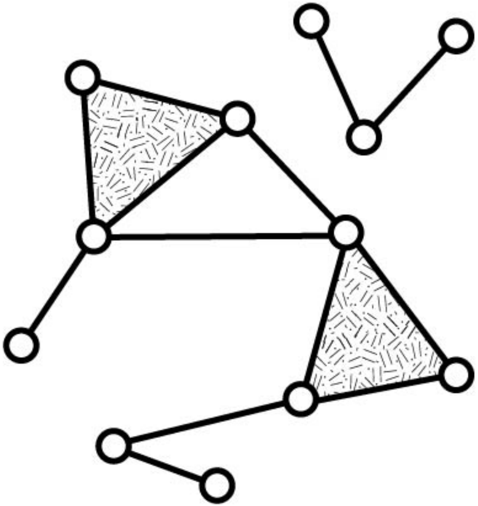


Simplicial complexes

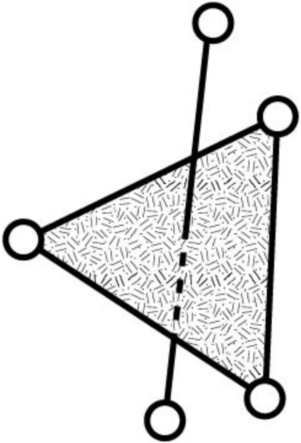
A **simplicial complex** K is a set of simplexes (any dimension) such that

1. Every face of a simplex from K is also in K, and
2. If two simplexes in K have a non-empty intersection, then said intersection is a face of each simplex.

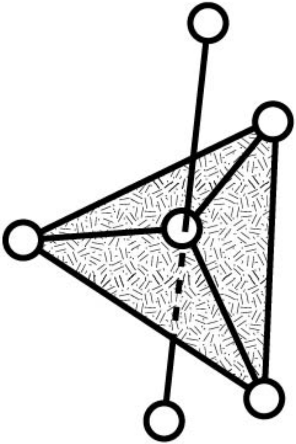
Examples of simplicial complexes



Non-examples of simplicial complexes



Non-examples of simplicial complexes



Abstract vs. geometric simplicial complexes

An **abstract simplicial complex** on a set S (such as $S = \{1, ..., n\}$) is a collection Δ of non-empty subsets of S, such that when $Y \subset X \in \Delta$, $Y \in \Delta$ as well.

Abstract vs. geometric simplicial complexes

An **abstract simplicial complex** on a set S (such as $S = \{1, ..., n\}$) is a collection Δ of non-empty subsets of S, such that when $Y \subset X \in \Delta$, $Y \in \Delta$ as well.

- The sets in Δ are the faces of the simplicial complex.
- The "intersection" property for (geometric) simplicial complexes is automatically satisfied by abstract simplicial complexes.
- Given a geometric simplicial complex, we can recover an abstract simplicial complex.

Abstract vs. geometric simplicial complexes

An **abstract simplicial complex** on a set S (such as $S = \{1, ..., n\}$) is a collection Δ of non-empty subsets of S, such that when $Y \subset X \in \Delta$, $Y \in \Delta$ as well.

- The sets in Δ are the faces of the simplicial complex.
- The "intersection" property for (geometric) simplicial complexes is automatically satisfied by abstract simplicial complexes.
- Given a geometric simplicial complex, we can recover an abstract simplicial complex. Also vice-versa.

Abstract simplicial complex non-example

- $K = \{\{0, 1, 2\}\}$

Abstract simplicial complex non-example

- $K = \{\{0, 1, 2\}\}$
- The issue is that the faces of $\{0, 1, 2\}$ aren't also in K.

Abstract simplicial complex non-example

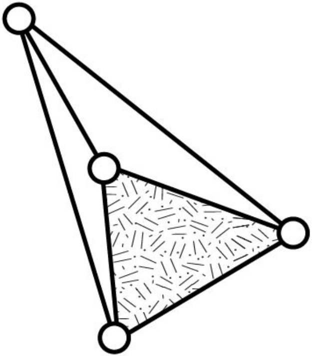
- $K = \{\{0, 1, 2\}\}$
- The issue is that the faces of $\{0, 1, 2\}$ aren't also in K.
- $K = \{\{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}$ is a "quick" fix.

Abstract simplicial complex example

$$K = \{\{0\}, \{1\}, \{2\}, \{3\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{1, 3\}, \{0, 3\}, \{2, 3\}, \{0, 2, 3\}\}$$

Abstract simplicial complex example

$$K = \{\{0\}, \{1\}, \{2\}, \{3\}, \{0,1\}, \{0,2\}, \{1,2\}, \{1,3\}, \{0,3\}, \{2,3\}, \{0,2,3\}\}$$



Cubes

- An (elementary) **interval** is an interval $I \subset \mathbb{R}$ of the form

$$I = [l, l+1], \text{ or } I = [l, l]$$

where $l \in \mathbb{Z}$.

