PERSISTENT COHOMOLOGY OF COVER REFINEMENTS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

MATHEMATICS

JULY 2022

By

Oleksandr Markovichenko

Dissertation Committee:

Yuriy Mileyko, Chairperson
Monique Chyba
Asaf Hadari
Elizabeth Gross
George Wilkens
Guylaine Poisson

ACKNOWLEDGMENTS

Firstly, I would like to express my deepest gratitude to my advisor, Dr. Yuriy Mileyko, for all his help, guidance, and friendship throughout all these years. This was a beautiful journey.

Also, I'm grateful to all my mentors who helped me to get where I am right now, including, but not limited to, Bogdan Rublev, Abdulaeva Natalia, and Andriy Anikushin.

Special thanks to all my friends for their care and love. Especially to Yulia.

Last but not least, I'm grateful to ZSU and all other people who support and defend my country in such a difficult time. Slava Ukrayini!

ABSTRACT

Topological data analysis (TDA) is a new approach to analyzing complex data which often helps reveal otherwise hidden patterns by highlighting various geometrical and topological features of the data. Persistent homology is a key in the TDA toolbox. It measures topological features of data that persist across multiple scales and thus are robust with respect to noise. Persistent homology has had many successful applications, but there is room for improvement. For large datasets, computation of persistent homology often takes a significant amount of time. Several approaches have been proposed to try to remedy this issue, such as witness complexes, but those approaches present their own difficulties.

In this work, we show that one can leverage a well-known data structure in computer science called a cover tree. It allows us to create a new construction that avoids difficulties of witness complex and can potentially provide a significant computational speed up. Moreover, we prove that the persistence diagrams obtained using our novel construction are actually close in a certain rigorously defined way to persistence diagrams which we obtain using the standard approach based on Čech complexes. This quantifiable coarse computation of persistence diagrams has the potential to be used in many applications where features with a low persistence are known to be less important.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

In recent years, data of various kinds have been collected at an extremely fast rate, and it has become clear that the standard techniques of data analysis are not always capable of handling the severe growing complexity of these data. Consequently, there has been active development of alternative tools for complex data analysis, in particular the ones employing methodology from combinatorial and algebraic topology, as they can potentially provide a better insight into the patterns hidden within the data than the standard tools and methods of data analysis. Topological data analysis (TDA) has thus become a fast-growing field which has already developed a number of topological and geometrical tools that can help us understand some important features of various complex data sets, in particular the features related to their shape.

Of course, this immediately brings up the question of what we mean by the "shape" of a discrete data set. Invoking the intuition behind a "zoomed out view" of data, we can think of a data set as representing a topological space obtained by drawing a ball with some fixed radius around each point and then taking the union of all these balls. However, it may not be at all clear how one should choose the radius for such balls. Hence, instead of fixing the radius we can consider all possible radii and get a family of topological spaces. When the radius changes from 0 to ∞ we can see, how the shape of these spaces changes, and how their topological features, e.g. holes of various dimensions, get created and destroyed during this process. While the field of topology offers a number of ways to characterize the shape of an object (within the category of topological spaces and continuous maps), one of the most computationally efficient yet descriptive ways is to compute its homology (or cohomology) groups. Of course, computational consideration also requires one to represent the union of balls using a more convenient structure, and it turns out such a structure is an appropriately constructed simplicial complex. This is the main idea behind one of the most widely used tool in TDA – persistent homology. This concept was initially introduced by Edelsbrunner et al. [10]. It has found a lot of applications in different areas: Bio-Science [9, 13]; sensor networks [6, 14]; analysis of breast cancer [18, 19]; and many others (see [16] for more examples)

With all of its successes, the persistent homology method does have some drawbacks. When the number of points and/or dimensions of the underlying space is too large, the computations may turn out to be prohibitively expensive. One can note, however, that if the usual set-up for persistent homology is used, i.e. when radii of the balls change from 0 to ∞, then in the beginning, for small radii, we compute a lot of features that capture "noise" – small holes, which do not carry any information about the true shape of the space underlying the data set. On the other hand, one can begin computations from the other end, that is, start with the balls of a very large radius and then decrease it to zero, using persistent cohomology instead of homology. However, in this case, the immediate computational slow down will be caused by the extremely large number of simplices needed to correctly represent the union of the balls.

The one possible way to avoid these problems is to use witness complexes, introduced in [5]. The idea behind the witness complexes is to select an appropriately chosen small subset of the data points, called landmarks, as a vertex set of a family of nested simplicial complexes and use all of the data points, referring to them as witnesses, to determine which simplices should be added to this family as the radius increases to infinity. Thus constructed family of witness simplicial complexes can be used to employ the usual persistent homology algorithm. There are variations of this approach, and its applications to real-world data can be found in [7, 8].

An astute reader can notice that the witness complex approach presents another problem: How many landmark points should one select and how does one go about selecting them? Several approaches have been suggested, but with very limited theoretical justification. Moreover, there are only very limited theoretical results allowing us to compare the results of the persistent homology computations using witness complexes and the ones obtained using simplicial complexes over the full set of data points.

In this work, we develop a new method for computing persistent homology and cohomology which allows us to circumvent the aforementioned problems for the price of some accuracy. This approach is based on the observation that a well-known computer science data structure called "cover tree" (which is mostly used for finding nearest neighbors) gives us some sort of a "discrete" set of covers of our data set, meaning that each element of a cover is represented as a point with descendants, thus giving us a discrete neighborhood. Importantly, it is a leveled data structure, that is, it has multiple levels, each containing a different number of points (subsets), starting from a single point at the highest level, and ending with all data points at a lower level. Furthermore, covers at lower levels can be regarded as refinements of covers at higher levels. Similar to the methodology described earlier, we were able to develop a technique for constructing continuous/combinatorial representation of the underlying unknown structure from these discrete covers. More specifically, we have developed a method for constructing a simplicial complex aimed at capturing the nerve of the unknown continuous cover underlying the discrete covers at each level, along with simplicial maps between the levels. Consequently, the persistent (co)homology algorithm can be applied to our construction. A central result of this work is to show that the results of the persistent (co)homology computations obtained using our novel construction are close to the results obtained using the standard approach. The rigorous formulation of this result employs the notion of interleaving between persistence modules obtained using the two approaches. Intuitively, this implies that the homological features computed with either of the two approaches appear and disappear at around similar scales.

CHAPTER 2
BACKGROUND AND PRELIMINARIES

The main premise of the topological data analysis is that data have shape and that shape matters. A typical assumption is that observed samples lie around a subspace (e.g a submanifold) of Euclidean space, and the goal is to get some insight into the topological properties of that space. Of course, trying to assess topology when given a set of discrete points leads to an immediate question of how those points should be connected. A well-established approach is to construct a simplicial complex using the given points as a vertex set. Once such construction is done, one can appeal to computationally tractable topological tools like homology to get an idea about the shape of the object of interest. However, when implementing this approach one runs into the problem of deciding which simplices should actually be added to our complex. One possible way of circumventing this issue is to use the concept of persistent homology. To understand our main result the reader needs to be fairly well familiar with certain families of simplicial complexes and persistent homology theory. Therefore, to keep thing mostly self contained, we will now cover the relevant background information.

2.1 Complexes, Homology, and Persistence

We start with the definitions of geometric and abstract simplices and simplicial complexes. More details about this topic can be found in [11]

2.1.1 Simplicial Complex

Let x_0, x_1 ... x_k be a set of affinely independent points in $\mathbb{R}^n$. A (geometric) **k-simplex** is a convex hull of $k+1$ affinely independent points, $\sigma = \mathrm{conv}\{x_0, x_1, ..., x_k\}$. Its dimension is $\dim \sigma = k$. Simplices of small dimension are typically referred to using specific names: 0-simplex is called a vertex, 1-simplex – an edge, 2-simplex – a triangle, and 3-simplex – a tetrahedron (Figure 2.1). A **face** of simplex σ is the convex hull of a subset of the points $x_0, \ldots, x_k$.

Definition 2.1. A (geometric) **simplicial complex** is a finite collection of simplices K, such that $\sigma \in K$ and γ is a face of σ implies $\gamma \in K$, and if σ_1, $\sigma_2 \in K$ then $\sigma_1 \bigcap \sigma_2$ is either empty or a face of both σ_1 and σ_2.

To not restrict ourselves only to Euclidean space $\mathbb{R}^n$, we can also think about simplicial complexes

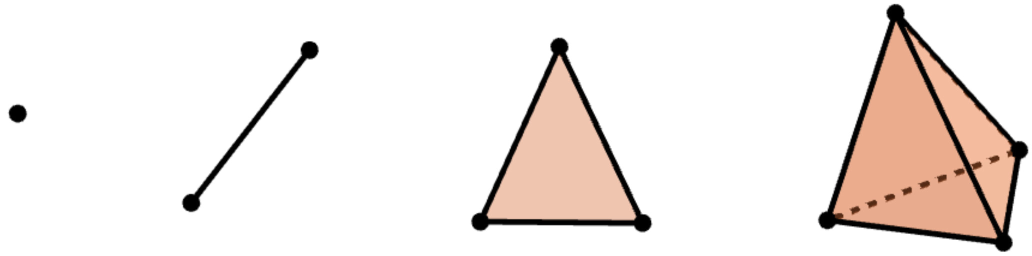

Figure 2.1: An example of simplices, from left to right: vertex; edge; triangle; tetrahedron

in a more abstract way:

Definition 2.2. An **abstract simplicial complex** is a finite collections of sets A, such that $\alpha \in A$ and $\beta \subset \alpha$ implies $\beta \in A$.

