Chapter 2: Simplicial Complex
Topics in Computational Topology: An Algorithmic View

One of the biggest issue of homotopy and homeomorphism is their computation, and also their somewhat un-intuitive definition (mainly for homotopy). Later we wil talk about another way to classify spaces, called homology groups, which we will focus on. It is weaker than homotopy, namely, homotopic spaces will have the same homology groups, but not vice versa.

The easiest way to introduce homology groups is to start with a nice space, like a triangulation for the surfaces. Basically, such spaces can be considered as a discretization of an underlying space; but they are also useful (as a way to representation) themselves. Specifically, we will talk about the so-called *simplicial complexes*. Section 1 introduces the concept, and in Section 2, we look at some common choices of simplicial complexes, which are especially useful when the input data is a set of discrete point samples from a metric space.

1 Simplicial Complex

A complex essentially is simply a collection of certain types of basic elements satisfying some properties (more precise form will follow later). In the case of simplicial complex, these basic elements are simplices. We first introduce the simplices and simplical complex in a geometric setting. Then we show that we can also consider it more abstractly (which makes the concepts more powerful in practice).

1.1 (Geometric) simplicial complex

Given a set of d points $P = \{p_0, \ldots, p_d\}$ in $\mathbb{R}^N$, P is said to be *geometrically (or linearly) independent* if the set of vectors $v_i = p_i - p_0$ are independent. Namely, $\sum a_i v_i = 0$ implies that $a_i = 0$ for every i.

Given $P = \{p_0, \ldots, p_d\}$, a point $p = \sum_{i=0}^{d} \alpha_i p_i$ is a *convex combination* of $\sum_i \alpha_i = 1$ and $\alpha_i \geq 0$. The *convex hull* of P, denoted by $\mathrm{CHull}(P)$, is the collection of convex combinations of P. A d-simplex σ is simply the convex hull of $d+1$ independent points P. We say that P is the set of vertices of σ or P spans σ. The dimension of σ is d. A 0-simplex is also called a vertex, a 1-simplex an edge (what is it?), a 2-simplex a triangle (why?: look at $p_0 + t_1(p_1 - p_0) + t_2(p_2 - p_0)$), and a 3-simplex a tetrahedron.

Properties of simplices. Let σ be a d-simplex spanned by $P = \{p_0, \ldots, p_d\}$.

P1 σ is convex.

P2 Any subset of P is also linearly independent. Hence their span also forms a simplex σ'. Note that every point in σ' also belongs to σ. So we have that $\sigma' \subseteq \sigma$. We call σ' *a face* of σ, and σ' is *a proper face* if it is not σ. So a d-simplex has $2^{d+1} - 1$ faces. The *boundary of* σ, denoted by $\mathrm{bd}\sigma$, is the union of all proper faces, and the *interior of* σ is the set $\mathrm{int}\sigma = \sigma - \mathrm{bd}\sigma$.

Now look at some examples: what are the faces of an edge? What are the faces of a triangle?

P3 Note that a point is in some face of σ iff some α_i is 0. So every point belongs to exactly the interior of one face, which is the face spanned by the set of points with positive α_is.

P4 $\sigma \sim \mathbb{B}^{d+1}$, bd$\sigma \sim \mathbb{S}^d$, and int$\sigma \sim \mathbb{R}^d$.

Finally, if $p = \sum \alpha_i p_i$, then $< \alpha_0, \ldots, \alpha_d >$ is *barycentric coordinates* of p w.r.t. P.

Definition 1.1 (Simplicial complex) *A* simplicial complex *is a finite collection of simplices K such that $\sigma \in K$ implies that any face of σ is also in K, and $\sigma_1, \sigma_2 \in K$ implies that $\sigma_1 \cap \sigma_2$ is either empty or a face of both.*

L is a subcomplex of K *if L is a simplicial complex and $L \subseteq K$.*

Basically, we are interested in a collection of simplices nicely put together; specially, they are closed under taking faces and that have no improper intersections. The second condition is also equivalent to that any pair of distinct simplices of K have disjoint interiors.

The *dimension of K* is the maximum dimension of its simplices. The *j-th skelenton of K*, denoted by $K^{(j)}$, is the set of simplices with dimension at most k. Note that $K^{(j)}$ necessarily forms a subcomplex of K. $K^{(0)}$ is also called the *vertices* of K.