Cubes

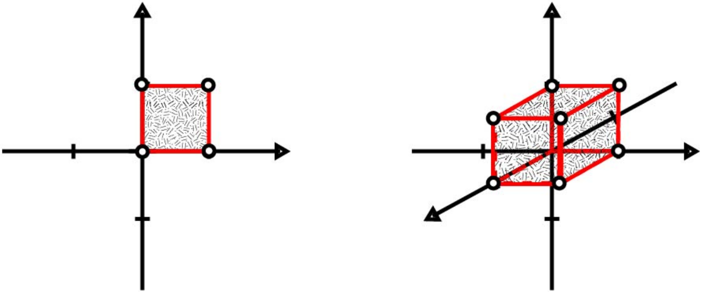
- An (elementary) **interval** is an interval $I \subset \mathbb{R}$ of the form

$$I = [l, l+1], \text{ or } I = [l, l]$$

  where $l \in \mathbb{Z}$.
- An n-**interval** is a product of n elementary intervals. An n-interval with $n > 1$ is **degenerate** if any of its factors is a singleton.

Examples of cubes, dim 0 and 1

Examples of cubes, dim 2 and 3



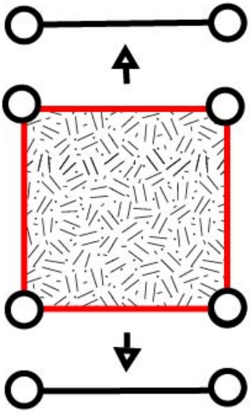
Faces and facets

- For a non-degenerate cube $I = [l_1, l_{1+1}] \times \cdots \times [l_k, l_{k+1}]$, the ith **upper and lower faces** are:
  - $\partial_i^+ I = [l_1, l_{1+1}] \times \cdots \times [l_{i+1}] \times \cdots \times [l_k, l_{k+1}]$,
  - $\partial_i^- I = [l_1, l_{1+1}] \times \cdots \times [l_i] \times \cdots \times [l_k, l_{k+1}]$.

Faces and facets

- For a non-degenerate cube $I = [l_1, l_{1+1}] \times \cdots \times [l_k, l_{k+1}]$, the ith **upper and lower faces** are:
  - $\partial_i^+ I = [l_1, l_{1+1}] \times \cdots \times [l_{i+1}] \times \cdots \times [l_k, l_{k+1}]$,
  - $\partial_i^- I = [l_1, l_{1+1}] \times \cdots \times [l_i] \times \cdots \times [l_k, l_{k+1}]$.
- For a degenerate cube $I = [l_1, l_{1+1}] \times \cdots [l_i, l_i] \times \cdots \times [l_k, l_{k+1}]$, we define the ith (upper and lower) face(s) to be empty.

Cubical face example

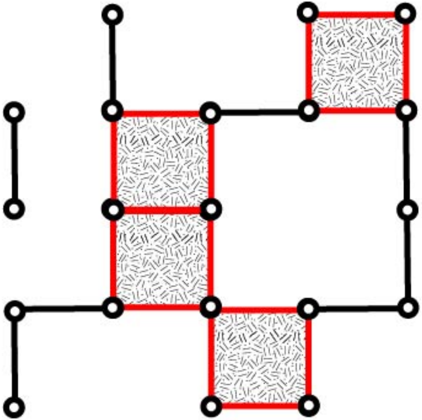


Cubical complexes

A **cubical complex** K is a set of cubes (any dimension) such that

1. Every face of a cube from K is also in K, and
2. If two cubes in K have a non-empty intersection, then said intersection is a face of each simplex.

Examples of cubical complexes



Filtrations

Suppose our (simplicial or cubical) complex comes with more information:

- each simplex is added to the complex at some time/index (dynamic formulation), or
- we have a function defined on the simplices/cubes (function formulation).

Filtrations

- We can use this extra information to **filter** our complex, and analyze pieces at different times (dynamic) or level sets (function).

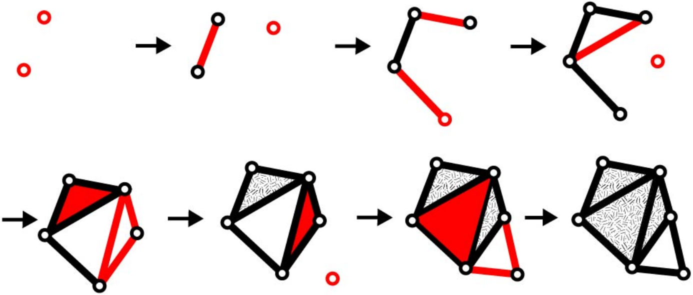
Filtrations

- We can use this extra information to **filter** our complex, and analyze pieces at different times (dynamic) or level sets (function).
- A **filtration** $\mathcal{F}$ of a complex K is a sequence of subsets $K_\alpha \subset K$, such that whenever $\alpha < \beta$, $K_\alpha \subset K_\beta$.

