Some Topics in Computational Topology

Yusu Wang
Ohio State University

AMS Short Course 2014
Introduction

- Much recent developments in computational topology
  - Both in theory and in their applications
  - E.g., the theory of persistence homology
    - [Edelsbrunner, Letscher, Zomorodian, DCG 2002], [Zomorodian and Carlsson, DCG 2005], [Carlsson and de Silva, FoCM 2010], …

- This short course:
  - A computational perspective:
    - Estimation and inference of topological information / structure from point clouds data
- Develop discrete analog for the continuous case
- Approximation from discrete samples with theoretical guarantees
- Algorithmic issues
Main Topics

- From PCDs to simplicial complexes
- Sampling conditions
- Topological inferences
Outline

- From PCDs to simplicial complexes
  - Delaunay, Čech, Vietoris-Rips, witness complexes
  - Graph induced complex

- Sampling conditions
  - Local feature size, and homological feature size

- Topology inference
  - Homology inference
  - Handling noise
  - Approximating cycles of shortest basis of the first homology group
  - Approximating Reeb graph
From PCD to Simplicial Complexes

Choice of Simplicial Complexes
to build on top of point cloud data
Delaunay Complex

- Given a set of points \( P = \{ p_1, p_2, \ldots, p_n \} \subset \mathbb{R}^d \)
- Delaunay complex \( Del(P) \)
  - A simplex \( \sigma = \{ p_{i_0}, p_{i_1}, \ldots, p_{i_k} \} \) is in \( Del(P) \) if and only if
    - There exists a ball \( B \) whose boundary contains vertices of \( \sigma \), and that
      the interior of \( B \) contains no other point from \( P \).
Delaunay Complex

- Many beautiful properties
  - Connection to Voronoi diagram
- Foundation for surface reconstruction and meshing in 3D
  - [Dey, Curve and Surface Reconstruction, 2006],
  - [Cheng, Dey and Shewchuk, Delaunay Mesh Generation, 2012]

- However,
  - Computationally very expensive in high dimensions
Čech Complex

- Given a set of points $P = \{ p_1, p_2, ..., p_n \} \subset \mathbb{R}^d$
- Given a real value $r > 0$, the Čech complex $C^r(P)$ is the nerve of the set $\{ B(p_i, r) \}_{i \in [1, n]}$
  - where $B_r(p_i) = B(p_i, r) = \{ x \in \mathbb{R}^d \mid d(p_i, x) < r \}$
  - i.e., a simplex $\sigma = \{ p_{i_0}, p_{i_1}, ..., p_{i_k} \}$ is in $C^r(P)$ if
    $$\bigcap_{0 \leq j \leq k} B_r(p_{i_j}) \neq \emptyset.$$
- The definition can be extended to a finite sample $P$ of a metric space.
Rips Complex

- Given a set of points $P = \{ p_1, p_2, \ldots, p_n \} \subset \mathbb{R}^d$
- Given a real value $r > 0$, the **Vietoris-Rips (Rips) complex** $R^r(P)$ is:
  $$\{ (p_{i_0}, p_{i_1}, \ldots, p_{i_k}) \mid B_r(p_{i_l}) \cap B_r(p_{i_j}) \neq \emptyset, \forall \ l, j \in [0, k] \}.$$
Rips and Čech Complexes

- Relation in general metric spaces
  - $C^r(P) \subseteq R^r(P) \subseteq C^{2r}(P)$
  - Bounds better in Euclidean space

- Simple to compute

- Able to capture geometry and topology
  - We will make it precise shortly
  - One of the most popular choices for topology inference in recent years

- However:
  - Huge sizes
  - Computation also costly
Witness Complexes

- A simplex $\sigma = \{q_0, ..., q_k\}$ is **weakly witnessed** by a point $x$ if $d(q_i, x) \leq d(q, x)$ for any $i \in [0, k]$ and $q \in Q \setminus \{q_0, ..., q_k\}$.
- is **strongly witnessed** if in addition $d(q_i, x) = d(q_j, x)$, $\forall i, j \in [0, k]$.