The sets in A represent simplices in the simplicial complex. If we have a geometric simplicial complex K, we can construct an abstract simplicial complex A by only considering the sets of vertices of each simplex in K. It is also well known that every abstract simplicial complex of dimension d has a geometric realization in $\mathbb{R}^{2d+1}$ (see e.g. [11]). Thus, we can still keep in mind the geometric picture when dealing with abstract simplicial complexes.

Returning to the original problem of connecting discrete data points, we now see that constructing a simplicial complex over such points is a viable approach, since we obtain an object possessing non-trivial topological and geometric properties. But what points should we connect into simplices? The intuition should tell us that points which we connect should not be far away from each other. Let's assume that we are given points form a Euclidean space. It seems reasonable to connect a set of points into a simplex if all of those points lie inside a ball of some radius $r > 0$, where r represents our threshold for "closeness". It turns out that the resuliting collection of simplicies does form a simplicial complex, which is called a **Čech complex** at scale r. Notice that a set of points being contained inside a ball with radius r is equivalent to having a nonempty intersection of the balls of radius r centered at these points.

To formalize the above discussion, consider a metric space $(\mathbb{X}, d)$ (e.g. the Euclidean space with the usual metric). Let's denote by $B_x(r) = \{y \in \mathbb{X} \mid d(x, y) \leq r\}$ a closed ball with radius r around a point x. The Čech complex over a set $S \subset \mathbb{X}$ at scale r is the abstract simplicial complex $\check{C}ech(r) = \{\sigma \subset S \mid \bigcap_{x \in \sigma} B_x(r) \neq \emptyset\}$ (Figure 2.2).

Note, that if $r_1 < r_2$ then points lying inside a ball of radius r_1 also lie inside the ball of radius

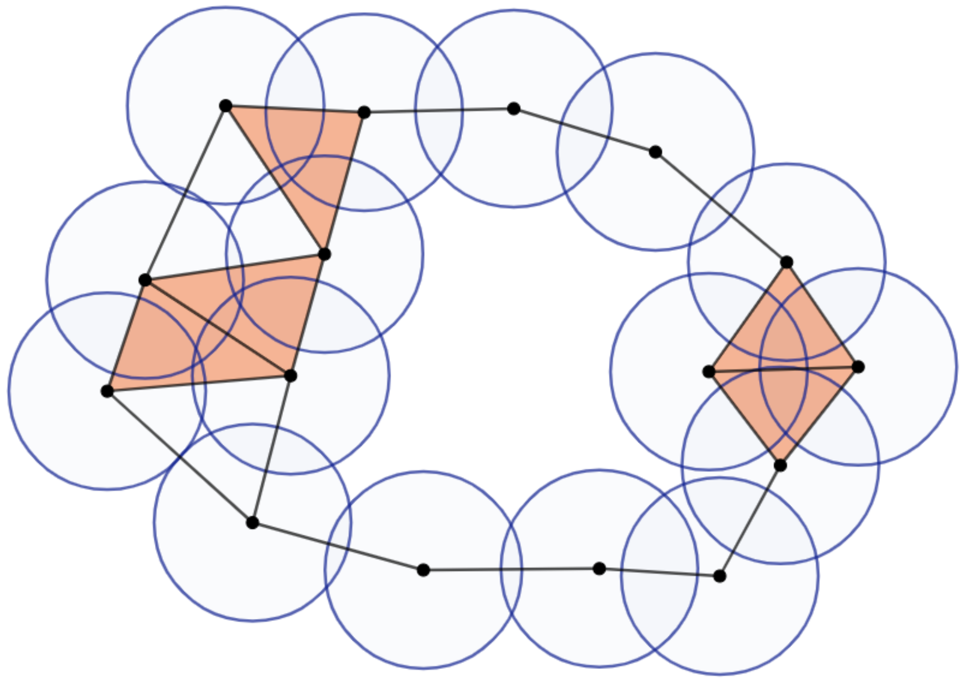

Figure 2.2: An example of construction of a Čech complex

r_2 with the same center. This implies that $\check{C}ech(r_1) \subseteq \check{C}ech(r_2)$. Thus, increasing the radius r from zero to infinity yields a nested family of vCech complexes. This property will enable us to employ persistent homology, as we discuss later.

Regarding the construction of vCech complexes from a practical viewpoint, we are faced with the problem of finding the smallest enclosing ball for a given set of points. This is a non trivial problem, and it can be computationally challenging, especially if we are working in a non-Euclidean space or a Euclidean space of a high dimension. One possible approach to simplify the construction is to connect a set of points into a simplex if all the pairwise distance are less or equal to $2r$, which is equivalent to saying that the edges of the simplex belong to the Čech complex at scale r. The resulting collection of simplices forms a simplicial complex knows as the **Vietoris-Rips complex**: $\textit{Vietoris-Rips}(r) = \{\sigma \subset S \mid d(x, y) < 2r \forall x, y \in \sigma\}$ (Figure 2.3). Thus, in order to build a Vietoris-Rips complex at scale r we only need to know pairwise distances between the points, which is clearly easier to compute that smallest enclosing balls.

One can see directly from the definitions that for every r, $\check{C}ech(r) \subset \textit{Vietoris-Rips}(r)$. Indeed, edges are added to both constructions using the the same criterion. For simplices of higher dimensions, two points lying inside a ball with radius r implies that the distance between them is less $2r$.

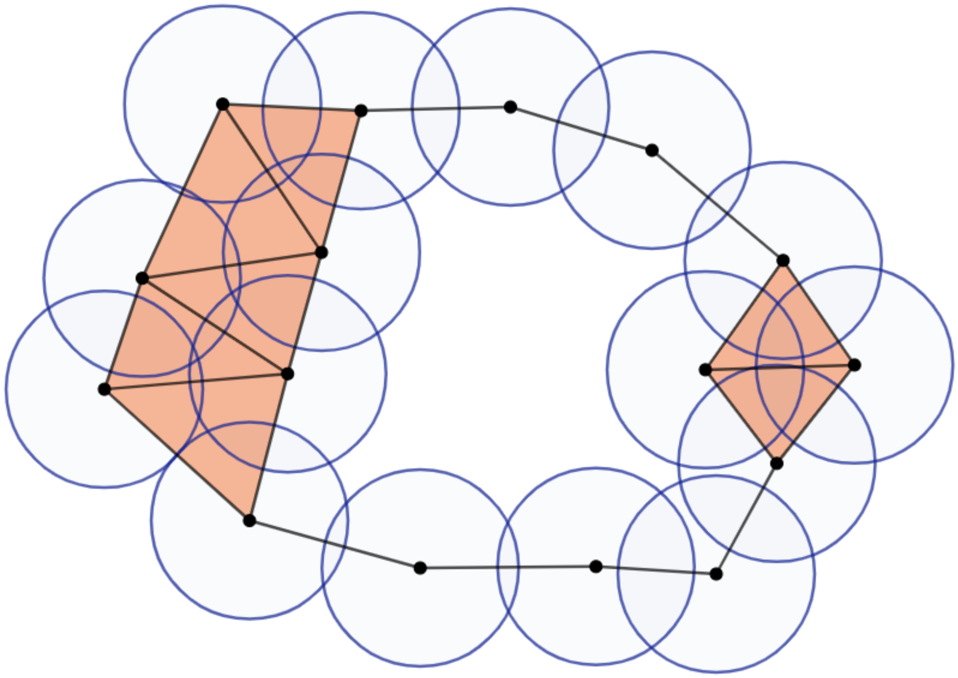

Figure 2.3: An example of construction of a Vietoris-Rips complex

Hence, for any simplex $\sigma \in \check{C}ech(r)$ pairwise distances between points in σ are less than or equal to $2r$, yielding $\sigma \in \textit{Vietoris-Rips}(r)$. We shall make use of this property later. Also, one can easily show that similarly to Čech complexes, if $r_1 < r_2$ then $\textit{Vietoris-Rips}(r_1) \subset \textit{Vietoris-Rips}(r_2)$.

Constructing Vietoris-Rips complexes over data points is a computationally viable task, letting us to obtain an object with topological and geometric properties. However, it is still not clear how to chose an appropriate radius r. Before delving deeper into this topic, let's first talk about topological properties which we are trying to recover. They are homology groups.