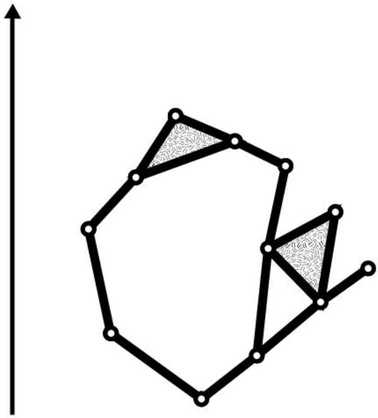
Filtrations

- We can use this extra information to **filter** our complex, and analyze pieces at different times (dynamic) or level sets (function).
- A **filtration** $\mathcal{F}$ of a complex K is a sequence of subsets $K_\alpha \subset K$, such that whenever $\alpha < \beta$, $K_\alpha \subset K_\beta$.
- Note that we haven't specified what the indices α are; they could by integers, reals, or any other object from a poset.

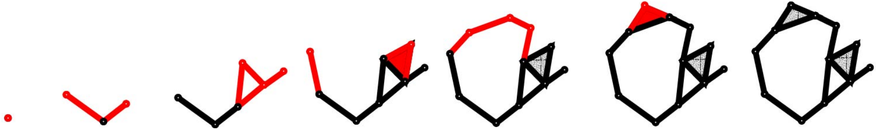
Examples of filtrations (indexed by integers)



Examples of filtrations (indexed by a function)



Examples of filtrations (indexed by a function)



Examples of filtrations

- We can also have filtrations over lattices:
  - the index set might be $\mathbb{Z} \times \mathbb{Z} = \{(m, n) \colon m, n \in \mathbb{Z}\}$,
  - the partial order is $(x_1, y_1) \leq (x_2, y_2)$ if $x_1 \leq x_2$ and $y_1 \leq y_2$.

Examples of filtrations

- We can also have filtrations over lattices:
  - the index set might be $\mathbb{Z} \times \mathbb{Z} = \{(m, n) \colon m, n \in \mathbb{Z}\}$,
  - the partial order is $(x_1, y_1) \leq (x_2, y_2)$ if $x_1 \leq x_2$ and $y_1 \leq y_2$.
- Example coming soon...

From data to filtrations

- In practice, we're given a point cloud and want to construct a complex from the points.

From data to filtrations

- In practice, we're given a point cloud and want to construct a complex from the points.
- There are many ways to do this, depending on what extra information you have:
  - Vietoris-Rips filtrations (need a metric),
  - Sub/super level set filtrations (need a function),
  - Cubical filtrations (need a voxelization and function),
  - Alpha-shape filtrations (need the points to be in $\mathbb{R}^2$ or $\mathbb{R}^3$),
  - Graph-based filtrations,
  - etc...

Vietoris-Rips complexes

- Given a finite set $S = \{x_1, ..., x_N\}$, a metric $d \colon S \times S \to [0, \infty)$, and a distance δ, the δ-**Vietoris-Rips complex** (δ-VR complex) is the abstract simplicial complex Δ_δ constructed by adding an n-simplex whenever $n+1$ points from S are pairwise within δ distance of each other.

Vietoris-Rips complexes

- Given a finite set $S = \{x_1, \ldots, x_N\}$, a metric $d\colon S \times S \to [0, \infty)$, and a distance δ, the δ-**Vietoris-Rips complex** (δ-VR complex) is the abstract simplicial complex Δ_δ constructed by adding an n-simplex whenever $n+1$ points from S are pairwise within δ distance of each other.
- I.e., $\{x_{i_1}, \ldots, x_{i_{n+1}}\} \in \Delta_\delta$ if $d(x_{i_l}, x_{i_k}) \leq \delta$ for all $1 \leq l, k \leq n+1$.

Vietoris-Rips complexes

- Given a finite set $S = \{x_1, \dots, x_N\}$, a metric $d\colon S \times S \to [0, \infty)$, and a distance δ, the δ-**Vietoris-Rips complex** (δ-VR complex) is the abstract simplicial complex Δ_δ constructed by adding an n-simplex whenever $n+1$ points from S are pairwise within δ distance of each other.
- I.e., $\{x_{i_1}, \dots, x_{i_{n+1}}\} \in \Delta_\delta$ if $d(x_{i_l}, x_{i_k}) \leq \delta$ for all $1 \leq l, k \leq n+1$.
- Can be thought of as: a disc of radius δ is placed around each x_i, and whenever the discs around x_i and x_j intersect, an edge is glued in, etc.

Vietoris-Rips filtrations

- Suppose Δ_δ is the VR complex constructed on (S, d) using distance δ. Note that $\Delta_\delta \subset \Delta_{\delta'}$ whenever $\delta \leq \delta'$.

Vietoris-Rips filtrations

- Suppose Δ_δ is the VR complex constructed on (S, d) using distance δ. Note that $\Delta_\delta \subset \Delta_{\delta'}$ whenever $\delta \leq \delta'$.
- The **Vietoris-Rips filtration** VR is the union $\cup_{\delta \in [0,\infty]} \Delta_\delta$, with the corresponding sequence of complexes $\{\Delta_\delta\}_{\delta \in [0,\infty]}$.