Definition 1.2 (Underlying space) *Given a simplicial complex K, its* underlying space $|K|$ *is the pointwise union of its cells; that is, $|K| = \bigcup_{\sigma \in K} \{x \in \sigma\}$. $|K|$ is also called the* polytope of K.

That is, $|K$ is the collection of the all points in K.

We can give each simplex its natural topology as a subspace of $\mathbb{R}^N$. We can then give $|K|$ a natural topology defined as: a subset A of $|K|$ is closed iff $A \cap \sigma$ is closed for any $\sigma \in K$. Alternatively, if K is finite, since $|K|$ is embedded, we can consider the subspace topology on K induced from $\mathbb{R}^N$. Note that these two topologies are the same only if K is finite. Otherwise, the first one is finer and more general than the second one. However, in this class, we will talk about only finite simplicial complexes. Hence from now on we consider the underlying space $|K|$ of any finite simplicial complex equipped with the natural induced homology form $\mathbb{R}^N$ (where K embeds into).

1.2 Abstract simplicial complex

The previous discussion relies on the embedding of K in a Euclidean space. We can in fact now define everything in a more abstract setting devoid of geoemtry. In this way, we do not need to worry about how to draw / embed the complex in space. Rather, we only want to focus on how different pieces (simplices) are glued together.

Definition 1.3 (Abstract simplicial complex) *Given a set A of elements, a collection S of finite non-empty subsets of A form an* abstract simplicial complex *if for any $\sigma(\subset A) \in S$, we have that all non-empty subsets $\sigma' \subset \sigma$ are also in S.*

The elements in A form the 0-skeleton of S and are called vertices *of S. Each subset σ of A in S is called a* simplex, *and the dimension of σ equals its cardinality minus 1; that is,* $\dim(\sigma) = |\sigma| - 1$. *A face of a simplex α is any non-empty subset $\beta \subseteq \alpha$.*

In particular, given a (geometric) simplicial complex K, we can construct an abstract simplicial complex S by throwing away all simplices and retaining only their geometric realization. We call S a *vertex scheme* of K, and K a *geometric realization* of S.

Although we remove the embedding information, but we actually do not really lose topological information. In particular, given any abstract simplicial complex A, there is a natural way to embed it (i.e, construct a geometric realization) in the standard simplex $\Delta \subseteq \mathbb{R}^N$, where $N = |VertA|$. Easy to check that no two simplex can intersect. In fact, we need a much smaller dimensional space to construct a geometric realization; see the following Theorem 1.4. However, the standard simplex in $\mathbb{R}^N$ is commonly used for simplicity (as each simplex from A will be embed into a different face of Δ).

Theorem 1.4 (Geometric Realization) *An abstract simplicial complex $\mathcal{K}$ of dimension d has a geometric realization K in $\mathbb{R}^{2d+1}$.*

Proof:[Proof taken from [14].]
First, we embed the set of vertices $Vert\mathcal{K}$ in $\mathbb{R}^{2d+1}$ as a set of points V in general position, that means, any set of $2d+2$ or fewer number of points are independent. Now if $a_0, \ldots, a_k$ span a k-simplex in $\mathcal{K}$, we construct the corresponding geometric simplex. The collection of them is K. To show that K is indeed a simplicial complex, we need to show that for any $\sigma, \tau \in K$, the intersection $\sigma \cap \tau$ is either empty or a face of both (by definition). To see that, pick any two simplices σ and τ. Note that the union of their vertices has cardinality $card(\sigma \cup \tau) = card(\sigma) + card(\tau) - card(\sigma \cap \tau) \leq 2d+2$. Hence vertices V' in $\sigma \cup \tau$ are affinely independent. Hence the intersection of $\sigma \cap \tau$ is either empty, or lying in the convex combination of some subset. This is because that since all vertices in V' are linearly independent, any point $x \in \mathrm{CHull}(V')$ has a unique linear combination of vertices from V'. If $x \in \mathrm{CHull}(vert(\sigma))$ and $x \in \mathrm{CHull}(vert(\tau))$, then x must be a linear combination of only vertices in $V'' = vert(\sigma) \cap vert(\tau)$, thus lying in the convex hull of V''. Hence they intersect at a face of σ and τ. $\blacksquare$

Definition 1.5 (Underlying space of abstract simplicial complex) *The underlying space of an abstract simplicial complex K, denoted also by $|K|$, is the the underlying space of its geometric realization into the standard simplex in $\mathbb{R}^N$ (with $N = |VertK|$).*

One can also use any other geometric realization, and any obtained underlying space are homeomorphic to each other; that is, the choice of geometric realization does not matter.

