Topological Data Analysis, Spring 2018

Simplicial Homology

Slightly rearranged, but mostly copy-pasted from

- Harer's and Edelsbrunner's Computational Topology,
- Verovsek's class notes.
- Gunnar Carlsson's Topology and Data.

In order to study and manipulate complex shapes it is convenient to discretize these shapes and to view them as the union of simple building blocks glued together in a 'clean fashion.' The building blocks should be simple geometric objects, for example, points, lines segments, triangles, tehrahedra and more generally simplices, or even convex polytopes. We will begin by using simplices as building blocks. The material presented in this chapter consists of the most basic notions of combinatorial topology, going back roughly to the 1900-1930 period and it is covered in nearly every algebraic topology book (certainly the 'classics'). Moreover, we have algorithms for computing homology when the space is a simplicial complex.

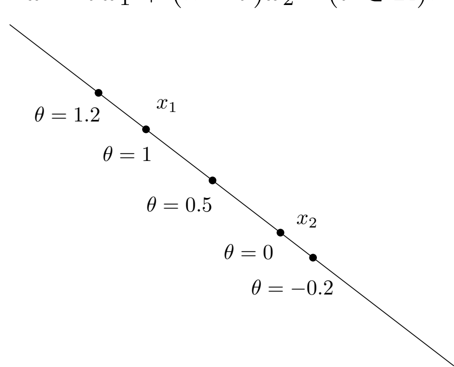
1 Simplicial Complexes

1.1 Affine and Convex Combinations

Definition 1.1 (Affine Combinations)**.** Let $u_0, u_1, \ldots, u_k$ be points in $\mathbb{R}^d$. A point $x = \sum_{i=0}^{k} \lambda_i u_i$ is an *affine combination* of the u_i if the λ_i sum to 1. The affine hull is the set of affine combinations. It is a k-plane if the $k+1$ points are *affinely independen*t by which we mean that any two affine combinations, $x = \lambda_i u_i$ and $y = \mu_i u_i$, are the same iff $\lambda_i = \mu_i$ for all i. The $k+1$ points are affinely independent iff the k vectors $u_i - u_0$, for $1 \leq i \leq k$, are linearly independent. In $\mathbb{R}^d$ we can have at most d linearly independent vectors and therefore at most $d+1$ affinely independent points.

Example 1.2. *Line* through x_1, x_2: all points

$$x = \theta x_1 + (1-\theta)x_2 \quad (\theta \in \mathbb{R}).$$

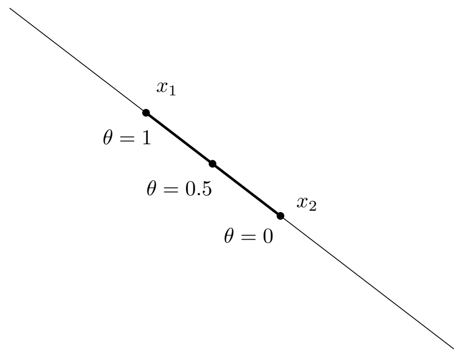


Definition 1.3 (Convex Combinations)**.** An affine combination, $x = \sum_{i=0}^{k} \lambda_i u_i$, is a convex combination if all λ_i are non-negative. The convex hull of a set S is the set of convex combinations of points in S.

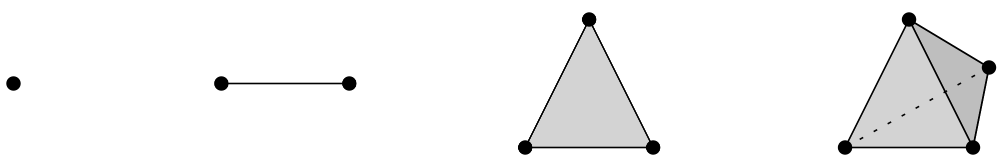
Example 1.4. *Line segment* between x_1 and x_2: all points

$$x = \theta x_1 + (1 - \theta)x_2 \quad (\theta \in \mathbb{R})$$

with $0 \le \theta \le 1$. The convex hull of x_1, x_2 is the line segment between x_1 and x_2.



Definition 1.5 (Simplex)**.** A *k-simplex* is the convex hull of $k + 1$ affinely independent points, $\sigma = \text{conv}\{u_0, u_1, \dots, u_k\}$. We sometimes say the u_i span σ. Its dimension is $\dim \sigma = k$. We use special names for the first few dimensions, *vertex* for 0-simplex, *edge* for 1-simplex, *triangle* for 2-simplex, and *tetrahedron* for 3-simplex.