2.1.2 Homology Groups

Intuitively, homology groups give us information about "holes" in the topological object. In Figure 2.4 we can see an annulus. We can think about the dark blue circle as a "representative" of the hole in the middle. But intuitively, the light blue circle should represent the same hole, since we can continuously deform one circle into the other. We also note that the purple circle in the figure is of a different kind. It does not represent any hole since it can be continuously deformed into a point. So, this trivial example suggests that every circle (or closed curve) is either contractible or represents some hole, and we need to understand which closed curves represent the same hole. Generalizing these ideas leads to the topic of homotopy theory, which is a powerful framework for describing the shape of topological spaces. The issue is that it is not computationally tractable. Remembering that we have to deal with simplicial complexes, we can try to capture similar ideas using simplicial constructions, which leads to the notion of homology groups.

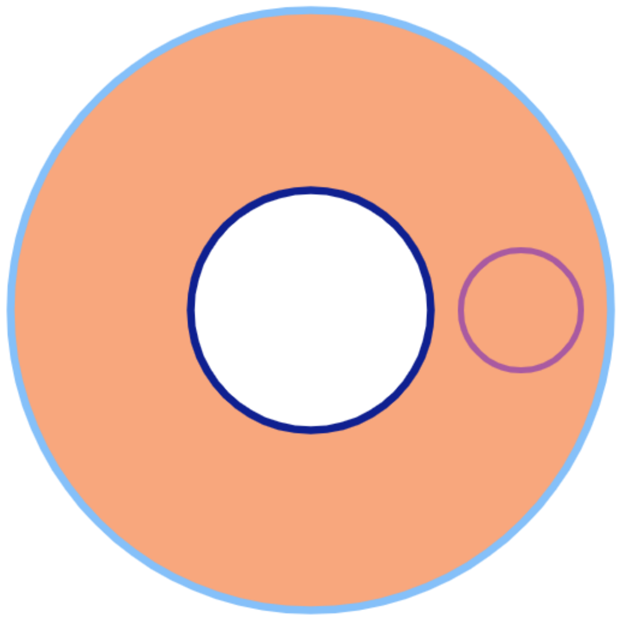

Figure 2.4: Circles as representatives of holes

Let K be a simplicial complex. A **p-chain** is a formal sum: $\sum a_i \sigma_i$, where σ_i is simplex of dimension p and a_i are coefficients. In computational settings these coefficients are often restricted to the values of 0 or 1, i.e. elements of $\mathbb{Z}_2$. In general, these coefficients can belong to a more complicated algebraic structure which can be such be a field, a ring, or a group. The standard choice is the set of integers, $\mathbb{Z}$. We can add different p-chains like polynomials. With this addition operation, p-chains form a group, denoted by $C_p = C_p(K)$. For convenience, in all of the examples in this section we will assume that the coefficients are elements of $\mathbb{Z}_2$.

Given a simplex σ with vertices $x_0, x_1, ..., x_p$, $\sigma = [x_0, x_1, ..., x_p]$ we can define a boundary operation: $\partial_p \sigma = \sum_{j=0}^{p} (-1)^j [x_o, x_1, ..., \widehat{x_j}, ..., x_p]$, where $[x_0, x_1, ..., \widehat{x_j}, ..., x_p]$ is a simplex of dimension $p-1$ with the vertex set $\{x_0, x_1, ..., x_p\} \setminus \{x_j\}$. For example, the boundary of a tetrahedron is sum of 4 triangles, the boundary of a triangle is sum of 3 edges – "sides" of the triangle, the boundary of an edge is the sum of two vertices – the endpoints, and the boundary of a vertex is the trivial chain, 0. The boundary of a p-chain is obtained by linearly extending the above operation, i.e. it is the sum of the boundaries of the corresponding simplices (multiplied by the corresponding coefficients, in general). It is a $(p-1)$-chain which we are going to call a **(p-1)-boundary**. All p-boundaries also form a group, denoted by B_p. The map $\partial_p : C_p \to C_{p-1}$ is a homomorphism which we shall refer to as a **boundary map** and, for convenience, will often drop the index p. The collection of all the chain groups and boundary maps between them is called a **chain complex**, and is typically denoted C_*.

Coming back to our toy example of an annulus, we can think about what kind of a chain would represent a hole if the annulus were represented by a simplicial complex. Intuitively, it should be a chain of edges forming a "cycle", i.e. with any two edges having a single endpoint in common (Figure 2.5). If we consider the boundary of this chain we note that the boundary of each edge consists of two vertices, but each vertex will be in the boundary of exactly two edges, thus canceling out since we are talking about $\mathbb{Z}_2$ group. Consequently, the boundary will be equal to 0.

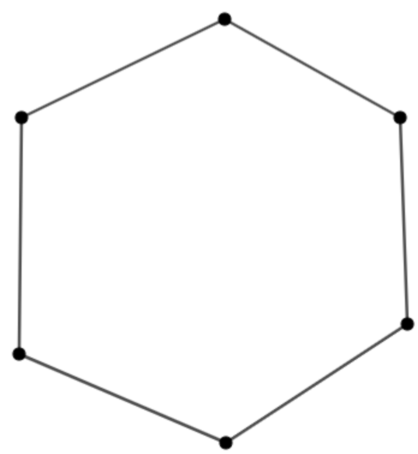

Figure 2.5: An example of a cycle of dimension 1

We can generalize these ideas to higher dimensions. Suppose our simplicial complex consists of exactly four triangles – faces of a tetrahedron, but not the tetrahedron itself. We can think about the void inside as a 2-dimensional hole, and we can view the collection (sum) of these 4 triangles as a representative of this hole. Let's find the boundary of this sum of 4 triangles. The boundary of each triangle consists of three edges, but each edge belongs to exactly two triangles. Again, since we are working with $\mathbb{Z}_2$ coefficients, the boundary will be zero. This gives us the idea of how to define a cycle. A **p-cycle** is a p-chain $c \in C_p$, such that $\partial c = 0$. All of the p-cycles form a group Z_p, a subgroup of C_p.

Note that in the above example we considered a simplicial complex that was actually a boundary. In fact, by performing careful calculations we can show that the boundary of the boundary of any simplex, and hence of any p-chain, is 0 (see e.g. [11]).

Lemma 2.1 (Fundamental Lemma of Homology)**.** $\partial_p \partial_{p+1} d = 0$ for all p and all $(p+1)$-chains d

The above lemma states that a p-boundary is also a p-cycle. One can also show that the collection of all p-boundaries, B_p, is a subgroup of C_p, and hence a subgroup of Z_p (Figure 2.6). And now we have all we need to define homology groups.

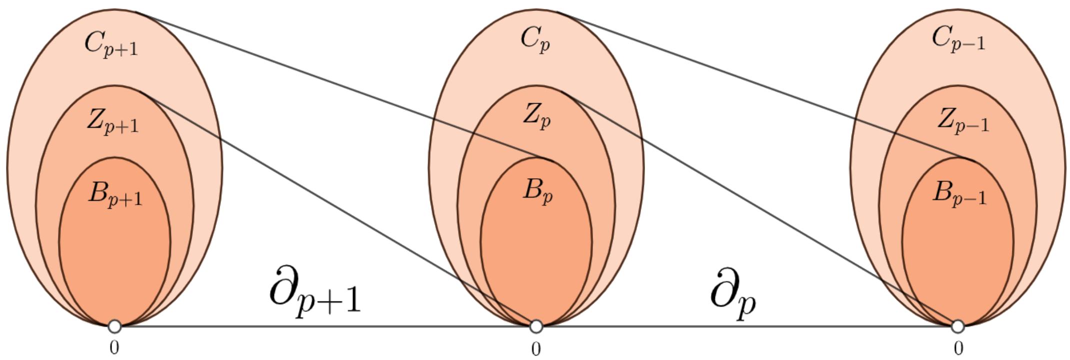

Figure 2.6: Homomorphism connections between chains, cycles and boundaries

Definition 2.3. The **p-th homology group** is the p-th cycle group modulo the p-th boundary group, $H_p = Z_p / B_p$. The **p-th Betti number** is the rank of this group, $\beta_p = \operatorname{rank} H_p$.

Intuitively, the p-th Betti number gives us the number of p-dimensional holes in a simplicial complex. Also, two cycles represent the same hole if they differ by a p-boundary, the boundary of a $(p+1)$-chain. Let's go back to our annulus example (again, assuming it is represented by a simplicial complex) and consider the boundary of the sum of all the triangles (which represent the interior of the annulus). This boundary will the sum of the two chains representing the dark blue and light blue circles. Since we are working with $\mathbb{Z}_2$, it implies that any one of these circles is the sum of the other one with the boundary of interior of the annulus. So, these two cycles differ by 1-boundary which puts them into the same homological class (element of 1-st homology group), as intended. The purple circle is the boundary of the small disk itself, so it is 1-boundary, hence it represents a zero element of the 1-st homology group, as intended. Thus, the 1-st homology group of the disk has the rank 1.