Stars and links. How do we talk about neighborhood in this simplicial complex setting? A *star* of a simplex τ is the set of simplices that have τ as a face, denoted by $\mathrm{St}\tau = \{\sigma \in K \mid \tau < \sigma\}$. Generally, the star is not closed under taking faces. We can make it into a complex by adding all missing faces. The result is the *closed star*, denoted by $\overline{\mathrm{St}}\tau$, which is the smallest subcomplex that contains the star. The $\mathrm{Lk}\tau$ consists of the set of simplices in the closed star that are disjoint from τ, $\mathrm{Lk}\tau = \{v \in \overline{\mathrm{St}}\tau \mid v \cap \tau = \emptyset\}$. If τ is a vertex then its link is just the difference between the closed star and star. More generally, it is the difference between the closed star of τ and the union of the star of all faces of τ.

Simplicial maps. Intuitively, simplicial complexes can be thought of discretization of spaces of interests. We now need in some sense an analog of "continuous functions" between space in this discrete setting. In particular, we will talk about the so-called simplicial maps between two simplicial complexes. It turns out that any continuous function between two spaces can be "approximated" by simplicial maps over appropriate complexes.

Definition 1.6 (Simplicial maps) *Given two simplicial complexes K and L, a* simplicial map *is a function $\phi : K \to L$ such that $\phi(Vert(K)) \subseteq Vert(L))$ and for any $\sigma = \{v_0, \ldots, v_d\} \in K$, $\phi(\sigma)$ is spanned by $\{\phi(v_0), \ldots, \phi(v_d)\}$.*

Intuitively, under a simplicial map ϕ, the images of vertices of a simplex K always span a simplex in L. Note that the image $\phi(\sigma)$ may have smaller dimension than σ. Give an example of a rectangle to a torus.

A simplicial map $\phi : K \to L$ is an *isomrophism* if it is bijective; in which case we say that K and L are *isomorphic*. This is basically a high-dimensional generalization of graph isomorphism. Give an example that maps a triangle to a solid triangle. This is a simplicial map and bijection on vertices only.

It turns out that a simplicial map ϕ induces a continuous function $\xi_\phi : |K| \to |L|$ such that $x = \sum \alpha_i v_i$ means that $\xi_\phi(x) = \sum \alpha_i \phi(v_i)$. That is, ξ_ϕ is linear when restricted to each simplex of K. We will not prove that this map is indeed continuous.

Lemma 1.7 *If $\phi : K \to L$ is an isomorphism, then its induced continuous function ξ_ϕ is a homeomorphism between $|K|$ and $|L|$.*

Example: Triangulation of manifolds.

2 Commonly Used Complexes

We now talk about several geometric or abstract complexes that are commonly used in practice. In particular, very often, the input is a set of points, say that sampled or approximating some hidden space. Points do not have any interesting topology, and we would like to enforce some connectivity on them and some topology. Specifically, assumed that we have a set of points $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$. In what follows, we use $\mathbb{B}(p, r)$ to denote the d-dimensional Euclidean ball centered at p with radius r, and $d(x, y)$ to denote the Euclidean distance between x and y.

2.1 Delaunay complex

One of the most well-known simplicial complex spanned by a set of points in $\mathbb{R}^d$ is the *Delaunay complex* (sometimes called the *Delaunay triangulation*, especially in low dimensions) defined as follows:

Definition 2.1 (Delaunary complex) *Given a set of points $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$, a k-simplex $\sigma = \langle p_{i_0}, \ldots, p_{i_k} \rangle$ is in the* Delaunay complex $\mathrm{Del}(P)$ *if and only if there exists a ball $\mathbb{B}$ whose boundary contains vertices of σ and that the interior of $\mathbb{B}$ contains no point from P. A simplex in* $\mathrm{Del}(P)$ *is also called a* Delaunay simplex.

The Delaunay complex is named after mathematician Boris Delaunay [9]. It has many beautiful properties, and has been well-studied, especially for the low-dimensional case when $d = 2$ or 3. Indeed, this concept is fundamental to the fields of surface reconstruction and meshing. See for examples books [10, 24]. The Delaunay complex is the dual complex of the so-called *Voronoi diagram* $\mathrm{Vor}(P)$ *of* P [23], which decomposes the space $\mathbb{R}^d$ into cells. Each cell is uniquely associated with an input point p_i, and contains the set of points from $\mathbb{R}^d$ that have p_i as the nearest neighbor among points in P; that is, the Voronoi cell of p_i is given by $\{x \in \mathbb{R}^d \mid d(x, p_i) \leq d(x, p)$, for any $p \in P\}$. See Figure 1 (a) for an illustration.