Vietoris-Rips filtrations

- Suppose Δ_δ is the VR complex constructed on (S, d) using distance δ. Note that $\Delta_\delta \subset \Delta_{\delta'}$ whenever $\delta \leq \delta'$.
- The **Vietoris-Rips filtration** *VR* is the union $\cup_{\delta \in [0,\infty]} \Delta_\delta$, with the corresponding sequence of complexes $\{\Delta_\delta\}_{\delta \in [0,\infty]}$.
- Note that, since $\Delta_\delta \subset \Delta_{\delta'}$ whenever $\delta \leq \delta'$,

$$\cup_{\delta \in [0,\infty]} \Delta_\delta = \Delta_\infty.$$

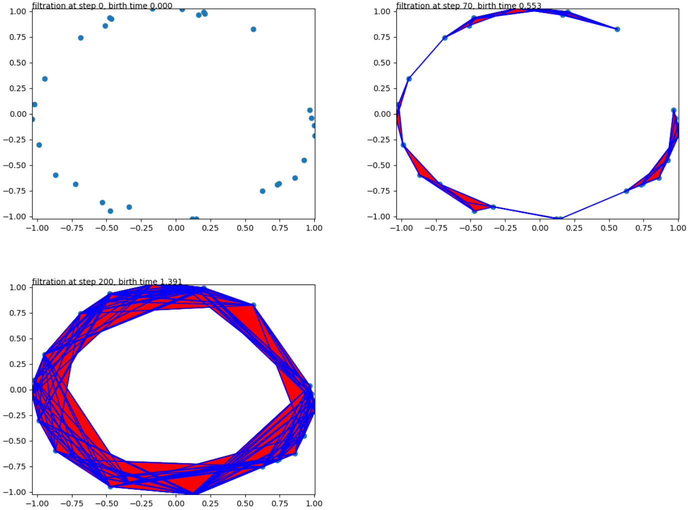
Simplicial filtration from data

- VR filtrations are the most basic to implement, and can be implemented for virtually any data set (as long as you have an underlying metric).

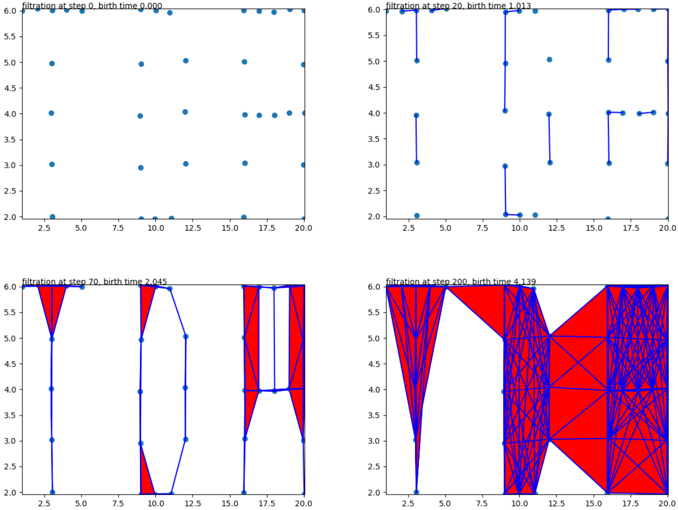
Simplicial filtration from data

- VR filtrations are the most basic to implement, and can be implemented for virtually any data set (as long as you have an underlying metric).
- Popular metrics are Euclidean (or any l^p), as long as your data is "vectorizable".

Example of VR filtration from data: noisy circle



Example of VR filtration from data: “TDA”



Example of VR filtration from data: "TDA"

The "end complex" Δ_∞ has:

- 37 0-simplices,
- 666 1-simplices,
- $7,770$ 2-simplices,
- $66,045$ 3-simplices,
- ...

Example of VR filtration from data: "TDA"

The "end complex" Δ_∞ has:

- 37 0-simplices,
- 666 1-simplices,
- $7,770$ 2-simplices,
- $66,045$ 3-simplices,
- ...
- 1 37-simplex.

Sub/superlevel sets

- Suppose we have a fixed simplicial (or cubical) complex K, and a function $f\colon K \to \mathbb{R}$. Here, we interpret the domain of f to be all possible simplices in K. We can filter K by considering sublevel sets of f, by setting $K_\delta^- = f^{-1}((-\infty, \delta])$.

Sub/superlevel sets

- Suppose we have a fixed simplicial (or cubical) complex K, and a function $f\colon K \to \mathbb{R}$. Here, we interpret the domain of f to be all possible simplices in K. We can filter K by considering sublevel sets of f, by setting $K_\delta^- = f^{-1}((-\infty, \delta])$.
- For this construction to be "well-defined", we need f to be increasing along face inclusions (otherwise an edge may appear without its boundary vertices, etc.).

Sub/superlevel sets

- Suppose we have a fixed simplicial (or cubical) complex K, and a function $f\colon K \to \mathbb{R}$. Here, we interpret the domain of f to be all possible simplices in K. We can filter K by considering sublevel sets of f, by setting $K_\delta^- = f^{-1}((-\infty, \delta])$.
- For this construction to be "well-defined", we need f to be increasing along face inclusions (otherwise an edge may appear without its boundary vertices, etc.).
- To be precise, we say f is **increasing along face inclusions** if whenever σ is a subface of τ, then $f(\sigma) \leq f(\tau)$.

