Complexes

6 Simplicial complex

A complex is a collection of some basic elements which satisfy certain properties. In a simplicial complex, these basic elements are simplices.

Definition 20 (simplex)**.** A k-*simplex* σ is the convex hull of a set P of $k+1$ affinely independent points. In particular, a 0-simplex is a *vertex*, a 1-simplex is an *edge*, a 2-simplex is a *triangle*, and a 3-simplex is a *tetrahedron*. A k-simplex is said to have *dimension* k. A *face* of σ is a simplex that is the convex hull of a nonempty subset of P. Faces of σ come in all dimensions from zero (σ's vertices) to k; σ is a face of σ. A *proper face* of σ is a simplex that is the convex hull of a proper subset of P; i.e. any face except σ. In particular, the $(k-1)$-faces of σ are called *facets* of σ; σ has $k+1$ facets. For instance, the facets of a tetrahedron are its four triangular faces.

Definition 21 (simplicial complex)**.** A *simplicial complex* $\mathcal{K}$, also known as a *triangulation*, is a set containing finitely[2] many simplices that satisfies the following two restrictions.

- $\mathcal{K}$ contains every face of every simplex in $\mathcal{K}$.
- For any two simplices $\sigma, \tau \in \mathcal{K}$, their intersection $\sigma \cap \tau$ is either empty or a face of both σ and τ.

Definition 22 (underlying space)**.** The *underlying space* of a complex $\mathcal{K}$, denoted $|\mathcal{K}|$, is the pointwise union of its cells; that is, $|\mathcal{K}| = \bigcup_{\sigma \in \mathcal{K}} \sigma$.

The above definition is very geometric which is why sometimes they are referred to as geometric simplicial complexes. There is a parallel notion of simplicial complexes that is devoid of geometry.

Definition 23 (abstract simplicial complex)**.** A collection $\mathcal{A}$ of subsets of a given set A is an abstract simplicial complex if every element $\sigma \in \mathcal{A}$ has all of its subsets $\sigma' \subseteq \sigma$ also in $\mathcal{A}$. The elements of A are the vertices of $\mathcal{A}$. Each (sub)set in $\mathcal{A}$ is a simplex whose dimension equals its cardinality.

An abstract simplicial complex $\mathcal{A}$ with m vertices can be embedded (geometrically realized) in $\mathbb{R}^{m-1}$ as a subcomplex of a geometric m-simplex. Thus, we define its underlying space as the underlying space of its geometric realization.

Definition 24 (k-skeleton)**.** The k-skeleton of a simplicial complex is the subcomplex formed by all of its k-dimensional simplices and their faces.

---

[2]Topologists usually define complexes so they have countable cardinality. We restrict complexes to finite cardinality here.

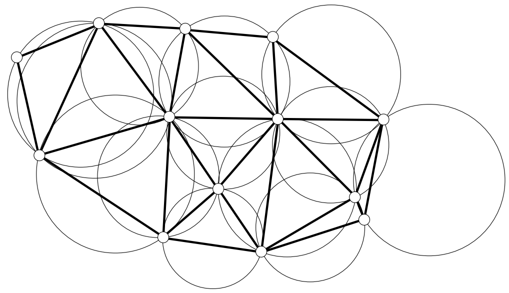

Figure 8: Every triangle in a Delaunay triangulation has an empty open circumdisk.

7 Delaunay complex

This is a special complex that can be constructed out of a point set $P \in \mathbb{R}^d$. This complex embeds in $\mathbb{R}^d$.

Definition 25 (Delaunay simplex; complex)**.** In the context of a finite point set $P \in \mathbb{R}^d$, a k-simplex σ is *Delaunay* if its vertices are in P and there is an open d-ball whose boundary contains its vertices and is *empty*—contains no point in P. Note that any number of points in P can lie on the boundary of this ball. But, for simplicity, we will assume that only the vertices of σ are on the boundary of its empty ball. A *Delaunay complex* of P, denoted Del P, is a simplicial complex with vertices in P in which every simplex is Delaunay and $|\mathrm{Del}\, P|$ coincides with the convex hull of P, as illustrated in Figure 9.