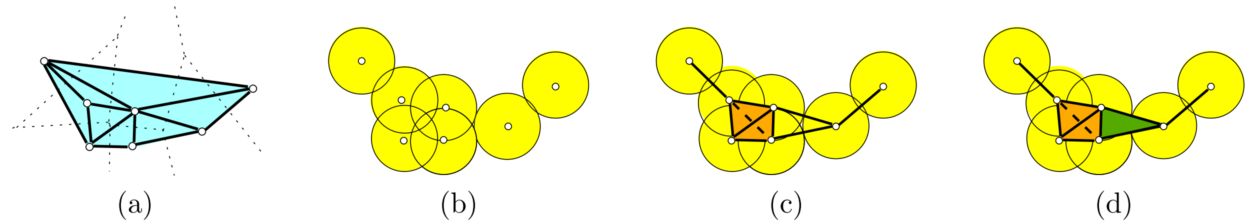


Figure 1: (a) Delaunay complex (solid lines). Dotted lines represent the dual Voronoi diagram. (b) The union of balls around the same set of points for a specific radius r. The Čech complex is in (c) while the Rips complex is in (d). Note that both the Čech and Rips complex contain the tetrahedron in the middle. The Rips complex contains the triangle on the right, while the Čech complex does not, as the three corresponding balls do not have common intersection.

For points in high-dimensional space, however, the construction of the Delaunay complex is expensive: It takes $O(n^{\lceil \frac{d}{2} \rceil} + n \log n)$ time in worst case to compute the Delaunay complex for a set of n points in $\mathbb{R}^d$ [6]. Furthermore, computing only the first-dimensional Delaunay simplices does not appear to be any faster, asymptotically, than computing the full Delaunay complex. This makes the complex less appealing for high dimensional data analysis.

2.2 Čech and Rips complexes

Definition 2.2 (Čech complex) *Given a set of points $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$ and a real value $r > 0$, a k-simplex $\sigma = \langle p_{i_0}, \ldots, p_{i_k} \rangle$ is in the* Čech complex $\mathsf{C}^r(P)$ *if and only if* $\bigcap_{0 \le j \le k} \mathbb{B}(p_{i_j}, r) \neq \emptyset$.

Note that the definition of Čech complex involves a parameter r, which can be useful in practice. In particular, one can imagine that we grow a ball around each point p_i in P, and look at the union of balls at scale r. See Figure 1 (b) and (c) for an illustration.

Nerves. It turns out that the Čech complex $\mathsf{C}^r(P)$ is the nerve of the set system $\{\mathbb{B}(p_i, r) \mid p_i \in P\}$. Given a finite collection of sets F, its *nerve* is defined as all non-empty subcollections whose sets have a non-empty common intersection:

$$NrvF := \{X \subseteq F \mid \bigcap X \neq \emptyset\}.$$

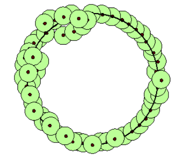
(Given an example of non-convex sets.) The following result is an important one and very useful in practice to approximate topology of some space.

Theorem 2.3 (Nerve Theorem) *Let F be a finite collection of closed, convex sets in Euclidean space. Then the nerve of F and the union of the sets in F have the same homotopy type.*

In fact, conditions in the Nerve Theorem can be relaxed a bit: as long as $\cup F$ is triangulable, all sets in F are closed, and all non-empty common intersections are contractible, then the theorem holds.

Since each Euclidean ball is convex, we have that the non-empty intersection of any collection of balls is also convex. Hence by the Nerve's Lemma (see e.g, Chapter 4.G of [20]), we obtain the following:

Claim 2.4 *The Čech complex $\mathsf{C}^r(P)$ is homotopy equivalent to the union of balls $\mathbb{U}^r(P) := \bigcup_{p_i \in P} \mathbb{B}(p_i, r)$; that is, $\mathsf{C}^r(P) \approx \mathbb{U}^r(P)$.*



This provides the motivation behind choosing the Čech complex as the simplicial complex built on points P sampled from X: Intuitively, one can consider it as using the union of balls around sample points in P to approximate the hidden domain X. See the right figure for an example, where given a set of points sampled around a circle, the union of balls approximates the hidden domain, which is the circle in this case. This is a very important result in manifold reconstruction. In particular, it turns out that if we sample a hidden manifold nicely, then the union-of-balls deformation contract to the hidden manifold. Hence we can recover topological information from the Čech complex.