Given a set of points $P = \{p_1, p_2, ..., p_n\} \subset R^d$ and a subset $Q \subseteq P$,

- The **witness complex** $W(Q, P)$ is the collection of simplices with vertices from $Q$ whose all subsimplices are weakly witnessed by a point in $P$.
- [de Silva and Carlsson, 2004] [de Silva 2003]
- Can be defined for a general metric space
- $P$ does not have to be a finite subset of points
Intuition

- $L$: landmarks from $P$, a way to subsample.
Witness Complexes

- Greatly reduce size of complex
  - Similar to Delaunay triangulation, remove redundancy

Relation to Delaunay complex

- \( W(Q, P) \subseteq Del Q \) if \( Q \subseteq P \subset R^d \)
- \( W(Q, R^d) = Del Q \)
- \( W(Q, M) = Del|_M Q \) if \( M \subseteq R^d \) is a smooth 1- or 2-manifold
  - [Attali et al, 2007]

However,

- Does not capture full topology easily for high-dimensional manifolds
Remark

- **Rips complex**
  - Capture homology when input points are sampled dense enough
  - But too large in size

- **Witness complex**
  - Use a subsampling idea
  - Reduce size tremendously
  - May not be easy to capture topology in high-dimensions

- **Combine the two?**
  - Graph induced complex
    - [Dey, Fan, Wang, SoCG 2013]
Subsampling

\[ P \quad R^\varepsilon(P) \]
Subsampling - cont

\[ P \quad Q \subseteq P \quad R^r(Q) \quad W(Q, P) \]
Subsampling - cont

\[ Q \subseteq P \quad R^r(Q) \quad W(Q,P) \]
Subsampling - cont

\( R^\varepsilon(P) \) \quad \( W(Q, P) \) \quad \( \mathcal{G}^r(Q, P) \)
Graph Induced Complex

- [Dey, Fan, Wang, SoCG 2013]
- $P$: finite set of points
- $(P, d)$: metric space
- $G(P)$: a graph

- $Q \subset P$: a subset
- $\pi(p)$: the closest point of $p \in P$ in $Q$
Graph Induced Complex

- **Graph induced complex** $G(P, Q, d): \{q_0, \ldots, q_k\} \subseteq Q$
  - if and only if there is a $(k+1)$-clique in $G(P)$ with vertices $p_0, \ldots, p_k$ such that $\pi(p_i) = q_i$, for any $i \in [0, k]$.
  - Similar to geodesic Delaunay [Oudot, Guibas, Gao, Wang, 2010]

- Graph induced complex depends on the metric $d$:
  - Euclidean metric
  - Graph based distance $d_G$
Graph Induced Complex

- Small size, but with homology inference guarantees
- In particular:
  - $H_1$ inference from a lean sample
Graph Induced Complex

- Small size, but with homology inference guarantees
- In particular:
  - $H_1$ inference from a lean sample
  - Surface reconstruction in $R^3$
  - Topological inference for compact sets in $R^d$ using persistence
Outline

- From PCDs to simplicial complexes
  - Delaunay, Čech, Vietoris-Rips, witness complexes
  - Graph induced complex

- Sampling conditions
  - Local feature size, and homological feature size

- Topology inference
  - Homology inference
    - Handling noise
Sampling Conditions
Motivation

- Theoretical guarantees are usually obtained when input points $P$ sampling the hidden domain “well enough”.

- Need to quantify the “wellness”.
- Two common ones based on:
  - Local feature size
  - Weak feature size
Distance Function

- $X \subset R^d$: a compact subset of $R^d$
- Distance function $d_X: R^d \rightarrow R^+ \cup \{0\}$
  - $d_X(x) = \min_{y \in X} d(x, y)$
  - $d_X$ is a 1-Lipschitz function
- $X^\alpha$: $\alpha$-offset of $X$
  - $X^\alpha = \{y \in R^d \mid d(y, X) \leq \alpha\}$
- Given any point $x \in R^d$
  - $\Gamma(x) := \{y \in X \mid d(x, y) = d_X(x)\}$
Medial Axis

- The *medial axis* \( \Sigma \) of \( X \) is the closure of the set of points \( x \in \mathbb{R}^d \) such that \( |\Gamma(x)| \geq 2 \)
- \( |\Gamma(x)| \geq 2 \) means that there is a medial ball \( B_r(x) \) touching \( X \) at more than 1 point and whose interior is empty of points from \( X \).
Local Feature Size

- The local feature size $lf s(x)$ at a point $x \in X$ is the distance of $x$ to the medial axis $\Sigma$ of $X$
  - That is, $lf s(x) = d(x, \Sigma)$

- This concept is adaptive
  - Large in a place without “features”

- Intuitively:
  - We should sample more densely if local feature size is small.

- The **reach** $\rho(X) = \inf_{x \in X} lf s(x)$

Courtesy of [Dey, 2006]
Gradient of Distance Function

- Distance function not differentiable on the medial axis
- Still can define a generalized concept of gradient
  - [Lieutier, 2004]
- For $x \in \mathbb{R}^d \setminus X$,
  - Let $c_X(x)$ and $r_X(x)$ be the center and radius of the smallest enclosing ball of point(s) in $\Gamma(x)$
  - The \textit{generalized gradient} of distance function
    $\nabla_X(x) = \frac{x - c_X(x)}{r_X(x)}$
- Flow lines induced by the generalized gradient
- Examples:
Critical Points

- A critical point of the distance function is a point whose generalized gradient $\nabla_X(x)$ vanishes.