In 3d, a Delaunay complex of a set of points in general position is made out of Delaunay tetrahedra and all of its lower dimensional faces.

Fact 1. *Every non-degenerate point set (no $d+2$ points are co-spherical) admits a unique Delaunay complex.*

Delaunay complexes are dual to the famous Voronoi diagrams which we will touch upon later.

8 Čech and Rips complex

Given a point set P with a metric, that is, pairwise distances in P are known, we can build an abstract simplicial complex with vertices in P which respects the metric.

Definition 26 (metric space)**.** A metric space is a pair (M, d) where M is a set and d is a distance function $d : M \times M \to \mathbb{R}$ satisfying the following properties:

1. $d(x, y) \geq 0 \;\forall (x, y) \in M \times M$
2. $d(x, y) = 0$ if and only if $x = y$
3. $d(x, y) = d(y, x) \;\; \forall (x, y) \in M \times M$
4. $d(x, y) \leq d(x, z) + d(z, y) \;\forall (x, y, z) \in M \times M \times M$

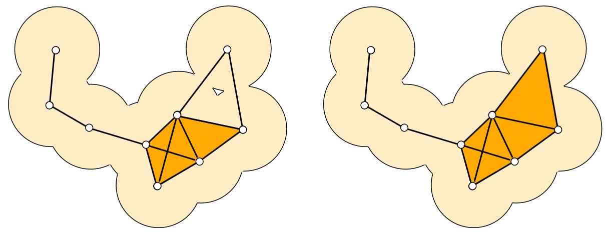

Figure 9: Čech complex $\mathcal{C}^r(P)$ and Rips complex $\mathcal{R}^r(P)$

An open geodesic ball of radius r centering a point $p \in M$ is the set $B(p, r) = \{x \in M \mid d(p, x) < r\}$.

Definition 27 (Vietoris-Rips complex)**.** Let (P, d) be a metric space where P is a point set. Given a real $r > 0$, the Vietoris-Rips (Rips in short) complex is the abstract simplicial complex $\mathcal{R}^r(P)$ where a simplex $\sigma \in \mathcal{R}^r(P)$ if and only if $d(p, q) \leq r$ for every pair of vertices of σ.

Notice that the 1-skeleton of $\mathcal{R}^r(P)$ determines all of its simplices. It is the completion (in terms of simplices) of its 1-skeleton.

The Rips complex is related to another complex called Čech complex which is often used in topological data analysis.

Definition 28 (nerve)**.** Let M be a topological space. Let $\mathcal{M}$ be a set of subsets of M. The nerve of $\mathcal{M}$ is the abstract simplicial complex $\mathcal{K}$ defined on the set $\mathcal{M}$ where a simplex $\{c_1, \ldots, c_k\} \subseteq \mathcal{M}$ is in $\mathcal{K}$ if

$$\cap_{i=1}^{k} c_i \neq \emptyset.$$

Definition 29. Let (M, d) be a metric space with a topology induced by its metric. Let P be a subset of M. Given a real $r > 0$, the Čech complex $\mathcal{C}^r(P)$ is defined to be the nerve of the set $\{B(p, r/2) \mid p \in P\}$ where

$$B(p, r/2) = \{x \in M \mid d(p, x) < r/2\}$$

is the metric open ball of radius $r/2$ centering p.

An easy but important observation is that the Rips and Čech complexes interleave.

Lemma 1. *Let P be a subset of a metric space (M, d). Then,*

$$\mathcal{C}^r(P) \subseteq \mathcal{R}^r(P) \subseteq \mathcal{C}^{2r}(P).$$

P . The first inclusion is obvious because if there is a point x in the intersection $\cap_{i=1}^{k} B(p_i, r/2)$, the distances $d(p_i, p_j)$ for every pair (i, j), $1 \leq i, j \leq k$, are at most r. It follows that for every simplex $\{p_1, \ldots, p_k\} \in \mathcal{C}^r(P)$ is also in $\mathcal{R}^r(P)$.