Sub/superlevel sets

- Suppose we have a fixed simplicial (or cubical) complex K, and a function $f\colon K \to \mathbb{R}$. Here, we interpret the domain of f to be all possible simplices in K. We can filter K by considering sublevel sets of f, by setting $K_\delta^- = f^{-1}((-\infty, \delta])$.
- For this construction to be "well-defined", we need f to be increasing along face inclusions (otherwise an edge may appear without its boundary vertices, etc.).
- To be precise, we say f is **increasing along face inclusions** if whenever σ is a subface of τ, then $f(\sigma) \leq f(\tau)$. Likewise, we say f is **decreasing along face inclusions** if whenever σ is a subface of τ, then $f(\sigma) \geq f(\tau)$.

Sub/superlevel sets

- Suppose we have a fixed simplicial (or cubical) complex K, and a function $f\colon K \to \mathbb{R}$. Here, we interpret the domain of f to be all possible simplices in K. We can filter K by considering sublevel sets of f, by setting $K_\delta^- = f^{-1}((-\infty, \delta])$.
- For this construction to be "well-defined", we need f to be increasing along face inclusions (otherwise an edge may appear without its boundary vertices, etc.).
- To be precise, we say f is **increasing along face inclusions** if whenever σ is a subface of τ, then $f(\sigma) \leq f(\tau)$. Likewise, we say f is **decreasing along face inclusions** if whenever σ is a subface of τ, then $f(\sigma) \geq f(\tau)$.
- Symbolically, $f\colon K \to \mathbb{R}$ is increasing if whenever $\sigma \leq \tau$, then $f(\sigma) \leq f(\tau)$.

Sub/superlevel set filtrations

- If $f\colon K \to \mathbb{R}$ is increasing, then $f^{-1}((-\infty, \delta])$ will be a subcomplex of K, and moreover a subcomplex of $f^{-1}((-\infty, \delta'])$ whenever $\delta \leq \delta'$.

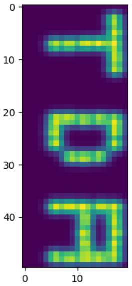
Sub/superlevel set filtrations

- If $f\colon K \to \mathbb{R}$ is increasing, then $f^{-1}((-\infty, \delta])$ will be a subcomplex of K, and moreover a subcomplex of $f^{-1}((-\infty, \delta'])$ whenever $\delta \leq \delta'$.
- Thus, we can filter K using the sequence of subcomplexes $\{f^{-1}((-\infty, \delta])\}_{\delta}$, giving rise to a **sublevel set filtration** on K.

Sub/superlevel set filtrations

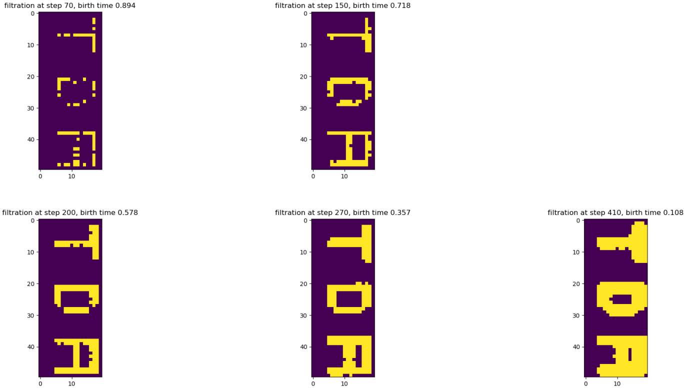
- If $f\colon K \to \mathbb{R}$ is increasing, then $f^{-1}((-\infty, \delta])$ will be a subcomplex of K, and moreover a subcomplex of $f^{-1}((-\infty, \delta'])$ whenever $\delta \leq \delta'$.
- Thus, we can filter K using the sequence of subcomplexes $\{f^{-1}((-\infty, \delta])\}_{\delta}$, giving rise to a **sublevel set filtration** on K.
- If $f\colon K \to \mathbb{R}$ is decreasing, then $f^{-1}([\delta, \infty))$ will be a subcomplex of K, and moreover a subcomplex of $f^{-1}([\delta', \infty))$ whenever $\delta' \leq \delta$.
- Thus, we can filter K using the sequence of subcomplexes $\{f^{-1}([\delta, \infty))\}_{\delta}$, giving rise to a **superlevel set filtration** on K.

Cubical filtration from data: voxelized "TDA", STD = 0.3

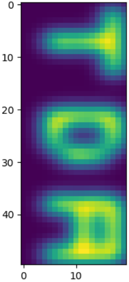


Data "voxelized" by putting a Gaussian at each point (x_T, y_T) of the "TDA" point set , and, for each grid center-point (x, y), computing $\exp(\frac{-(x_T-x)^2-(y_T-y)^2}{\sigma^2})$, $\sigma = 0.3$. We filter based on these values.