- A critical point is either in $X$ or in its medial axis $\Sigma$. 
Weak Feature Size

- Given a compact $X \subset \mathbb{R}^d$, let $C \subset \mathbb{R}^d$ denote
  - the set of critical points of the distance function $d_X$ that are not in $X$

- Given a compact $X \subset \mathbb{R}^d$, the \textit{weak feature size} is
  - $\text{wfs}(X) = \inf_{x \in X} d(x, C)$

- Equivalently,
  - $\text{wfs}(X)$ is the infimum of the positive critical value of $d_X$

- $\rho(X) \leq \text{wfs}(X)$
Why Distance Field?

- Theorem [Offset Homotopy] [Grove’93]
  If $0 < \alpha < \alpha'$ are such that there is no critical value of $d_X$ in the closed interval $[\alpha, \alpha']$, then $X^{\alpha'}$ deformation retracts onto $X^\alpha$. In particular, $H(X^\alpha) \cong H(X^{\alpha'})$.

- Remarks:
  - For the case of compact set $X$, note that it is possible that $X^\alpha$, for sufficiently small $\alpha > 0$, may not be homotopy equivalent to $X^0 = X$.
  - Intuitively, by above theorem, we can approximate $H(X^\alpha)$ for any small positive $\alpha$ from a thickened version (offset) of $X^\alpha$.
  - The sampling condition makes sure that the discrete sample is sufficient to recover the offset homology.
Typical Sampling Conditions

- Hausdorff distance $d_H(A, B)$ between two sets $A$ and $B$
  - infinumum value $\alpha$ such that $A \subseteq B^\alpha$ and $B \subseteq A^\alpha$

- No noise version:
  - A set of points $P$ is an $\epsilon$-sample of $X$
    if $P \subseteq X$ and $d_H(P, X) \leq \epsilon$

- With noise version:
  - A set of points $P$ is an $\epsilon$-sample of $X$ if $d_H(P, X) \leq \epsilon$

- Theoretical guarantees will be achieved when $\epsilon$ is small with respect to local feature size or weak feature size
Outline

- From PCDs to simplicial complexes
  - Delaunay, Čech, Vietoris-Rips, witness complexes
  - Graph induced complex

- Sampling conditions
  - Local feature size, and homological feature size

- Topology inference
  - Homology inference
  - Handling noise
Homology Inference from PCD
Problem Setup

- A hidden compact $X$ (or a manifold $M$)
- An $\epsilon$-sample $P$ of $X$
- Recover homology of $X$ from some complex built on $P$
  - will focus on Čech complex and Rips complex
Union of Balls

- $X^\alpha = \bigcup_{x \in X} B(x, \alpha)$
- $P^\alpha = \bigcup_{p \in P} B(p, \alpha)$

- Intuitively, $P^\alpha$ approximates offset $X^\alpha$

- The Čech complex $C^\alpha(P)$ is the Nerve of $P^\alpha$

- By Nerve Lemma, $C^\alpha(P)$ is homotopy equivalent to $P^\alpha$
Smooth Manifold Case

- Let $X$ be a smooth manifold embedded in $\mathbb{R}^d$

**Theorem [Niyogi, Smale, Weinberger]**

Let $P \subset X$ be such that $d_H(X, P) \leq \epsilon$. If $2\epsilon \leq \alpha \leq \sqrt{\frac{3}{5}} \rho(X)$, there is a deformation retraction from $P^\alpha$ to $X$.

**Corollary A**

Under the conditions above, we have

$$H(X) \cong H(P^\alpha) \cong H(C^\alpha(P)).$$
How about using Rips complex instead of Čech complex?

Recall that

\[ C^r(P) \subseteq R^r(P) \subseteq C^{2r}(P) \]

inducing

\[ H(C^r) \to H(R^r) \to H(C^{2r}) \]

Idea [Chazal and Oudot 2008]:

- Forming interleaving sequence of homomorphism to connect them with the homology of the input manifold \( X \) and its offsets \( X^\alpha \)
Convert to Cech Complexes

- Lemma A [Chazal and Oudot, 2008]:
  - The following diagram commutes:
    \[
    \begin{array}{ccc}
    H(P^\alpha) & \xrightarrow{i_*} & H(P^\beta) \\
    h_* & & h_* \\
    H(C^\alpha) & \xrightarrow{i_*} & H(C^\beta)
    \end{array}
    \]