Any subset of affinely independent points is again affinely independent and therefore also defines a simplex. A face of σ is the convex hull of a non-empty subset of the u_i and it is proper if the subset is not the entire set. We sometimes write $\tau \le \sigma$ if τ is a face and $\tau < \sigma$ if it is a proper face of σ. If τ is a (proper) face of σ we call σ a (proper) coface of τ. Since a set of size $k + 1$ has 2^{k+1} subsets, including the empty set, σ has $2^{k+1} - 1$ faces, all of which are proper except for σ itself. The boundary of σ, denoted as $\text{bd}\,\sigma$, is the union of all proper faces, and the interior is everything else, $\text{int}\,\sigma = \sigma \setminus \text{bd}\,\sigma$. A point $x \in \sigma$ belongs to $\text{int}\,\sigma$ iff all its coefficients λ_i are positive. It follows that every point $x \in \sigma$ belongs to the interior of exactly one face, namely the one spanned by the points u_i that correspond to positive coefficients λ_i.

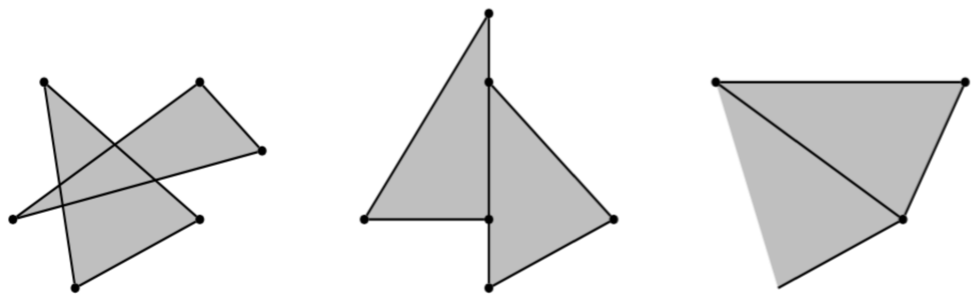
1.2 Simplicial complexes

We are interested in sets of simplices that are closed under taking faces and that have no improper intersections.

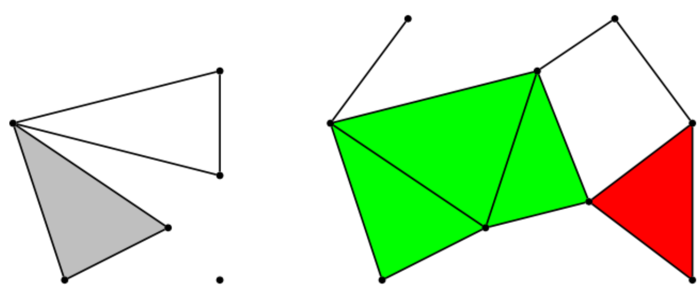
Example 1.6. A simplicial complex is a finite collection of simplices K such that

1. Every face of a simplex in K also belongs to K.
2. For any two simplices σ_1 and σ_2 in K, if $\sigma_1 \cap \sigma_2 \neq \emptyset$, then $\sigma_1 \cap \sigma_2$ is a common face of both σ_1 and σ_2.

The dimension of the simplicial complex K is the maximum of the dimensions of all simplices in K. If dim $K = d$, then every face of dimension d is called a cell and every *face* of dimension $d-1$ is called a *facet*.



On the left, the intersection of the two 2-simplices is neither an edge nor a vertex of either triangle. In the middle case, two simplices meet along an edge which is not an edge of either triangle. On the right, there is a missing edge and a missing vertex. Below are examples of simplicial complexes.



The union $|K|$ of all the simplices in K is a subset of $\mathbb{R}^d$. We can equip $|K|$ with the topology inherited from the ambient Euclidean space in which the simplices live. The resulting topological space $|K|$ is called the *geometric realization* of K.

Definition 1.7. Given any complex, K_2, a subset $K_1 \subseteq K_2$ of K_2 is a subcomplex of K_2 iff it is also a complex.

For any complex, K, of dimension d, for any i with $0 \leq i \leq d$, the subset

$$K^{(i)} = \{\sigma \in K \mid \dim \sigma \leq i\}$$

is called the i-skeleton of K. Clearly, $K^{(i)}$ is a subcomplex of K. We also let

$$K^i = \{\sigma \in K \mid \dim \sigma = i\}.$$

Observe that K^0 is the set of vertices of K and K^i is not a complex.