Note that the Čech complex is not embedded in $\mathbb{R}^d$: In fact, it could have simplices of dimension up to $n-1$, where n is the number of points in P. On the other hand, since the Čech complex $\mathsf{C}^r(P)$ is homotopy equivalent to the union of balls $\mathsf{U}^r(P)$ which is a subset of $\mathbb{R}^d$, $\mathsf{C}^r(P)$ has trivial homology for dimension larger than d. Hence it is not necessary to compute the k-skeleton of Čech complex $\mathsf{C}^r(P)$ for $k > d$.

Nevertheless, compared to the Delaunay complex, it tends to have much larger number of simplices. Consider a set of k points Q in the plane contained within a disk of radius r. If these points are in generic positions, then the Delaunay triangulation is an embedded complex in the plane. Hence the number of simplices in $\mathrm{Del}(Q)$ is $\Theta(k)$. However, for the parameter r, the Čech complex $\mathsf{C}^r(Q)$ contains $\Theta(k^3)$ triangles, since the intersection of any three balls centered at points in Q is non-empty.

There is a related concept, called the *alpha complex*, that also encodes the structure of union of (not necessarily congruent) balls: it was originally introduced in [15] for points in $\mathbb{R}^2$ and extended in [13, 16]. Different from Čech complex but similar to the Delaunay triangulation, the alpha complex has a natural geometric realization, and can have much smaller size than Čech complex. The alpha complexes have found many applications in three-dimensions, especially in structural molecular biology. In high dimensions, it has the same computational issue as the Delaunay complexes.

Finally, a straightforward algorithm to compute of the k-th skeleton of the Čech complex takes time $O(n^{k+1})$ time: While this is expensive for large k, in practice, one often is only interested in low dimensional topological information, say $k = 1$ or 2, but the ambient dimension d could be much larger.

The Vietoris-Rips complex is a variant of the Čech complex, defined as follows.

Definition 2.5 (Rips complex) *Given a set of points $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$ and a real value $r > 0$, a k-simplex $\sigma = \langle p_{i_0}, \ldots, p_{i_k} \rangle$ is in the* Vietoris-Rips (Rips) complex $\mathsf{R}^r(P)$ *if and only if $\mathbb{B}(p_{i_j}, r) \cap \mathbb{B}(p_{i_{j'}}, r) \neq \emptyset$ for any $j, j' \in [0, k]$.*

In other words, points $p_{i_0}, \ldots, p_{i_k}$ span a k-simplex if and only if the Euclidean balls with radius r centered at these points have pairwise intersection. See Figure 1 (d) for an illustration.

An alternative way to think of the Rips complex $\mathsf{R}^r(P)$ is as follows: An edge (1-simplex) $\langle p_{i_0}, p_{i_1} \rangle$ is in $\mathsf{R}^r(P)$ if and only if $\mathbb{B}(p_{i_0}, r) \cap \mathbb{B}(p_{i_1}, r) \neq \emptyset$. Indeed, the 1-skeleton of $\mathsf{R}^r(P)$ is the same as the 1-skeleton of the Čech complex $\mathsf{C}^r(P)$. The Rips complex $\mathsf{R}^r(P)$ is the clique complex induced by its 1-skeleton: that is, for every k-clique in the 1-skeleton of $\mathsf{R}^r(P)$ (which can be considered as a graph), the vertices involved spans a $k-1$-simplex in $\mathsf{R}^r(P)$. In other words, the Rips complex is the *clique complex*, also called the *flag complex*, induced by its edge set.

It is easy to see that $\mathsf{C}^r(P) \subseteq \mathsf{R}^r(P)$. More generally, the Rips complex and the Čech complex are related as follows (see e.g, [5]):

Claim 2.6 *For a set of points P in a metric space,*

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

For the case where $P \subset \mathbb{R}^d$, we have

$$\mathsf{C}^r(P) \subseteq \mathsf{R}^r(P) \subseteq \mathsf{C}^{\sqrt{\frac{2d}{d+1}}r}(P).$$

Given that the Rips complex is completely defined by its 1-skeleton (as it is the clique complex induced by it), once one computes the 1-skeleton of $\mathsf{R}^r(P)$, the information of the entire complex is implicitly encoded. This and the simplicity of the Rips complex make it one of the most popular type of simplicial complexes studied in Computational topology. At the same time, however, the size of the Rips complex is often huge in practice, even just the 2-skeleton of it. This has severely impacted the application of the Rips complex for high dimensional point cloud data. There have been various recent work aiming to tackle this issue: For example, Sheehy proposed an algorithm to sparsify the input point cloud data while still approximating the so-called persistent homology of a sequence of Rips complexes [22], with follow-up work in [4, 12]. See also [21] for interesting discussion and comparison of various Rips-like zigzag sequence for homology inference. Succinct representation and simplification of clique complexes (such as Rips complex) has been studied in [2, 25].