To illustrate further properties of homology groups, let us consider several more examples in Figure 2.7. It can be shown that the 0-dimensional homology group captures (path) connected components of the space. There is only one connected component in all of the examples in the figure, so $\beta_0 = 1$. In the Figure 2.7 (a) we have a 2-dimensional sphere. We can notice that every 1-cycle actually bounds a disk. In other words, every 1-cycle is boundary of some 2-chain, for example the red cycle is a boundary of the "hat" above it. So the H_1 is trivial. Also, the whole surface of the sphere (when represented by triangles) is a 2-cycle which is not the boundary of any 3-chain, since there are no 3-chains. Consequently, this cycle represents the only non-trivial element of H_2. In other words, the sphere has the only one "hole" of dimension 2. In the Figure 2.7 (b) we have a torus. It has 2 different cycles of dimension 1 representing nontrivial homology classes (red and green ones). We can notice that both red cycles belongs to the same homology class – they are the boundary of the tube connecting them. Since we are working in $\mathbb{Z}_2$, it implies that one is the sum of the other with the boundary of the tube. So $\beta_1 = 2$. Similarly to the previous example, the whole surface of the torus is a 2-cycle, which is not the boundary of any 3-chain, so $\beta_2 = 1$. Lastly, in Figure 2.7 (c) we have a solid torus. Now, the red cycles represent a trivial homology class, since it bounds a disk, i.e. the red cycles are boundaries themselves. The green one is still represents a nontrivial homology class, so $\beta_1 = 1$. The surface of the whole solid torus is still a cycle, but now that this cycle is a boundary of interior of the torus (represented by the sum of all the tetrahedra). Hence H_2 is trivial and $\beta_2 = 0$.

2.1.3 Persistent Homology

Armed with Vietoris-Rips and Čech complexes along with (simplicial) homology theory, we can now return to the problem of capturing the shape of a discrete set of data points. Let's consider a simple example of a sample from some subspace of the Euclidean plane shown in Figure 2.8.

Our goal is now more concrete – we are trying to recover homology groups (with $\mathbb{Z}_2$) of the (unknown) subspace from the given sample. Since we are working in the plane, we are interested in the 1-st homology group. We know how to construct a simplicial complex (Čech or Vietoris-Rips) at some scale $r > 0$, and we can compute its homology groups. For convenience, let's use the Vietoris-Rips complex. Thus, we add edges if the corresponding endpoints are within distance $2r$, and we add triangles if all their edges have been added. We show the resulting simplicial complex in Figure 2.9. We did not fill in the triangles to keep the picture visually simpler, but it should be

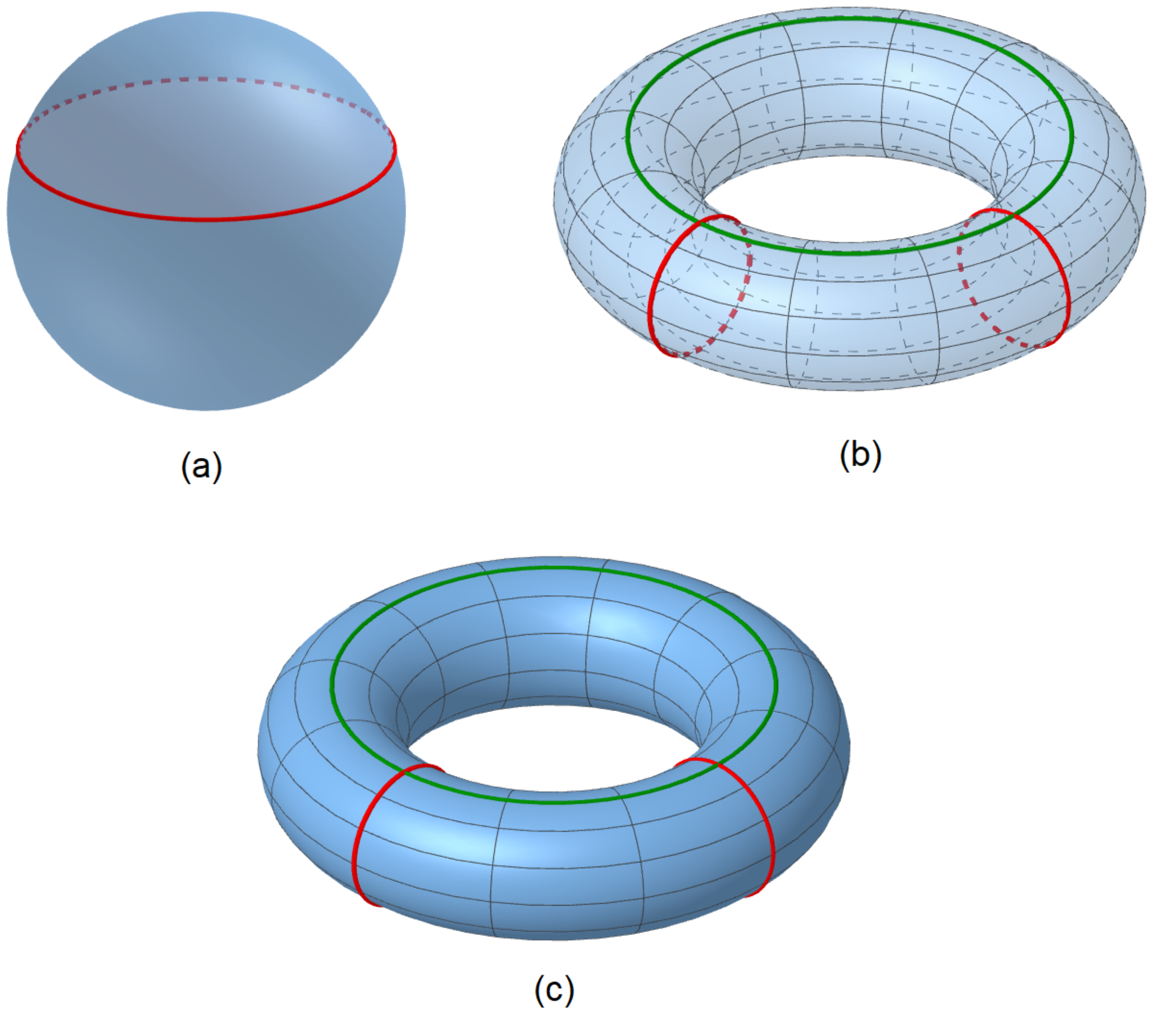

Figure 2.7: An example of homology groups and their representatives in: (a) - 2-sphere; (b) - torus; (c) - filled torus

clear which triangles are contained in the complex. We can also notice that there is a tetrahedron in this complex.

We got lucky as our simplicial complex does seem to capture the right connectivity, but in general it is unclear how to chose the radius r in the Vietoris-Rips construction. If r is too small, then every point will create a separate connected component (Figure 2.10 (a)). If r is too big, than we will get contractible complex. ((Figure 2.10 (b)).

The idea behind persistent homology is to not focus on any specific value of the radius, but to look at the all radii from 0 to ∞. As we discussed already, if $r_1 < r_2$, then $\textit{Vietoris-Rips}(r_1) \subseteq \textit{Vietoris-Rips}(r_2)$. Hence, we will get a family of nested simplicial complexes and, at least intuitively

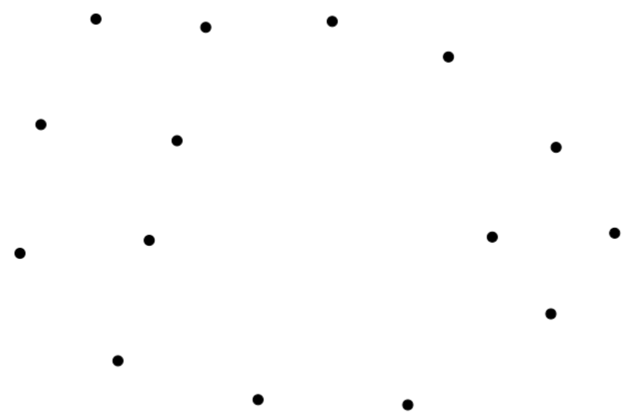

Figure 2.8: Some random data set

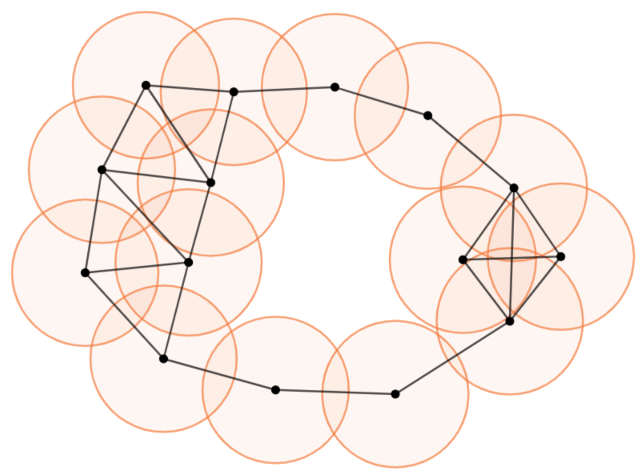

Figure 2.9: An example of Vietoris-Rips complex for some appropriate radius selection

for now, we can to track at what radius/scale a hole, i.e. a 1-cycle representing a non-trivial homology class, appears (is born) and at what scale it disappears (dies). Figure 2.11 illustrates this idea, as we can see that on the left there is a small non-trivial 1-cycle born, but it becomes a boundary, i.e. dies, on the right, as we increase the radius.