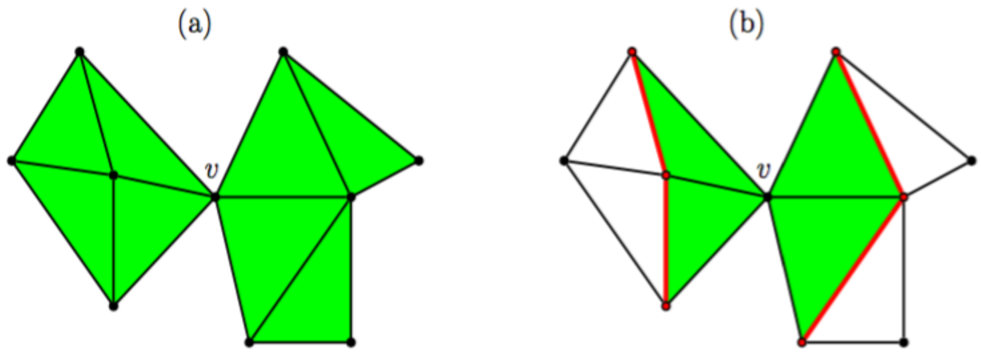
A subset of a simplicial complex K useful when talking about local neighborhoods is the *star* of a simplex τ

$$\text{St}\ \tau = \{\sigma \in K \mid \tau \leq \sigma\}.$$

consisting of all cofaces of τ. 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*, $\overline{\text{St}}\ \tau$, which is the smallest subcomplex that contains the star. The *link* consists of all simplices in the closed star that are disjoint from τ,

$$\text{Lk}\ \tau = \{v \in \overline{\text{St}}\ \tau \mid v \cap \tau = \emptyset\}.$$

It τ is a vertex then the link is just the difference between the closed star and the star. More generally, it is the closed star minus the stars of all faces of τ.

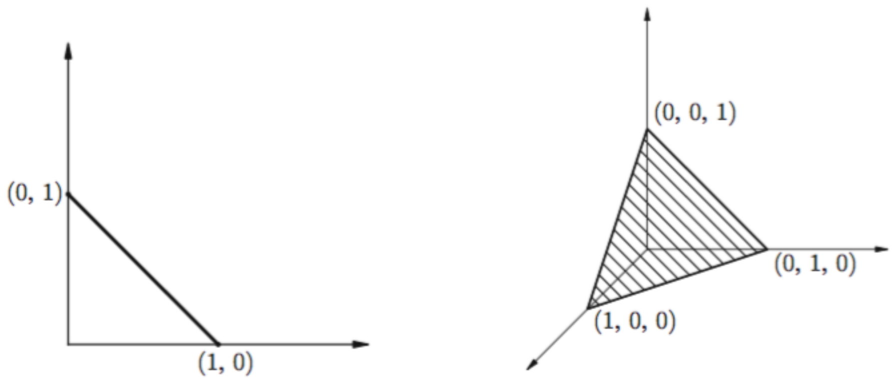


1.3 Abstract simplicial complex

It is often easier to construct a complex abstractly and to worry about how to put it into Euclidean space later, if at all.

Definition 1.8 (Abstract simplicial complex)**.** By an abstract simplicial complex A, we will mean a pair $A = (V, \Sigma)$, where V is a finite set called the vertices of A, and where Σ is a subset (called the simplices) of the collection of all non-empty subsets of V, satisfying the conditions that if $\sigma \in \Sigma$, and $\emptyset \neq \tau \subseteq \sigma$, then $\tau \in \Sigma$. Simplices consisting of exactly two vertices are called edges.

Given a (geometric) simplicial complex K, we can construct an abstract simplicial complex A by throwing away all simplices and retaining only their sets of vertices. We call A a vertex scheme of K. Symmetrically, we call K a geometric realization of A. Constructing geometric realizations is surprisingly easy if the dimension of the ambient space is sufficiently high.



Theorem 1.9 (Geometric Realization Theorem)**.** *Every abstract simplicial complex $A = (V, \Sigma)$ with $n+1$ vertices has a geometric realization in $\mathbb{R}^{n+1}$.*

Proof. Assume that V has $n+1$ points. Let φ be a bijection between V and the set of vertices of any n-simplex in $\mathbb{R}^{n+1}$. Define a subcomplex by

$$K = \{\{\varphi(v_0)\varphi(v_1)\ldots\varphi(v_k)\} \mid \{v_0, v_1, \ldots, v_k\} \text{ is a } k \text{ simplex of } A\}.$$

Then K is a geometric realization of A. □

We denote the geometric realization of A by $|A|$.