Cubical filtration from data: voxelized "TDA", STD = 0.3

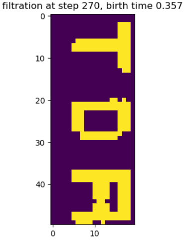


Cubical filtration from data: voxelized "TDA", STD $= 1.0$

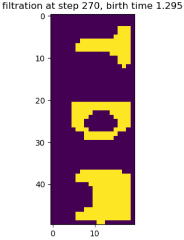


Data "voxelized" by putting a Gaussian at each point (x_T, y_T) of the "TDA" point set , and, for each grid center-point (x, y), computing $\exp(\frac{-(x_T-x)^2-(y_T-y)^2}{\sigma^2})$, $\sigma = 1.0$. We filter based on these values.

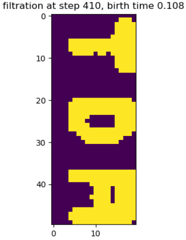
Cubical filtration from data: voxelized "TDA", STD = 1.0



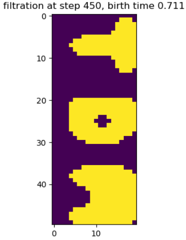
vs



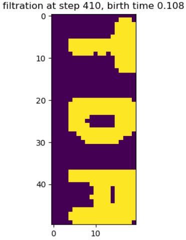
Cubical filtration from data: voxelized "TDA", STD = 1.0



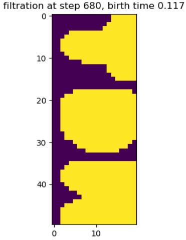
vs



Cubical filtration from data: voxelized "TDA", STD = 1.0



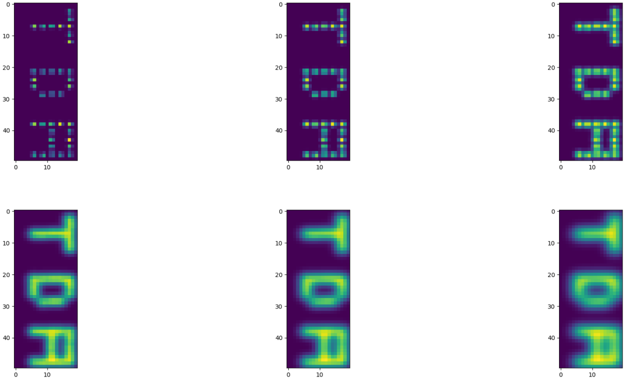
vs



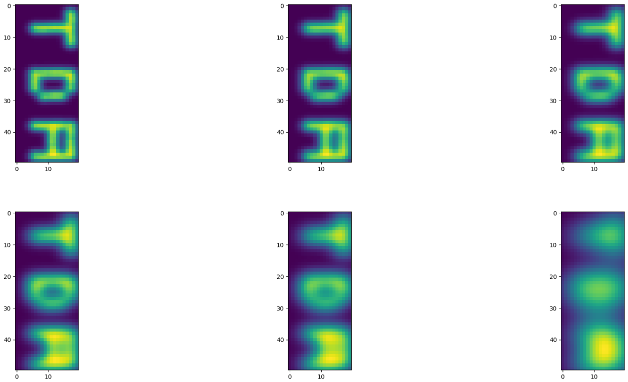
Cubical filtration from data: voxelized "TDA" 2.0

What if instead of using a superlevel set filtration, we filter by σ?

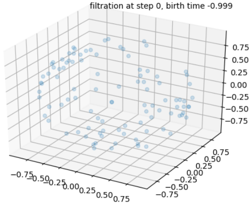
Cubical filtration from data: voxelized “TDA” 2.0

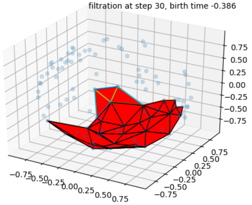


Cubical filtration from data: voxelized “TDA” 2.0

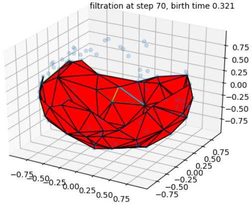


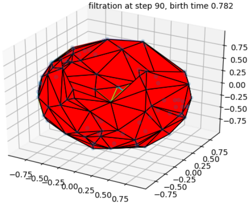
Simplicial filtration from data: sphere with height function





Simplicial filtration from data: sphere with height function



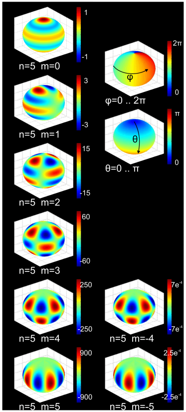


Simplicial filtration from data: sphere with spherical harmonic

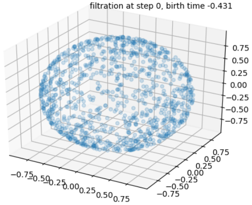
Spherical harmonics are solutions u to the PDE

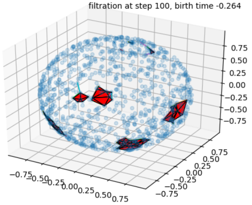
$$\partial_x^2 u + \partial_y^2 u + \partial_z^2 u = 0.$$