The subsequent birth-death process is illustrated in Figure 2.12. In picture (a), we can see that there is one more 1-cycle born, but then it is "split" into two cycles in picture (b). This split is nothing more but a birth of another 1-cycle representing a different homology class. And in the

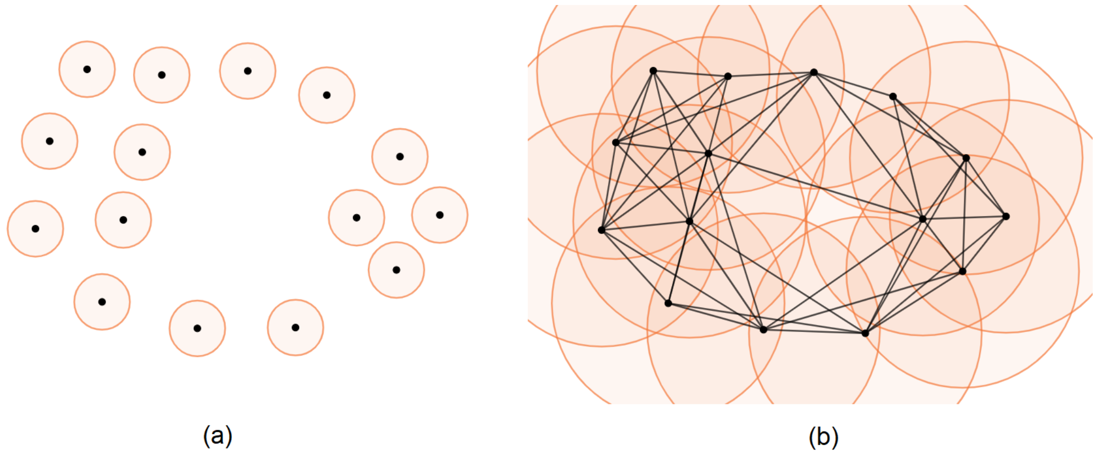

Figure 2.10: An example of Vietoris-Rips complex for some inappropriate radius selection: (a) - the radius is too small; (b) - the radius is too big

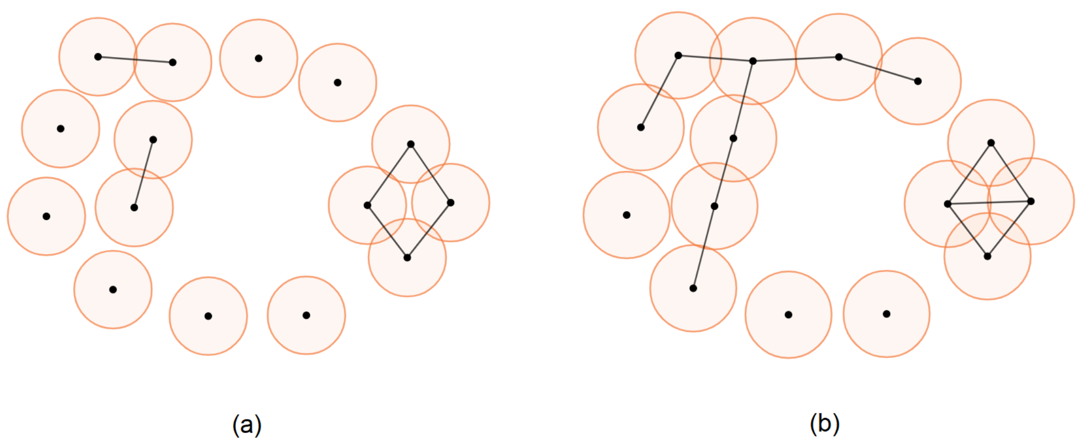

Figure 2.11: The birth (a) and death (b) of the first hole

picture (c) we can see that one of these cycles, the top one, died because two triangles were added to the complex. Keeping in mind that we are actually interested in homology classes, not just cycles, we should ask ourselves how to decide which of the two homology classes dies: the one that born firstly or the one that born lastly? In this simple example we can employ **elder rule** – the younger homology class dies and the older one continues to live. In picture (d), we can see how the older one dies as well.

In the Figure 2.13 (a) we can see how the new big hole is born, and dies on picture (b).

In general, the elder rules works only for 0-dimensional homology classes. The formal tracking of births and deaths of higher dimensional homology classes is much more complicated, but can be done when considering homology with field coefficients (see [11] for details). As a result, for each

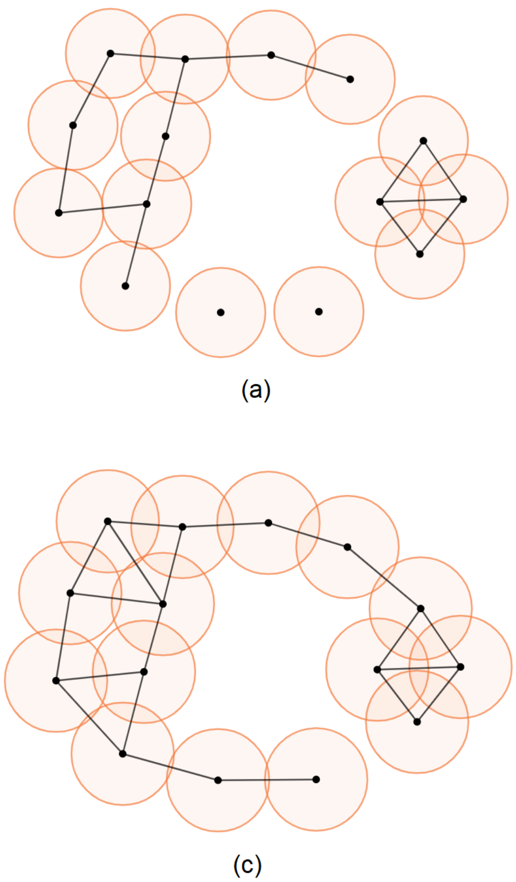

Figure 2.12: The births and deaths of the second and third holes

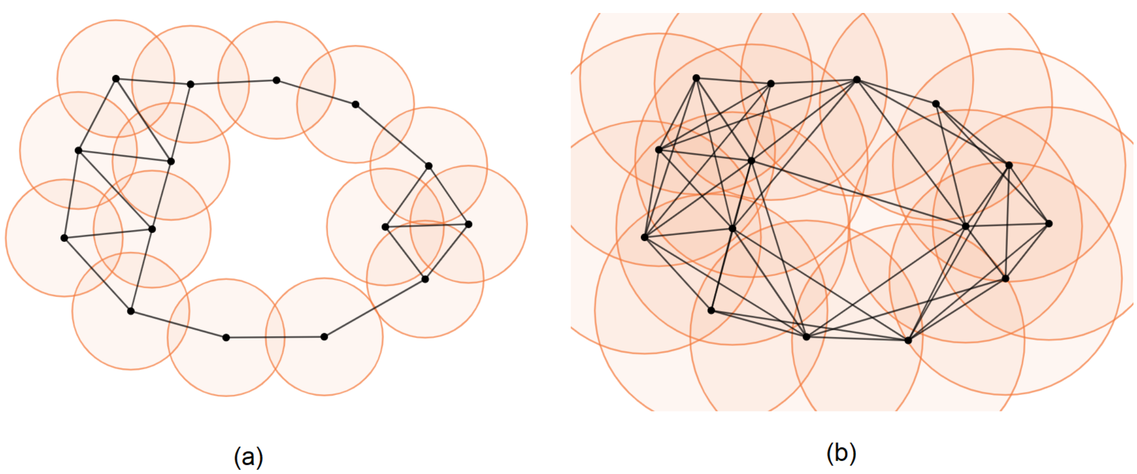

Figure 2.13: The birth and death of the last fourth hole

nontrivial homology class we can say when it is born and when it dies. Assuming that b and d denote the birth and the death of some homology class, we can regard these values as coordinates of a point in the plane, (b, d). The collection of all such points constitutes a **persistence diagram**, with the difference $d - b$ being the persistence of the corresponding homology class. Obviously, for any point in the diagram we have $d \geq b$, so no point lies below the diagonal. In Figure 2.14 we can see four points corresponding to the homology classes that discussed earlier. If the point is close to the diagonal, i.e. has small persistence, then the corresponding homology class lived for a short rage of scales. Oftentimes, such homology classes represent noise. By contrast, a point which is far away from the diagonal, i.e. has large persistence, represent a homology class that lived across a large range of scales. Such a homology class may correspond to a true homology class of the underlying topological space. In general, the distinction between noise and true homology may be illusive, but there are results stating that under certain conditions one can determine which points in a persistence diagram represent true homology of the underlying space (see [11] for details).