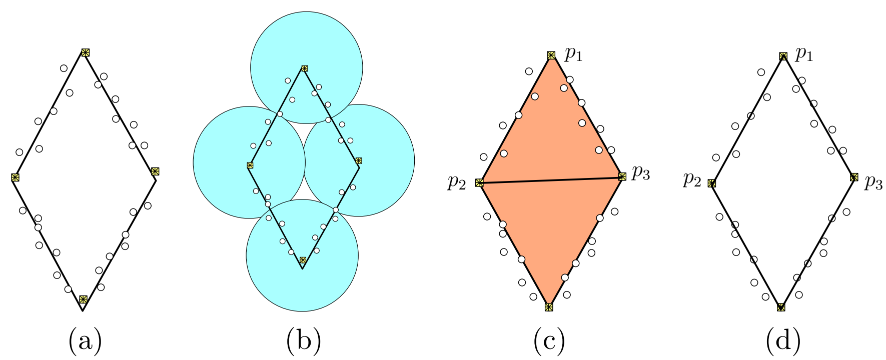


Figure 2: (a) A set of points P sampling a curve. The four square points are subsampled landmarks Q. (b) The union of balls around subsamples in Q for some radius r. Its corresponding Rips complex is shown in (c), containing two triangles. Shrinking the radius r however will disconnect poits p_1 and p_3. The witness complex, shown in (d), has the same topology as the hidden curve. For this particular example, the graph induced complex is the same as (d).

2.3 Witness complex

One idea to reduce the size of the input points is via subsampling, and then build a complex on the subsamples $Q \subseteq P$. However, ideally one should still be able to use information from P as one builds a complex on Q. Indeed, see Figure 2, for a sparse subset Q, directly using Euclidean distance between points to build a Rips complex ignore the structure captured by the denser points in P. One of the first work that introduced the idea of building a complex on a subsample while leveraging information from original points is an elegant structure called the *witness complex*, proposed in [8, 7]. It is defined as follows.

Definition 2.7 *A simplex $\sigma = \langle q_0, \dots, q_k \rangle$ is* weakly witnessed *by a point x if $d(q_i, x) \le d(q, x)$ for any $i \in [0, k]$ and $q \in Q \setminus \{q_0, \dots, q_k\}$. It is* strongly witnessed *by a point x if in addition to being weakly witnessed by x, we also have that $d(q_i, x) = d(q_j, x)$ for any $i, j \in [0, k]$.*

Definition 2.8 *Given a set of points $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$ and a subset $Q \subseteq P$, the* witness complex $\mathsf{W}(Q, P)$ *is the collection of simplices with vertices from Q whose all faces are weakly witnessed by some point in P.*

We remark that the original definition [8] is for any metric space, while here we only present the case where points are sampled from the Euclidean space. We note that it is possible that a simplex is weakly witnessed by some point in P, but not all its faces. See Figure 1 (d) for an example: Here, note that the three points $\{p_0, p_1, p_2\}$ are actually weakly witnessed by several points in P (pick an arbitrary point above the segment p_2p_3). However, this triangle is not in the witness complex because one of its face, the edge p_2p_3, is not weakly witnessed by any point in P. Thus we have the requirement that a weakly witnessed complex is in $\mathsf{W}(Q, P)$ if and only if all its faces (subsimplices) are also weakly witnessed by some point in P, to enforce the condition of being a simplicial complex. In the case where points are sampled from the Euclidean space, it turns out that such simplices are exactly strongly witnessed simplices [7].