To prove the second inclusion, consider a simplex $\{p_1, \ldots, p_k\} \in \mathcal{R}^r(P)$. Since by definition of the Rips complex $d(p_i, p_1) \leq r$ for every p_i, $i = 1, \ldots, k$, we have $\cap_{i=1}^{k} B(p_i, r) \supset p_1 \neq \emptyset$. Then, by definition, $\{p_1, \ldots, p_k\}$ is also a simplex in $\mathcal{C}^{2r}(P)$. □

9 Witness complex

The Rips and Čech complexes are often too large to handle. For example, the Rips complex with n points in $\mathbb{R}^d$ can have $\Omega(n^d)$ simplices. In practice, also they become too large to handle even in dimension as low as three. Just to have a sense of the scale of the problem, we note that the Rips or Čech complex built out of a few thousand points often has triangles in the range of millions. The witness complex defined by de Silva and Carlsson [5] sidesteps this problem by a subsampling strategy. Given a point sample P from a metric space, we subsample P with a subset $Q \subseteq P$ and then build a complex on Q instead of P.

Definition 30. Let Q be a finite subset of a metric space (M, d). A simplex $\sigma = \{q_1, \ldots, q_k\}$ is weakly witnessed by $x \in M$ if $d(q, x) \leq d(p, x)$ for every $q \in \{q_1, \ldots, q_k\}$ and $p \in Q \setminus \{q_1, \ldots, q_k\}$. The simplex σ is strongly witnessed if, additionally, $d(q_1, x) = \cdots = d(q_k, x)$.

The following fact is proved in [6].

Proposition 2. *A simplex σ is strongly witnessed if and only if every subsimplex $\tau \leq \sigma$ is weakly witnessed.*

We now define the witness complex using the notion of weak witnesses.

Definition 31. Let P be a finite subset of a metric space (M, d). For a subset $Q \subseteq M$, consider all simplices with vertex set in Q that are strongly witnessed by a point in P. The witness complex $\mathcal{W}(Q, P)$ is defined as the collection of all these simplices.

Observe that a simplex which is weakly witnessed may not have all its subsimplices weakly witnessed (think about an example). This is why the definition above forces the condition of strong witness. A different complex that can be built is by considering all weakly witnessed simplices and their faces. This has not been proposed in the original definition of [**?**].

When the metric space is a Euclidean space $(\mathbb{R}^k, d)$, we have some connections of the witness complex to the Dealunay complex. By definition, we know the following:

Fact 2. *Let Q be a finite subset of $(\mathbb{R}^k, d)$. Then a simplex σ is in the Delaunay triangulation* Del Q *if and only if σ is strongly witnessed by a point in $\mathbb{R}^k$.*

By combining the above fact and the observation that every simplex in a witness complex is strongly witnessed, we have the following result which was observed by de Silva [6].

Proposition 3. *If P is a finite subset of $(\mathbb{R}^k, d)$ and $Q \subseteq P$, then $\mathcal{W}(Q, P) \subseteq$* Del Q.

One important implication of the above observation is that the witness complexes for point samples in an Euclidean space are embedded in that space.

The concept of the witness complex has a parallel in the concept of the restricted Delaunay triangulation. When the set P in Proposition 3 is not necessarily a finite subset, but only a subset of $\mathbb{R}^k$, and Q is a finite point set, what can we say about $\mathcal{W}(Q, P)$?

Proposition 4.

1. $\mathcal{W}(Q, \mathbb{R}^k) = \mathrm{Del}|_{\mathbb{R}^k}\, Q := \mathrm{Del}\, Q$ *[6].*

2. $\mathcal{W}(Q, M) = \mathrm{Del}|_M\, Q$ *if* $M \subseteq \mathbb{R}^k$ *is a smooth* 1- *or* 2-*manifold [1].*

3. $\mathcal{W}(Q, P) = \mathrm{Del}|_M\, Q$ *where* P *and* Q *are sufficiently dense sample of a* 1-*manifold* M *in* $\mathbb{R}^2$ *and the result does not extend to other cases of submanifolds embedded in Euclidean spaces [4].*

10 Graph induced complex

The witness complex does not capture the topology of a manifold even if the input sample is dense except for smooth curves in the plane. One can modify them with extra structures such as putting weights on the points and changing the metric to weighted distances to tackle this problem as shown in [2]. But, this becomes clumsy in terms of the 'practicality' of a solution. We study another complex called *graph induced complex*(GIC) introduced by Dey, Fan, and Wang [3] which also works on the notion of subsampling, but is more powerful in capturing topology and in some case geometry. The adavantage of GIC over the witness complex can be attributed to the fact that GIC is not necessarily a subcomplex of the Delaunay complex and hence contains few more simplices which aid topology inference. But, for the same reason, it may not embed in the Euclidean space where its input vertices lie.