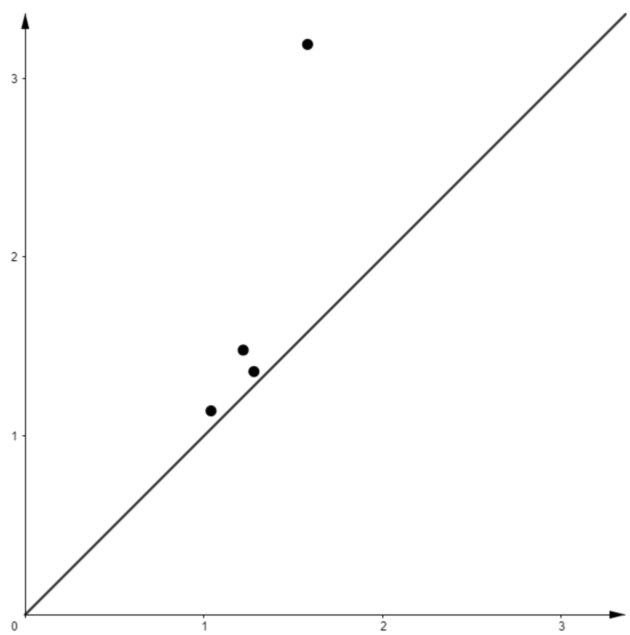

Figure 2.14: The persistence diagram of the example above

The ideas described above are a part of the **persistent homology** theory. We shall now describe it more more detail.

First, we need to define a **simplicial map**. Given simplicial complexes A and B, a map $f : A \to B$ is called simplicial if it maps the vertex set of A into the vertex set of B and for any simplex

$\sigma \in A$ we have $f(\sigma) \in B$. This is equivalent to saying that the image of the set of vertices of a simplex is a set of vertices of a simplex. It may be useful to consider an example of a map that is not simplicial. Suppose a simplicial complex A consists of three points: A_1, A_2, A_3 and a triangle on these points (with all its faces), and let a simplicial complex B consist of three points B_1, B_2, B_3 with only three edges connecting these points (Figure **??**). If f maps A_1 to B_1, A_2 to B_2, and A_3 to B_3, then this is not a simplicial map. Indeed, if σ is the triangle in A, then $f(\sigma)$ should be the simplex with vertices B_1, B_2, B_3, but there is no such simplex in B. Note that mapping A_1 to B_1, A_2 to B_2, and A_3 to B_2 yields a simplicial map which maps the triangle to the edge with vertices B_1, B_2.

It is a well known fact (see e.g. [11]) that a simplicial map $f : K_1 \to K_2$ between simplicial complexes K_1 and K_2 induces homomorphisms between the corresponding homology groups, i.e. we get a homomorphism $f_{p*} : H_p(K_1) \to H_p(K_2)$ for each homological dimension p. The idea of the proof is that the simplicial map f induces a map between p-chains which maps cycles to cycles and boundaries to boundaries. Hence it induces the map between quotient groups, i.e. p-th homology groups. For concreteness, we shall restrict ourselves to coefficients in $\mathbb{Z}_2$, but the approach works with any fields coefficients. If the homological dimension is clear from the context, we denote the induced homomorphism simply by f_*.

Suppose we have the finite family of simplicial complexes K_i, $i = 1, \ldots, n$, and a family of simplicial maps $f_i : K_i \to K_{i+1}$. In TDA, one typically considers a nested family of simplicial complexes $K_i \subseteq K_{i+1}$, which is called a **filtration**. Vietoris-Rips complexes considered earlier are an example of a filtration. In such a case, all simplicial maps f_i are just inclusion maps. In general, we can consider any sequence of simplicial maps. Let $f^{i,j} = f^i \circ f^{i+1} \circ ... \circ f^{j-1}$ be the resulting simplicial map between K_i and K_j and let $f^{i,j}_{p*} : H_p(K_i) \to H_p(K_j)$ be the corresponding induced homomorphism. Note that we have the following sequence of homology groups

$$H_p(K_1) \xrightarrow{f^1_{p*}} H_p(K_2) \xrightarrow{f^2_{p*}} ... \xrightarrow{f^n_{p*}} H_p(K_n)$$

in each homological dimension p. Since we consider field coefficients, these are actually vector spaces connected by linear maps. Such a sequence is referred to as a **persistence module**.

Definition 2.4. The p-th persistent homology groups are the images of the homomorphisms induced by the simplicial maps, $H^{i,j}_p = \mathrm{im} f^{i,j}_{p*}$ for $0 < i < j \leq n$. The corresponding p-th persistent Betti numbers are the ranks of these groups, $\beta_p^{i,j} = \text{rank} H_p^{i,j}$.

The persistent homology group $H_p^{i,j}$ consists of the homology classes present in K_i and still alive at K_j. If γ is a class in $H_p(K_i)$, we say it was born at K_i if $\gamma \notin H_p^{i-1,i}$. Moreover, we say γ dies at K_j if it merges with some older class as we go from K_{j-1} to K_j, i.e. $f_{p*}^{i,j-1}(\gamma) \notin H_p^{i-1,j-1}$ but $f_{p*}^{i,j}(\gamma) \in H_p^{i-1,j}$ (Figure 2.15). If γ is born in K_i and died in K_j then the persistence is $\text{pers}(\gamma) = j - i$. When considering filtrations where the simplicial complexes K_i correspond to real parameters r_i, as in the case of Čech or Vietoris-Rips complexes, the persistence is typically defined as $\text{pers}(\gamma) = r_j - r_i$. Note that in the case of Čech or Vietoris-Rips complexes the values r_i correspond to the scales at which new simplices are added to the complex.

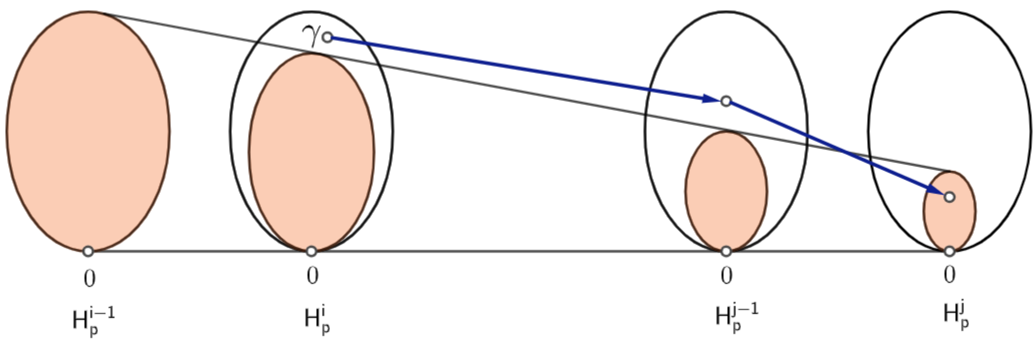

Figure 2.15: An example of a call γ which is born at K_i and dies entering K_j

Notice that $\beta_p^{i,j} - \beta_p^{i,j-1}$ is the number of homology classes that are born at of before K_i and die at K_j. Similarly, $\beta_p^{i-1,j} - \beta_p^{i-1,j-1}$ is the number of homology classes that are born at or before K_{i-1} and die at K_j. Thus, if we consider the difference of these two quantities we will get the number of homology classes that are born at K_i and die at K_j, $\mu_p^{i,j} = (\beta_p^{i,j} - \beta_p^{i,j-1}) - (\beta_p^{i-1,j} - \beta_p^{i-1,j-1})$. As we alluded to earlier, we can visualize births and deaths as points in $\mathbb{R}^{\not\vDash}$, with the x-coordinate corresponding to birth and the y-coordinate corresponding to death. Note that some points may have infinite y-coordinate, since a homology class may never die, and several homology classes can have the same birth and death values. Hence, we are looking at a multiset of points (i.e. points with multiplicities) in the extended $\overline{\mathbb{R}}^2$. These points together with all points from diagonal $x = y$, where each point on a diagonal has an infinite multiplicity, gives us the **p-th persistence diagram**. A point (r_i, r_j) in this diagram has multiplicity $\mu_p^{i,j}$, which is the number of homology classes it represents. The difference between coordinates is then the persistence of the corresponding homology classes. We will explain the need to include points on the diagonal shortly.

It is important to note that persistence diagrams allows us to recover persistent Betti numbers, $\beta_p^{i,j}$, and since we are considering homology with fields coefficients, this means that the diagrams encodes all the information about persistent homology group:

Lemma 2.2 (Fundamental Lemma of Persistent Homology)**.** Let $K_1 \xrightarrow{f_1} K_2 \xrightarrow{f_2} ... \xrightarrow{f_{n-1}} K_n$ be a simplicial filtration. For every pair of indices $1 \leq k \leq l \leq n$ and every homological dimension p, the p-th Betti numbers $\beta_p^{k,l} = \sum_{i \leq k} \sum_{j > l} \mu_p^{i,j}$.