- Corollary B
  - Let \( P \subset X \) be s.t. \( d_H(X, P) \leq \epsilon \). If \( 2\epsilon \leq \alpha \leq \alpha' \leq \sqrt{\frac{3}{5}} \rho(X) \),
    \[ H(X) \cong H(C^\alpha) \cong H(C^\beta) \]
    where the second isomorphism is induced by inclusion.
Lemma B:

Given a sequence $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$ of homomorphisms between finite dimensional vector spaces, if $\text{rank}(A \rightarrow \ldots)$

Rips and Čech complexes:

$$C^\alpha(P) \subseteq R^\alpha(P) \subseteq C^{2\alpha}(P) \subseteq R^{2\alpha}(P) \subseteq C^{4\alpha}(P)$$

$$\Rightarrow H(C^\alpha) \rightarrow H(R^\alpha) \rightarrow H(C^{2\alpha}) \rightarrow H(R^{2\alpha}) \rightarrow H(C^{4\alpha})$$

Applying Lemma B

$$\text{rank}(H(R^\alpha) \rightarrow H(R^{2\alpha})) = \text{rank}(H(C^{2\alpha})) = \text{rank}(H(X))$$
The Case of Compact

- In contrast to Corollary A, now we have the following (using Lemma B).

- **Lemma C [Chazal and Oudot 2008]:**
  Let $P \subset \mathbb{R}^d$ be a finite set such that $d_H(X, P) < \epsilon$ for some $\epsilon < \frac{1}{4} wfs(X)$. Then for all $\alpha, \beta \in [\epsilon, wfs(X) - \epsilon]$ such that $\beta - \alpha \geq 2\epsilon$, and for all $\lambda \in (0, wfs(X))$, we have $H(X^\lambda) \cong \text{image}(i_*)$, where $i_* : H(P^\alpha) \to H(P^\beta)$ is the homomorphism between homology groups induced by the canonical inclusion $i : P^\alpha \to P^\beta$. 
The Case of Compacts

- One more level of interleaving.
- Use the following extension of Lemma B:

  Given a sequence \( A \to B \to C \to D \to E \to F \) of homomorphisms between finite dimensional vector spaces, if \( \text{rank}(A \to F) = \text{rank}(C \to D) \), then \( \text{rank}(B \to E) = \text{rank}(C \to D) \).

Theorem [Homology Inference] [Chazal and Oudot 2008]:

Let \( P \subset \mathbb{R}^d \) be a finite set such that \( d_H(X, P) < \epsilon \) for some \( \epsilon < \frac{1}{9} \text{wfs}(X) \). Then for all \( \alpha \in \left[2\epsilon, \frac{1}{4}(\text{wfs}(X) - \epsilon)\right] \) all \( \lambda \in (0, \text{wfs}(X)) \), we have \( H(X^\lambda) \cong \text{image}(j_*) \), where \( j_* \) is the homomorphism between homology groups induced by canonical inclusion \( j: \mathbb{R}^\alpha \to \mathbb{R}^{4\alpha} \).
Theorem [Homology Inference] [Chazal and Oudot 2008]:

Let $P \subset \mathbb{R}^d$ be a finite set such that $d_H(X, P) < \epsilon$ for some $\epsilon < \frac{1}{9} \text{wfs}(X)$. Then for all $\alpha \in \left[2\epsilon, \frac{1}{4} (\text{wfs}(X) - \epsilon)\right]$ all $\lambda \in (0, \text{wfs}(X))$, we have $H(X^\lambda) \cong \text{image}(j_*)$, where $j_*$ is the homomorphism between homology groups induced by canonical inclusion $j: \mathbb{R}^\alpha \rightarrow \mathbb{R}^{4\alpha}$. 
Summary of Homology Inference

- $X^\alpha$ homotopy equivalent to $X^\beta$
  - Critical points of distance field
- $P^\alpha$ approximates $X^\alpha$ (may be interleaving)
  - E.g., [Niyogi, Smale, Weinberger, 2006], [Chazal and Oudot 2008]
- $C^\alpha$ homotopy equivalent to $P^\alpha$
  - Nerve Lemma
- $H(C^\alpha)$ interleaves $H(X^\alpha)$ at homology level
  - [Chazal and Oudot 2008]
- $R^\alpha$ and $C^\alpha$ interleave
- Derive homology inference from the interleaving sequence of homomorphisms
Outline

- From PCDs to simplicial complexes
  - Delaunay, Čech, Vietoris-Rips, witness complexes
  - Graph induced complex

- Sampling conditions
  - Local feature size, and homological feature size

- Topology inference
  - Homology inference
  - Handling noise
References


References – cont.


References – cont.