Definition 32. Let (P, d) be a metric space where P is a finite set and $G(P)$ be a graph with vertices in P. Let $Q \subseteq P$ be a subset where $\nu : P \to Q$ be the vertex map given by $\nu(p) = argmin\ d(p, Q)$. The graph induced complex $\mathcal{G}(G(P), Q)$ is the simplicial complex containing a k-simplex $\sigma = \{q_1, \ldots, q_{k+1}\}$ if and only if there exists a $(k+1)$-clique $\{p_1, \ldots, p_{k+1}\} \subseteq P$ so that $\nu(p_i) = q_i$ for each $i \in \{1, 2, \ldots, k+1\}$. To see that it is indeed a simplicial complex, observe that a subset of a clique is also a clique.

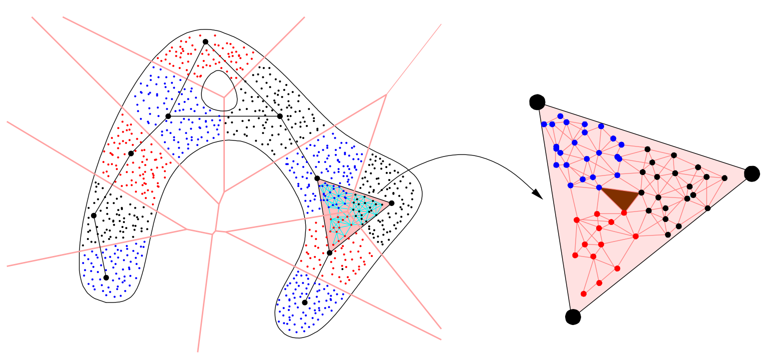

Figure 10: A graph induced complex shown with bold vertices, edges, and a shaded triangle on left. The input graph within the shaded triangle is shown on right. The three differently colored vertices of the input graph (shown inside the shaded triangle on right) cause the shaded traingle to be in the graph induced complex.

Given a neighborhood graph on a point data P equipped with a metric, one can build a graph induced complex on a subsample $Q \subseteq P$ by throwing in a simplex with a vertex set $V \subseteq Q$ if a set of points in P, each being closest to exactly one vertex in V, forms a clique. Figure 10 shows a graph induced complex for a point data in the plane where d is the Euclidean metric.

Input graph $G(P)$. The input point set P can be a finite sample of a subset X of an Euclidean space, such as a manifold or a compact subset. In this case, we will consider the input graph $G(P)$ to be the neighborhood graph $G^\alpha(P) := (P, E)$ where there is an edge $\{p, q\} \in E$ if and only if $\|p - q\| \leq \alpha$. The intuion is that if P is a sufficiently dense sample of X, then $G^\alpha(P)$ captures the local neighborhoods of the points in X. To emphasize the dependence on α we will use the notation $\mathcal{G}^\alpha(P, Q) := \mathcal{G}(G^\alpha(P), Q)$.

Subsample Q. Of course, the ability of capturing the topology of the sampled space after subsampling with Q depends on the quality of Q. We will quantify this quality with a parameter $\delta > 0$.

Definition 33. A subset $Q \subseteq P$ is called a δ-sample of a finite set P if the following condition holds:

- $\forall p \in P$, there exists a $q \in Q$, so that $d(p, q) \leq \delta$.

Q is called δ-sparse if the following condition holds:

- $\forall (q, r) \in Q \times Q,\ d(q, r) \geq \delta$.

The first condition ensures Q to be a good sample of P with respect to the parameter δ and the second condition enforces that the points in Q cannot be too close relative to the distance δ.