Thus, to compare the shape of two data sets we can compare the homology of the corresponding simplicial filtrations by comparing the resulting persistence diagrams. To make such a comparison rigorous, we need to define a notion of the distance between persistence diagrams.

Definition 2.5. Let X and Y be two persistence diagrams. The **bottleneck distance** between these diagrams is defined as

$$W_\infty(X, Y) = \inf_{\nu : X \to Y} \sup_{x \in X} ||x - \nu(x)||_\infty,$$

where ν ranges over all bijections between X and Y.

One can see that the above bottleneck distance is well defined only if a set of bijections between any two diagrams is not empty. This is where our decision to include points on the diagonal to the diagrams comes into play, as it provides such a guarantee by allowing to us map off-diagonal points to the diagonal.

To make the bottleneck distance truly useful we need it to be close for the diagrams which are obtained for samples that are close (in Hausdorff distance). The corresponding result is given in [4], and actually covers a more general case of subsets of a metric space. An even more general result is given in [3], and it shows that the bottleneck distance for the diagrams obtained from close persistence modules remains close. We shall need a slightly modified and simplified version of this result, which we shall now state. Two persistence modules $V_1 \to \cdots \to V_n$ and $W_1 \to \cdots \to W_n$ are said to be 1-interleaved if there exist linear maps $\phi_i : V_i \to G_{i+1}$ and $\psi_i : W_i \to V_{i+1}$ such that for any integer $k \geq 0$ the following diagrams commute:

$$\begin{tikzcd} V_{i-1} \arrow[rrr] \arrow[dr, "\phi_{i-1}"'] & & & V_{i+k+1} \\ & W_i \arrow[r] & W_{i+k} \arrow[ur, "\psi_{i+k}"'] & \end{tikzcd} \qquad \begin{tikzcd} & V_i \arrow[r] & V_{i+k} \arrow[dr, "\phi_{i+k}"] & \\ W_{i-1} \arrow[ur, "\psi_{i-1}"] \arrow[rrr] & & & W_{i+k+1} \end{tikzcd}$$

The horizontal maps in the above diagrams are simply compositions of the connecting maps of the modules. The following lemma follows from the result in [3] regarding weakly interleaved persistence modules.

Lemma 2.3. If two persistence modules are 1-interleaved then the bottleneck distance between the corresponding persistence diagrams is bounded by 3.

2.1.4 Persistent Cohomology

Looking back at the earlier examples illustrating ideas behind persistent homology, we can notice that computations at small scales, which happen at the beginning of the standard persistent homology algorithm, are likely to mostly produce homological classes of small persistence that represent noise rather than true homology of the underlying space. That's because we expect relevant homology classes to have q relatively large persistence and hence die later in the process. Hence, we may ask ourselves if we could develop an alternative algorithm which also captures persistent homology information but instead of starting at the smaller scales and continuing to the larger scale it goes in the opposite direction, i.e. starts at the larger scales and moves towards the smaller scales. A possible advantage of such an approach would be capturing relevant persistent homology information earlier in the computations, thus possibly reducing the computational cost by simply avoiding computing persistent homology at smaller scales.

Theoretical underpinnings of the above "top-to-bottom" approach are provided by the theory of cohomology, which is a notion dual to homology. In cohomology theory, instead of considering a chain complex C_* one considers a cochain complex obtained by considering dual groups, $C^p = \mathrm{Hom}(C_p, G)$, where C_p are considered with integer coefficients and G is a fixed group, which, for simplicity, we once again assume to be Z_2, although the persistence framework can work with any field. Dualizing boundary maps $\partial_p : C_p \to C_{p-1}$ we obtain coboundary maps $\delta_{p-1} : C^{p-1} \to C^p$, and since $\partial\partial = 0$ we get $\delta\delta = 0$ as well. Hence, similarly to homology, we can define cocycle groups $Z^p = \ker \delta^p$ and coboundary groups $B^p = \mathrm{im}\ \delta^{p-1}$ (Figure 2.16). Consequently, we can define cohomology groups:

Definition 2.6. The p-th cohomology group is the quotient $\mathrm{H}^p = Z^p/B^p$ for all p.

Similarly to homology theory, a simplicial map $f : K_1 \to K_2$ induces a homomorphism between cohomology groups, $f^{p*} : H^p(K_2) \to H^p(K_1)$. Hence, given a sequence of simplicial complexes connected by simplicial maps the resulting cohomology groups and indiced homomorphisms also

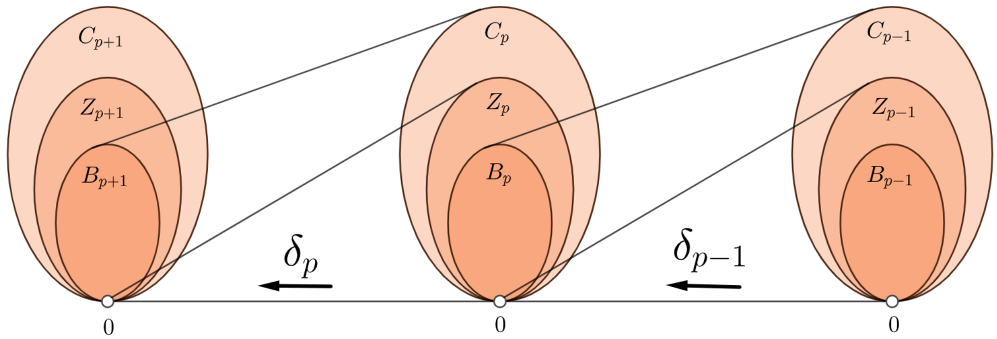

Figure 2.16: Homomorphism connections between cochains, cocycles and coboundaries

form a persistence module, except that arrows point in the opposite direction. As a result, we can define persistent cohomology as persistent homology of such a module, and it gives us maps going from larger scales to smaller scales.

The definition of cohomology groups suggests that there should be a relation between persistent homology and cohomology. It turns out that persistent cohomology captures the same information as persistent homology (see e.g. [11]):

Theorem 2.1 (Universal Coefficient Theorem.)**.** Given a simplicial complex K and a field G, there are maps $H^p(K) \to \mathrm{Hom}(H_p(K), G) \to H_p(K)$ in which the fist map is a natural isomorphism and the second is an isomorphism that is not natural.

Naturality of the first map means that if we have another simplicial complex L and a simplicial map $K \to L$ then the following diagram commutes:

$$\begin{array}{ccc} H^p(K) & \longrightarrow & \mathrm{Hom}(H_p(K), G) \\ \uparrow & & \uparrow \\ H^p(L) & \longrightarrow & \mathrm{Hom}(H_p(L), G) \end{array}$$

The second map is not natural because it depends on the choice of bases.

2.2 Cover Tree

It is important to note that existing algorithms for computing standard persistent homology and cohomology for a data set require us to compute the corresponding simplicial complexes, e.g. Čech or Vietoris-Rips, at the largest scale of interest. If the data set consists of a lot of points lying in a high dimensional space, then the sheer number of simplices may make the computation extremely expensive. A possible way to deal with this problem is to construct simplicial complexes whose vertex sets are subsets of the data points rather than the whole set. However, this approach raises two natural questions: how to choose the vertex sets of our simplicial complexes and how to determine which of the vertices span simplices? When considered previously, these questions led to the development of so called witness complexes [5]. However, the choice of the vertex set for these complexes, which is the same at all scales, is done in an ad hoc manner. Here, we take a different approach. We are going to employ a well-known data structure in computer science called "cover tree", which is typically used to perform a nearest neighbor search. [1]. For the sake of self-containment, we shall now provide a brief description of this data structure, following closely the exposition in [1].

Definition 2.7. Given a finite subset S of a metric space with metric d, a *cover tree* T on S is a leveled tree where each level is a "cover" for the level beneath it. Each level is indexed by an integer scale i which decreases as the tree is descended. Every node in the tree is associated with a point in S. Each point in S may be associated with multiple nodes in the tree; however, we require that any point appears at most once at every level. Let C_i denote the set of points in S associated with the nodes at level i. The cover tree obeys the following invariants for all i:

(i) (Nesting) $C_i \subset C_{i-1}$. This implies that once a point $p \in S$ appears in C_i then every lower level in the tree has a node associated with p;

(ii) (Covering tree) For every $p \in C_{i-1}$, there exists a $q \in C_i$ such that $d(p,q) < 2^i$ and the node in level i associated with q is a **parent** of the node in level $i-1$ associated with p, and respectively the node in level $i-1$ associated with p is a **child** of the node in level i associated with q.

(iii) (Separation) For all distinct $p, q \in C_i$, $d(p,q) > 2^i$.

Important Note: We will commit a slight abuse of terminology and identify nodes with their associated points, keeping in mind the the distinction made above. Since a point can appear in at most one node in the same level, no confusion can occur.