The natural equivalent of continuous maps between topological spaces are simplicial maps between simplicial complexes, which we now introduce. Let K be a simplicial complex with vertices $u_0, u_1, \ldots, u_n$. Every point $x \in |K|$ belongs to the interior of exactly one simplex in K. Letting $\sigma = \text{conv}\{u_0, u_1, \ldots, u_k\}$ be this simplex, we have $x = \sum_{i=0}^{k} \lambda_i u_i$ with $\sum_{i=0}^{k} \lambda_i = 1$ and $\lambda_i > 0$ for all i. Setting $b_i(x) = \lambda_i$ for $0 \le i \le k$ and $b_i(x) = 0$ for $k+1 \le i \le n$, we have $x = \sum_{i=0}^{n} b_i(x) u_i$, and we call the $b_i(x)$ the *barycentric coordinates* of x in K.

We use these coordinates to construct a piecewise linear, continuous map from a particular kind of map between the vertices of two simplicial complexes. A vertex map is a function $\varphi \colon \text{Vert}K \to \text{Vert}L$ with the property that the vertices of every simplex in K map to vertices of a simplex in L. Then φ can be extended to a continuous map $f \colon |K| \to |L|$ defined by

$$f(x) = \sum_{i=0}^{n} b_i(x)\varphi(u_i),$$

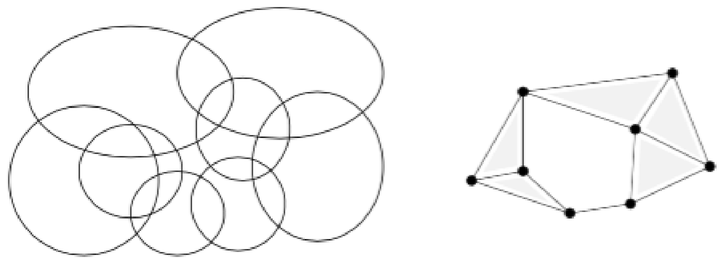
the *simplicial map* induced by φ.

2 Point Cloud Triangulations and the Nerve Lemma

We now discuss a variety of ways to triangulate a collection of points. Finite metric spaces have no interesting topology themselves, so what we will do is try to deduce the shape of a point cloud based on sets that cover this point cloud. To get from the continuous world of open and closed sets set to simplicial complexes, we use a contruction called the nerve of a covering.

Definition 2.1 (Nerve of a Covering)**.** Let X be a topological space, and let $\mathscr{U} = \{U_i\}_{i \in I}$ be any covering of X. The *nerve* of $\mathscr{U}$, denoted by $\mathcal{N}\mathscr{U}$, is the abstract simplicial complex with vertex set I, where a family $\{i_0, \ldots, i_k\}$ spans a k-simplex if and only if $U_{i_0} \cap \ldots \cap U_{i_k} \neq \emptyset$.

Example 2.2. The picture depicts a covering and the nerve.



The main use of the nerve is the following. Suppose we want to represent some topological space X in a combinatorial fashion. We cover X by some collection of sets $\mathscr{U}$ and then take the nerve. Hopefully this nerve will faithfully represent the original space, and in certain situations this is guaranteed. Recall that a space is called contractible if it is homotopically equivalent to a point.

Theorem 2.3 (NERVE THEOREM)**.** *Suppose that X and $\mathscr{U}$ are as above, and suppose that the covering consists of finitely many open sets. Suppose further every nonempty intersection of sets in $\mathscr{U}$ is contractible. Then $\mathcal{N}\mathscr{U}$ is homotopy equivalent to X.*

This implies that the nerve has homotopy groups isomorphic to the underlying space. One now needs methods for generating coverings.

2.1 Čech Complex

When the space in question is a finite metric space X, one covering is given by the family $B_r(X) = \{B_r(x)\}_{x \in X}$, for some $r > 0$. We will denote this construction by $\check{\mathrm{C}}(X, r)$, and refer to it as the Čech complex attached to X and r.

We note three facts about Čech complexes before moving on:

1. Since the sets $B_r(x)$ are all convex, the Nerve Theorem applies, and hence $\check{\mathrm{C}}(X, r)$ has the same homotopy type as the union of r-balls around the points in X. This latter set can be thought of as a 'thickening' by r of the point set X.