Simplicial filtration from data: sphere with spherical harmonic

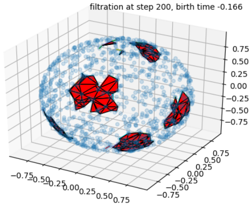


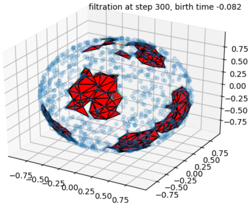
Simplicial filtration from data: sphere with spherical harmonic



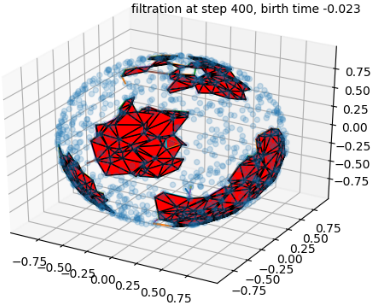


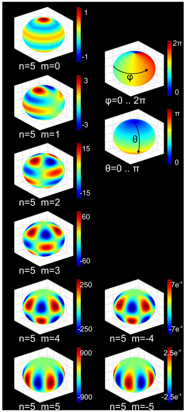
Simplicial filtration from data: sphere with spherical harmonic





Simplicial filtration from data: sphere with spherical harmonic





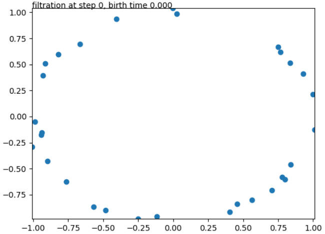
Alpha shapes

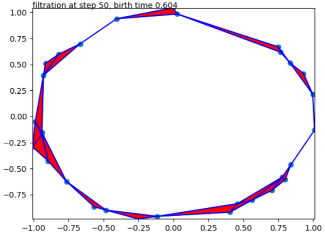
- Given a collection of points $\{x_1, \ldots, x_n\}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$, the ith **Voronoi cell** $V(x_i)$ is the set

$$V(x_i) = \{y \colon \text{dist}(x_i, y) \leq \text{dist}(x_j, y) \text{ for all } j \neq i\}.$$

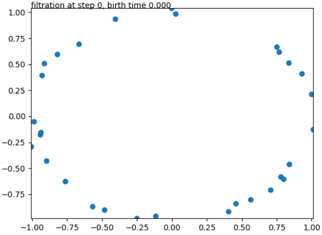
  The cells consist of points closer to the given x_i than any other x_j.
- The Voronoi cells of a collection of points tile $\mathbb{R}^2$ or $\mathbb{R}^3$.
- We construct an **alpha filtration** in the same way as we do a VR filtration, except that each growing disc is intersected with the center point's Voronoi cell.