Points in Q are often referred to as *landmarks* [8]. The definition of a witnessed complex bears similarity with the Delaunay complex. Recall that a simplex σ is Delaunay if there exists a Euclidean ball whose boundary contains all vertices of σ, and whose interior does not contain any other point from P. In contrast, a strongly witnessed complex σ spanned by landmarks in Q means that there exists an Euclidean ball *centered at a point p in P* (i.e, witnessed by p), such that its boundary contains all vertices of σ, and its interior is free of other points from Q. Indeed, we have the following relations:

Claim 2.9 ([7]) *Given an arbitrary set of points Q from $\mathbb{R}^d$, $\mathsf{W}(Q, \mathbb{R}^d) = \mathrm{Del}(Q)$.*
In general, given two set of points $Q \subseteq P \subset \mathbb{R}^d$, we have $\mathsf{W}(Q, P) \subseteq \mathrm{Del}(Q)$.

The witness complex has many nice properties, and can for example be used to reconstruct a smooth curve or surface from its point samples, which is much stronger than simply recovering the homological information [1, 19]. Its size can be much smaller than the Čech or Rips complex. However, due to the condition of being witnessed, it may contain too few simplices to capture the topology of a sampled space in dimensions three or more [19]. To tackle this issue, some modifications of the witness complex were suggested in [3, 19], which enlarges the witness complexes to enable topology inference, but also makes it more complicated and costly to compute.

2.4 Graph-induced complex

First, we introduce the graph induced complex [11] in a more abstract setting:

Definition 2.10 *Let $G(V)$ be a graph with vertex set V. Let $L \subset V$ be called a set of* landmarks. *Let $\nu : V \to L$ be a vertex map where $\nu(V) = L \subseteq V$. The* graph induced complex *$\mathcal{G}(G(V), L, \nu)$ is defined as the simplicial complex where a k-simplex $\sigma = \{v'_0, v'_1, \ldots, v'_k\}$ is in $\mathcal{G}(G(V), L, \nu)$ if and only if there exists a $(k+1)$-clique in G spanned by $\{v_0, v_1, \ldots, v_k\} \subseteq V$ so that $\nu(v_i) = v'_i$ for each $i \in \{0, 1, \ldots, k\}$.*

Now assume that the vertices of the input graph are associated with a metric structure (P, d); that is, $d : P \times P \to \mathbb{R}^+$ satisfies (i) $d(x, y) \geq 0$, (ii) $d(x, y) = 0$ iff $x = y$; (iii) $d(x, y) = d(y, x)$ and (iv) $d(x, y) \leq d(x, z) + d(z, y)$.

Definition 2.11 *Let (P, d) be a metric space where P is a finite point set and let $Q \subset P$ be a subset. Let $\nu_d : P \to Q$ denote the nearest point map where $\nu_d(p)$ is a point in $argmin_{q \in Q} d(p, q)$.*

Given a graph $G(P)$ with vertex set P, we define its graph induced complex as $\mathcal{G}(G(P), Q, d) := \mathcal{G}(G(P), Q, \nu_d)$.

A preliminary version of the graph induced complex $\mathcal{G}(G(P), Q, d)$ was introduced in [17] for the application of sensor network routing. Also, the idea follows the concept of *combinatorial Delaunay triangulation* of [8] and is a discrete version of the dual of the geodesic Voronoi diagram as studied in [18]. In [11], the concept is formally formulated, and several theoretical results about its homology inference ability are obtained.

Intuitively, the nearest point map ν_d decompose points in P into a set of "discrete" version of Voronoi cells, one for each point $q \in Q$. A set of k points $q_1, \ldots, q_k$ in Q spans a simplex in $\mathcal{G}(G(P), Q, d)$ if and only there exists a k-clique from the graph $G(P)$ with one point from each Voronoi cell $\mathrm{Vor}(q_i)$. See Figure 3 (taken from [11]) for an illustration, where the colors indicate points in the "discrete" Voronoi cells, and the detailed input graph contained in the shaded region on the left is shown on the right.

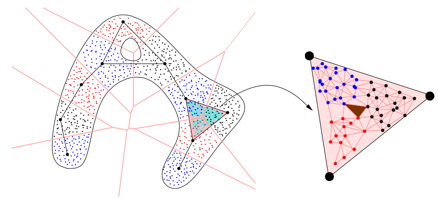


Figure 3: Image taken from [11]

The definition of the graph induced complex requires an input graph $G(P)$ with vertex set P. In the context where we are given a set of points P sampled from a hidden domain X, one can use the 1-skeleton of the Rips complex $\mathsf{R}^\alpha(P)$ for some small α as $G(P)$. Note that this 1-skeleton can be computed in $O(n^2)$ time where $n = |P|$. Similar to the witness complex, the size of the graph induced complex can be orders of magnitude smaller than that of the Rips complex (under comparable parameters); see Figure 2 in [11] for empirical evidence. At the same time, it was shown in [11] that the graph induced complex can recover homological information of the hidden domain for appropriated parameters.