2. Given two radii $r < r'$, we obviously have the inclusion $\check{\mathrm{C}}(X, r) \subseteq \check{\mathrm{C}}(X, r')$. Hence, if we let r from 0 to ∞, we produce a nested family of simplicial complexes, starting with a set of n vertices and ending with an enormous n-simplex, where $n = |X|$.

3. The Čech complex is massive at large r, both in size and in dimension. As we will see later, most of the information it provides is also redundant, in that we can come up with a much smaller complex with the same homotopy type if we are little smarter.

As stated above, it is often a little nasty to compute the Čech complex. A much more convenient object is the following.

2.2 Vietoris-Rips Complex

Given a point cloud X and a fixed number $r \geq 0$, we define the Vietoris-Rips complex of X and r to be:

$$\mathrm{VR}(X, r) = \{\sigma \subseteq X \mid B_r(x) \cap B_r(x) \neq \emptyset, \forall x, y \in \sigma\}.$$

In other words, $\mathrm{VR}(X, r)$ consists of all subsets of X whose diameter is no greater than $2r$. From the definitions, it is trivial to see that $\check{\mathrm{C}}(X, r) \subseteq \mathrm{VR}(X, r)$. In fact, the two subcomplexes share the same 1-skeleton (vertices and edges) and the Rips complex is the largest simplicial complex that can be formed from the 1-skeleton of the Čech. In other words, the Rips complex will in general be even larger than the Čech. Of course, for $r < r'$, we again have the inclusion $\mathrm{VR}(X, r) \subseteq \mathrm{VR}(X, r')$.

Choosing Rips complex over Čech is not without cost. The penalty for this simplicity is that it is not immediately clear what is encoded in the homotopy type of the complex. Example in Figure 1 shows that the Rips complex does is not necessarily homotopy equivalent to the union of balls.

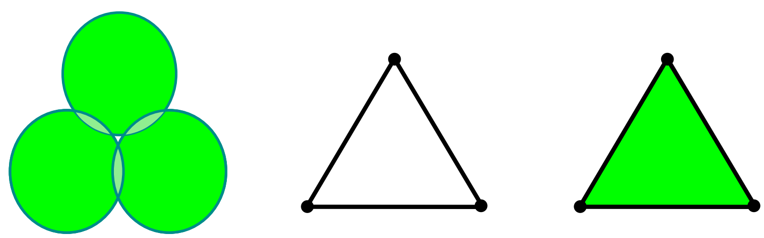


Figure 1: Observe a metric space X with three pairwise equidistant vertices. The Čech and Vietoris-Rips complexes of the covering on the left each have three edges. The Čech complex does not have any 2-simplices since the sets do not have a common intersection. The Rips complex, however, includes the triangle since all points have pairwise distance less than twice the radius of the balls.

If X is a finite subset of $\mathbb{R}^n$, with the standard metric, then

$$\check{C}(X, r) \subseteq \mathrm{VR}(X, r) \subseteq \check{C}(X, 2r).$$

Intuitively, the Rips complex of X is nested between two Čech complexes for X. For this reason we can take it to be a fairly good approximation.

2.3 Delaunay Triangulation

Even the Vietoris-Rips complex is computationally expensive, though, due to the fact that its vertex set consists of the entire metric space in question. We now give a construction which drastically limits the number of simplices, as well as reducing the dimension to that of the ambient space for points in general position. Given a finite point set $X \subseteq \mathbb{R}^d$, we define the *Voronoi cell* of a point $p \in X$ to be:

$$V_p = \{x \in \mathbb{R}^d \,|\, \mathrm{d}(x, p) \leq d(x, q), \forall q \in X\}.$$

The collection of all Voronoi cells is called the Voronoi diagram of X; we note that it covers the entire ambient space $\mathbb{R}^d$.

We then define the Delaunay triangulation of S to be (isomorphic to) the nerve of the collection of Voronoi cells; more precisely,

$$\mathrm{Del}(X) = \{\sigma \subseteq X \mid \bigcap_{p \in \sigma} V_p \neq \emptyset\}.$$

We note that a set of vertices $\sigma \subseteq X$ forms a simplex in $\mathrm{Del}(S)$ iff these vertices all lie on a common $(d-1)$-sphere in $\mathbb{R}^d$. Assuming general position, we do in fact get a simplicial complex.