Metric d. The metric d assumed in the metric space (P, d) will be of two types, (i) the Euclidean metric denoted d_E, (ii) the graph metric d_G derived from the the input graph $G(P)$ where $d_G(p, q)$ is the shrtest path distance between p and q in the graph $G(P)$ assuming its edges have non-negative weights such as their Euclidean lengths.

When equipped with appropriate metric, the GIC can decipher the topology from data. It retains the simplicity of the Rips complex as well as the sparsity of the witness complex. It does not build a Rips complex on the subsample and thus is sparser than the Rips complex with the same set of vertices. This fact makes a real difference in practice as experiments show.

Figure 11 shows experimental results on two data sets, 40,000 sample points from a Klein bottle in $\mathbb{R}^4$ and 15,000 sample points from the primary circle of natural image data considered in $\mathbb{R}^{25}$. The graphs connecting any two points within $\alpha = 0.05$ unit distance for Klein bottle and $\alpha = 0.6$ unit distance for the primary circle were taken as input for the graph induced complexes. The 2-skeleton of the Rips complexes for these α parameters have $608, 200$ and $1, 329, 672, 867$ simplices respectively. These sizes are too large to carry out fast computations.

For comparisons, we constructed the graph induced complex, sparsified Rips complex, and the witness complex on the same subsample determined by a parameter δ. The parameter δ is also used in the graph induced complex and the witness complex. The edges in the Rips complex built on the same *subsample* were of lengths at most $\alpha + 2\delta$. We varied δ and observed

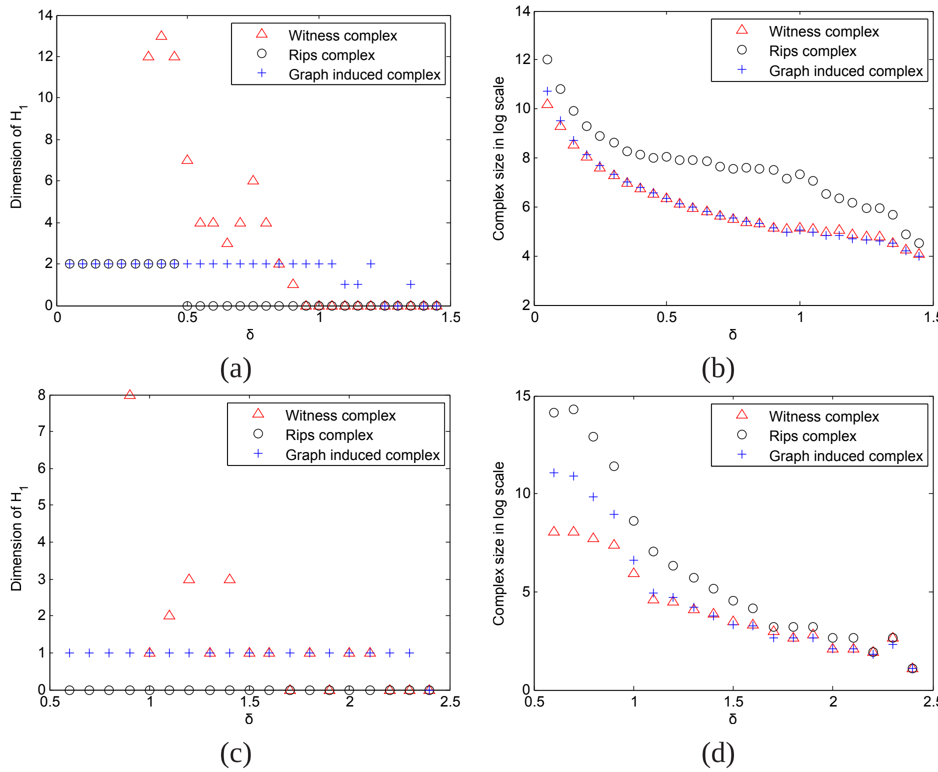

Figure 11: Comparison results for Klein bottle in $\mathbb{R}^4$ (top row) and primary circle in $\mathbb{R}^{25}$ (bottom row). The estimated β_1 for three complexes are shown on the left, and their sizes are shown on log scale on right.