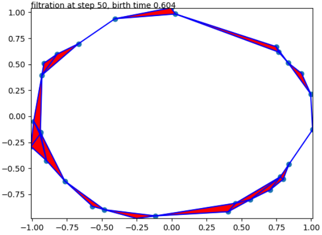
Alpha filtration example: noisy circle



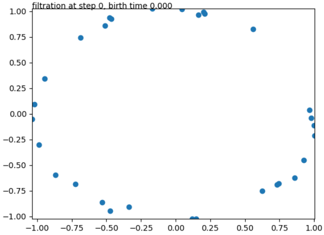


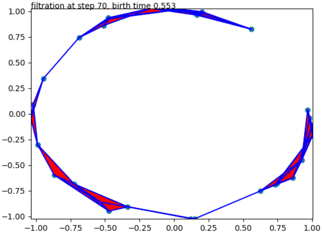
Alpha filtration example: noisy circle



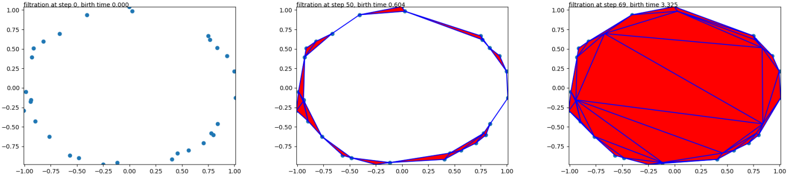


vs

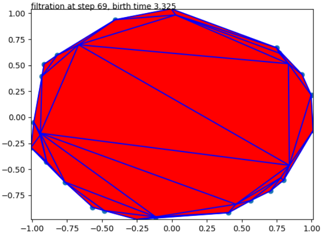




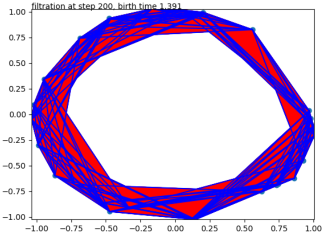
Alpha filtration example: noisy circle



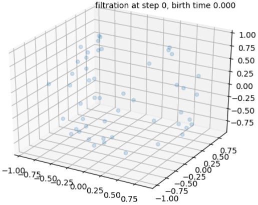
Alpha filtration example: noisy circle

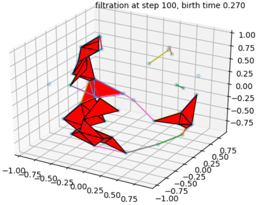


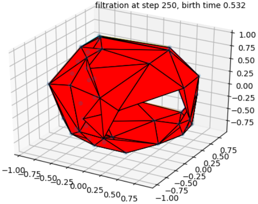
vs



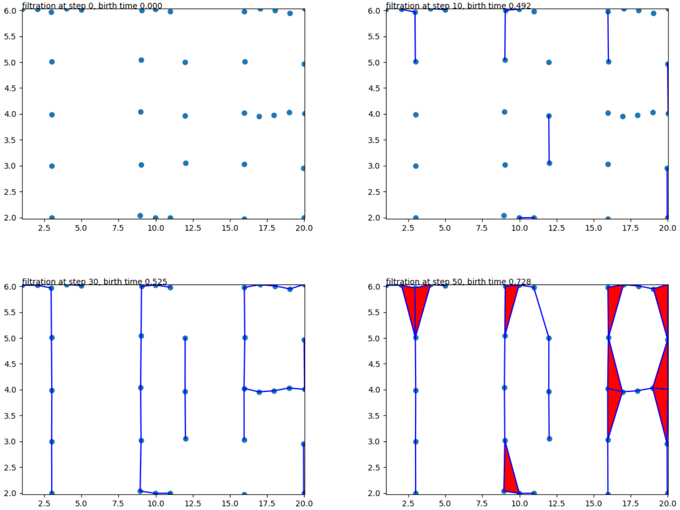
Alpha filtration example: noisy sphere



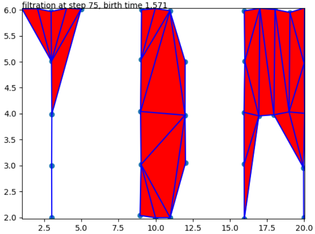


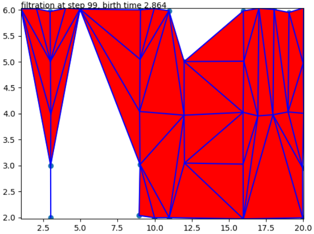


Alpha filtration example: "TDA"

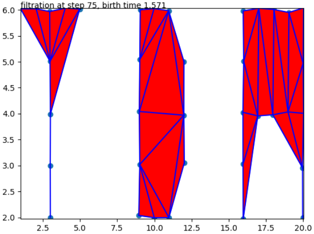


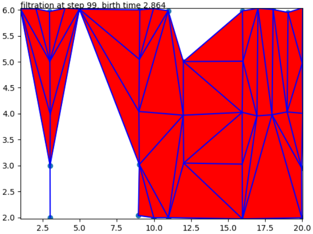
Alpha filtration example: “TDA”



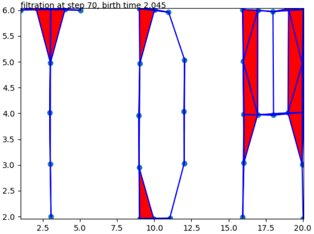


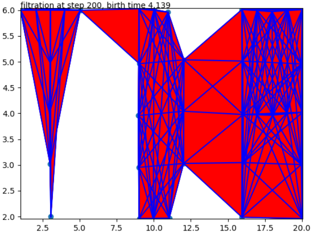
Alpha filtration example: “TDA”



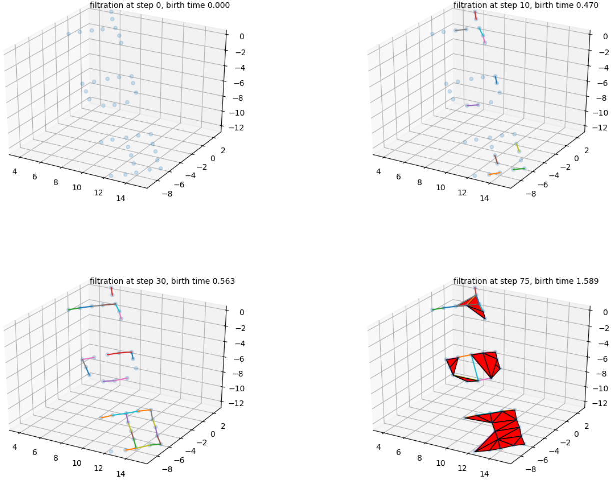


vs

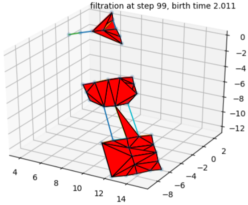


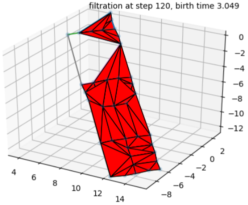


Alpha filtration example: “TDA” 3D



Alpha filtration example: "TDA" 3D





How can other kinds of data be adapted?

- For data you're interested in, discuss which kinds of filtrations/complexes might be well adapted.
- In addition (or otherwise), here are some other kinds of data:
  - Gene expression data/microarrays
  - Text documents
  - Video clips
  - Audio clips
  - Sensor networks
  - Graph based data