2.4 Alpha Complexes

We again let X be a finite set of points in $\mathbb{R}^d$ and fix some radius r. As seen above, the complex $\check{\mathrm{C}}(X, r)$ has the same homotopy type as the union of r-balls X_r, but requires far too many simplices for large r. We now define a much smaller complex, $\mathrm{Alpha}(X, r)$, which is geometrically realizable in $\mathbb{R}^d$, and gives the correct homotopy type.

First, for each $p \in X$, we intersect the r-ball around p with its Voronoi region, to form $R_r(p) = B_r(p) \cap V_p$. These sets are convex and their union still equals X_r. We then define the Alpha complex of X and r to be the nerve of the collection of these sets:

$$\mathrm{Alpha}(X, r) = \{\sigma \subseteq X \mid \bigcap_{p \in \sigma} R_r(p) \neq \emptyset\}$$

By the Nerve Theorem, $\mathrm{Alpha}(X, r)$ has the same homotopy type as X_r. On the other hand, it is certainly much smaller, both in cardinality and dimension, than $\check{\mathrm{C}}(X, r)$; for example, $\mathrm{Alpha}(X, r)$ will always be a subcomplex of $\mathrm{Del}(X)$, and thus can have dimension no larger than the ambient dimension. As usual, we have inclusions $\mathrm{Alpha}(X, r) \subseteq \mathrm{Alpha}(X, r')$, for all $r < r'$, and we note that $\mathrm{Alpha}(X, \infty) = \mathrm{Del}(X)$.

2.5 Witness Complexes

The definition of the Delaunay complex makes sense for any metric space, in particular for finite metric spaces. However, for finite metric spaces, it generically produces degenerate (i.e. discrete) complexes, with no one dimensional simplices. This is due to the fact that for finite metric spaces, it is generically the case that each value of the distance is taken only for one pair of points, so one does not have any points which are equidistant between a pair of landmarks. In order to make the method useful for finite metric spaces, we therefore modify the definition of the Delaunay complex to accommodate pairs of points which are 'almost' (as permitted by the introduction of a parameter ϵ) equidistant from a pair of landmark points.

Let X be any metric space, and suppose we are given a finite set L of points in X, called the landmark set, and a parameter $\epsilon > 0$. For every point $x \in X$, we let m_x denote the distance from this point to the set L, i.e. the minimum distance from x to any point in the landmark set. Then we define the strong witness complex attached to this data to be the complex $\mathrm{W}^s(X, L, \epsilon)$ whose vertex set is L, and where a collection $\{l_0, \ldots, l_k\}$ spans a k-simplex if and only if there is a point $x \in X$ (the witness) so that $\mathrm{d}(x, l_i) \leq m_x + \epsilon$ for all i. We can also consider the version of this complex in which the 1-simplices are identical to those of $\mathrm{W}(X, L, \epsilon)$, but where the family $\{l_0, \ldots, l_k\}$ spans a k-simplex if and only if all the pairs (l_i, l_j) are 1-simplices. We will denote this by $\mathrm{W}^s_{VR}(X, L, \epsilon)$ and call it the *lazy witness complex.* .

There is a modified version of this construction, which is quite useful, called the weak witness construction. Suppose we are given a metric space X, and a set of points $L \subseteq X$. Let $\Lambda = \{l_0, \ldots, l_k\}$ denote a finite subset of a metric space L. We say a point $x \in X$ is a weak witness for Λ if $\mathrm{d}(x, l) \geq \mathrm{d}(x, l_i)$ for all i and all $l \notin \Lambda$. Given $\epsilon \geq 0$, we will also say that x is an ϵ-weak witness for Λ if $\mathrm{d}(x, l) + \epsilon \geq d(x, l_i)$ for all i and all $l \notin \Lambda$.

Let X, L, and ϵ be as above. We construct the weak witness complex $\mathrm{W}^w(X, L, \epsilon)$ for the given data by declaring that a family $\Lambda = \{l_0, \ldots, l_k\}$ spans a k-simplex if and only if Λ and all its faces admit ϵ weak witnesses. This complex also clearly has a version in which a k-simplex is included as a simplex if and only if all its 1-faces are, and we will denote this version by $\mathrm{W}^w_{VR}(X, L, \epsilon)$.

It is clearly the case that if we have $0 \leq \epsilon \leq \epsilon'$, then we have an inclusion

$$\mathrm{W}^s(X, L, \epsilon) \to \mathrm{W}^s(X, L, \epsilon')$$

and similarly for W^s_{VR} , W^w, and W^w_{VR}.