the rank of the one dimensional homology group (β_1). As evident from the plots, the graph induced complex captured β_1 correctly for a significantly wider range of δ (left plots) while its size remained comparable to that of the witness complex (right plots). In some cases, the graph induced complex could capture the correct β_1 with remarkably small number of simplices. For example, it had $\beta_1 = 2$ for Klein bottle when there were 278 simplices for $\delta = 0.7$ and 154 simplices for $\delta = 1.0$. In both cases Rips and witness complexes had wrong β_1 while the Rips complex had a much larger size ($\log_e$ scale plot) and the witness complex had comparable size. This illustrates why the graph induced complex can be a better choice than the Rips and witness complexes.

Constructing a GIC. One may wonder how to efficiently construct the graph induced complexes in practice. Experiments show that the following procedure runs quite efficiently in practice. It takes advantage of computing nearest neighbors within a range and, more importantly, computing cliques only in a sparsified graph.

Let the ball $B(q, \delta)$ in metric d be called the δ-cover for the point q. A graph induced complex $\mathcal{G}^\alpha(P, Q, d)$ where Q is a δ-sparse δ-sample can be built easily by identifying δ-covers with a rather standard greedy (farthest point) iterative algorithm. Let $Q_i = \{q_1, \ldots, q_i\}$ be the point set sampled so far from P. We maintain the invariants (i) Q_i is δ-sparse and (ii) every point $p \in P$ that are in the union of δ-covers $\bigcup_{q \in Q_i} B(q, \delta)$ have their closest point $\nu(p) = \text{argmin}_{q \in Q_i} d(p, q)$ in Q_i identified. To augment Q_i to $Q_{i+1} = Q_i \cup \{q_{i+1}\}$, we choose a point $q_{i+1} \in P$ that is outside the δ-covers $\bigcup_{q \in Q_i} B(q, \delta)$. Certainly, q_{i+1} is at least δ units away from all points in Q_i thus satisfying the first invariant. For the second invariant, we check every point p in the δ-cover of q_{i+1} and update $\nu(p)$ to be q_{i+1} if its distance to q_{i+1} is smaller than the distance $d(p, \nu(p))$. At the end, we obtain a sample $Q \subseteq P$ whose δ-covers cover the entire point set P and thus is a δ-sample of (P, d) which is also δ-sparse. Next, we construct the simplices of $\mathcal{G}^\alpha(P, Q, d)$. This needs identifying cliques in $G^\alpha(P)$ that have vertices with different closest points in Q. We delete every edge pp' from $G^\alpha(P)$ where $\nu(p) = \nu(p')$. Then, we determine every clique $\{p_1, \ldots p_k\}$ in the remaining sparsified graph and include the simplex $\{\nu(p_1), \ldots, \nu(p_k)\}$ in $\mathcal{G}^\alpha(P, Q, d)$. The main saving here is that many cliques of the original graph are removed before it is processed for clique computation.

References

[1] D. Attali, H. Edelsbrunner, and Y. Mileyko. Weak witnesses for Delaunay triangulations of submanifolds. *Proc. ACM Sympos. on Solid and Physical Modeling* (2007), 143–150.

[2] J.-D. Boissonnat and L. J. Guibas and S. Y. Oudot. Manifold reconstruction in arbitrary dimensions using witness complexes. *Proc. 23rd ACM Sympos. Comput. Geom.* (2007), 194–203.

[3] T. K. Dey, F. Fan, and Y. Wang. Graph induced complex on point data. *Proc. 29th Annu. Sympos. Comput. Geom.* (2013), 107–116.

[4] L. J. Guibas and S. Y. Oudot. Reconstructing using witness complexes. *Discrete Comput. Geom.* **30** (2008), 325–356.

[5] V. de Silva and G. Carlsson. Topological estimation using witness complexes. *Proc. Sympos. Point Based Graphics* (2004), 157–166.

[6] V. de Silva. A weak definition of Delaunay triangulation. arxiv:cs.CG/0310031, Oct. 2003.