References

[1] D. Attali, H. Edelsbrunner, and Y. Mileyko. Weak witnesses for Delaunay triangulations of submanifolds. In *Symposium on Solid and Physical Modeling*, pages 143–150, 2007.

[2] D. Attali, A. Lieutier, and D. Salinas. Efficient data structure for representing and simplifying simplicial complexes in high dimensions. *Int. J. Comput. Geometry Appl.*, 22(4):279–304, 2012.

[3] J.-D. Boissonnat, L. J. Guibas, and S. Oudot. Manifold reconstruction in arbitrary dimensions using witness complexes. *Discrete & Computational Geometry*, 42(1):37–70, 2009.

[4] M. Buchet, F. Chazal, S. Y. Oudot, and D. Sheehy. Efficient and robust topological data analysis on metric spaces. *CoRR*, abs/1306.0039, 2013.

[5] F. Chazal and S. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proc. 24th ACM Sympos. on Comput. Geom.*, pages 232–241, 2008.

[6] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf. *Computational Geometry: Algorithms and applications.* Springer, 2 edition, 2000.

[7] V. de Silva. A weak definition of Delaunay triangulation. *CoRR*, cs.CG/0310031, 2003.

[8] V. de Silva and G. Carlsson. Topological estimation using witness complexes. In *Sympos. Point-Based Graphics*, pages 157–166, 2004.

[9] B. Delaunay. Sur la aphère vide. a la memoire de georges voronoi. *Izv. Akad. Nauk SSSR, Otdelenie Mathematicheskii i Estestvennyka Nauk*, 7:793–800, 1934.

[10] T. K. Dey. *Curve and Surface Reconstruction: Algorithms with Mathematical Analysis (Cambridge Monographs on Applied and Computational Mathematics).* Cambridge University Press, New York, NY, USA, 2006.

[11] T. K. Dey, F. Fan, and Y. Wang. Graph induced complex on point data. In *Proc. 29th Annu. ACM Sympos. Compu. Geom.*, pages 107–116, 2013.

[12] T. K. Dey, F. Fan, and Y. Wang. Computing topological persistence for simplicial maps. In *Proc. 30th Annu. ACM Sympos. Compu. Geom.*, 2014. To appear.

[13] H. Edelsbrunner. The union of balls and its dual shape. *Discrete & Computational Geometry*, 13:415–440, 1995.

[14] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction.* Amer. Math. Soc., Providence, Rhode Island, 2010.

[15] H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel. On the shape of a set of points in the plane. *IEEE Trans. Inform. Theory*, 29(4):551–558, 1983.

[16] H. Edelsbrunner and E. P. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graph.*, 13(1):43–72, 1994.

[17] Q. Fang, J. Gao, L. J. Guibas, V. de Silva, and L. Zhang. GLIDER: gradient landmark-based distributed routing for sensor networks. In *INFOCOM*, pages 339–350, 2005.

[18] J. Gao, L. J. Guibas, S. Oudot, and Y. Wang. Geodesic delaunay triangulation and witness complex in the plane. In *Proc. 19th ACM-SIAM Sympos. Discrete Algorithms*, pages 571–580, 2008.

[19] L. J. Guibas and S. Oudot. Reconstruction using witness complexes. *Discrete & Computational Geometry*, 40(3):325–356, 2008.

[20] A. Hatcher. *Algebraic Topology.* Cambridge U. Press, New York, 2002.

[21] S. Y. Oudot and D. Sheehy. Zigzag zoology: rips zigzags for homology inference. In *Proc. 29th Annu. ACM Sympos. Compu. Geom.*, pages 387–396, 2013.

[22] D. Sheehy. Linear-size approximations to the vietoris-rips filtration. *Discrete & Computational Geometry*, 49(4):778–796, 2013.

[23] G. Voronoi. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. *J. Reinne Angew. Math.*, 134:198–287, 1908.

[24] S. wing Cheng, T. K. Dey, and J. R. Shewchuk. *Delaunay Mesh Generation.* CRC Press, 2012.

[25] A. Zomorodian. The tidy set: A minimal simplicial set for computing homology of clique complexes. In *Proc. ACM Symposium of Computational Geometry*, pages 257–266, 2010.