Instead of using scales 2^i, we can generalize the definition a little bit and use a^i, where $a > 1$ is some constant. However, our main result in this work uses the standard choice of scales, 2^i. We have a reason to believe that this result will also hold true for some $a \in (1, 2)$, which would be a slight improvement, but we only managed to find a proof for $a = 2$.

The cover tree consist of infinite number of layers, where C_∞ consists of one "root" node, and $C_{-\infty} = S$. Note, that we can put every point $p \in S$ into correspondence with exactly one node q for every level i. To see this, suppose that point p first time appeared in the cover tree at level j, C_j. If $j \geq i$, then this point also will be the node of C_i, so q is going to be the child of itself at level i. If $j < i$, then p has exactly one parent p_1 at level $j+1$, p_1 has exactly one parent p_2 on the level $j+2$, and so on. Thus, in the end we can put into correspondence point p and point $p_{i-j} = q \in C_j$. We are going to say, that q is a parent of p at level i, and also that p is an descendant of q at level i. Note, that for any point $q \in C_i$ and any of its descendant p, $d(p, q) < 2^{i+1}$. Indeed, since we can build a finite sequence of parent-child connections between p and q, we obtain $d(p, q) < \sum_{k=0}^{\infty} 2^{i-k} = 2^{i+1}$. We will appeal to this fact frequently later in this work.

The cover tree gives us a sequence of levels, C_i, which correspond to subsets of S, and each point in such a subset represents the set of its descendants. If we can manage to construct a simplicial complex T_i over the nodes of C_i for every i and connect these simplicial complexes with simplicial maps between T_i and T_{i+1}, then we will be able to compute the corresponding persistent (co)homology. Keeping in mind that the standard persistent (co)homology computations based on Čech or Vietoris-Rips complexes allow us to provide certain guarantees regarding recovered (co)homology classes, we may ask ourselves if it is possible to construct T_i and the connecting simplicial maps in such a way that the resulting persistent (co)homology is similar the one obtained using the standard way. The next chapter is devoted to providing an affirmative answer to this question.

CHAPTER 3
PERSISTENT COHOMOLOGY OF COVER REFINEMENTS VIA COVER TREE

The goal of this chapter is to develop a procedure for constructing a sequence of simplicial complexes, T_i, at each the level i of the cover tree, C_i, along with a sequence of connecting simplicial maps between them. In addition, we will show that the resulting persistence module is 1-interleaved with the corresponding persistence module obtained using the regular Čech complexes. Throughout this chapter, we assume that our data set S is a finite subset of a Euclidean space, and all the lemmas implicitly assume that all the points under consideration belong to this Euclidean space.

Recall that the cover allows us to select the subset of S at each level of the cover tree. We would like to such a subset at level i as the vertex set for our simplicial complex T_i. The general idea remains the same as earlier – connect points into a simplex if they are "close enough". To understand what "close enough" might mean in this case, let us imagine that the nodes of the i-th level are tree trunks in a forest. Each tree has its own family of leaves at the top, similarly to how each node has a family of descendants. Then we can say that a set of trees is close enough if we can find a set of leaves, one leaf from each tree, all of which lie inside some relatively small ball. Using this intuition, we can provide the formal definition of our simplicial complex T_i. Consider a subset of points $A \subset C_i$. We add simplex $\overline{\sigma}$ with the vertices from A to the simplicial complex T_i if and only if there exists a set of points $B \subseteq S$, such that a radius of a minimal ball containing B is not greater than 2^{i-1}, and for any point $a \in A$ there exist a point $b \in B$, such that b is a descendant of a at level i. Note that since the radius of the minimal ball containing B is not greater than 2^{i-1}, the set B spans a simplex $\sigma \in \check{C}ech(2^{i-1})$. Let $K_i = \check{C}ech(2^i)$. We say that $\sigma \in K_{i-1}$ **creates** simplex $\overline{\sigma} \in T_i$. Notice that such a simplex $\sigma \in K_{i-1}$ might not be unique.

We should mention that the choice of the value 2^{i-1} as a bound on the smallest enclosing ball of the creator simplex is not arbitrary. As me mentioned earlier, we would like to obtain an interleaving between the persistence modules constructed using the cover tree and the usual Čech complexes. Choosing the value of 2^{i-1} makes this possible. Let's start by reiterating that creator simplices for a simplex at scale 2^i belong to the V̌cech complex at scale 2^{i-1}. Moreover, we can notice that we can define a simplicial map $h_i : K_{i-1} \to T_i$: $h_i(a) = \overline{a}$, where a is any vertex in K_{i-1}, and $\overline{a}$ is a parent of a at level i. To see that this map is well defined, consider any simplex $\sigma \in K_{i-1}$. Let $\overline{\sigma}$

be the simplex on the parents of vertices of σ at level i, i.e., $h_i(\sigma) = \overline{\sigma}$. We want to ensure that $\overline{\sigma}$ is indeed in T_i. But it is clear from the construction that σ creates $\overline{\sigma}$.

Similarly, we can define a map $g_i : T_{i-1} \to T_i$ from a vertex to its parent at level i. Again, we need to make sure that this map is well defined. Suppose $g_i(\overline{\sigma}) = \widehat{\sigma}$. We want to show that $\widehat{\sigma}$ is indeed in T_{i+1}. Since $\overline{\sigma} \in T_i$, there exists $\sigma \in K_{i-1}$ such that σ creates $\overline{\sigma} \in T_i$. Notice that σ also creates $\widehat{\sigma} \in T_{i+1}$.

We can see now that we have sequence of simplicial complexes connected by simplicial maps, ... $\xrightarrow{g_{i-1}} T_{i-1} \xrightarrow{g_i} T_i \xrightarrow{g_{i+1}} T_{i+1} \xrightarrow{g_{i+2}}$ .... Hence, we can apply the usual procedure to compute persistent homology of the corresponding persistence module, ... $\xrightarrow{g_{i-1*}} H_p(T_{i-1}) \xrightarrow{g_{i*}} H_p(T_i) \xrightarrow{g_{i+1*}} H_p(T_{i+1}) \xrightarrow{g_{i+2*}}$ .... We also have the persistence module obtained using Čech complexes, ... $\xrightarrow{f_{i-1*}} H_p(K_{i-1}) \xrightarrow{f_{i*}} H_p(K_i) \xrightarrow{f_{i+1*}} H_p(K_{i+1}) \xrightarrow{f_{i+2*}}$ ..., with the maps f_{i*} induced by inclusion. Our main result shows that the two persistence modules are 1-interleaved.

Theorem 3.1. For every p, there exist maps $\alpha_i : H_p(K_{i-1}) \longrightarrow H_p(T_i)$ and $\beta_i : H_p(T_{i-1}) \longrightarrow H_p(K_i)$ such that for any $k \in \mathbb{Z}_0^+$ the following 2 diagrams commute:

$$\begin{tikzcd}
H_p(T_{i-1}) \arrow[rrr, "g_*"] \arrow[dr, "\beta_i"'] & & & H_p(T_{i+k+1}) \\
& H_p(K_i) \arrow[r, "f_*"] & H_p(K_{i+k}) \arrow[ur, "\alpha_{i+k+1}"'] &
\end{tikzcd}$$

$$\begin{tikzcd}
& H_p(T_i) \arrow[r, "g_*"] & H_p(T_{i+k}) \arrow[dr, "\beta_{i+k+1}"] & \\
H_p(K_{i-1}) \arrow[ur, "\alpha_i"] \arrow[rrr, "f_*"] & & & H_p(K_{i+k+1})
\end{tikzcd}$$

Let us briefly describe why this interleaving implies that the corresponding persistence diagrams are within bottleneck distance three. Fix the homological dimension p. Using the interleaving, we see that if some homology class lives across $k \geq 3$ consecutive groups $H_p(K_i)$, $H_p(K_{i+1})$,..., $H_p(K_{i+k-1})$, then there is a corresponding homology class that lives across at least $k-2$ consecutive groups $H_p(T_{i+1})$, $H_p(T_{i+2})$,..., $H_p(T_{i+k-2})$, and vice versa. In terms of persistence diagrams, it means that if D_T is the persistence diagram in dimension p computed for the persistence modules obtained from the cover tree and D_K is the persistence diagram in dimension p computed for the persistence module obtained using Čech complexes, then for any $x = (x_1, x_2) \in D_T$, if $|x_2 - x_1| \geq 3$ there is corresponding $y \in D_K$, such that $||x - y||_\infty \leq 2$, and vice versa. However, since we are looking for bijections between all such x's and y's, the bottleneck distance becomes bounded by three: $W_\infty(D_T, D_K) \leq 3$ (for details see [3]). We should note that since we use indices as coordinates of the points in the persistence diagram, we are technically dealing with the logarithmic scale (the actual scale is given by the corresponding powers of 2).

We discussed earlier, there is a well defined map $K_{i-1} \to T_i$. We know that this map induces a map between homology groups $H_p(K_{i-1}) \to H_p(T_i)$. This is going to be our map α_i.

Now, consider the following diagram:

$$\begin{tikzcd}
 & H_p(T_i) \arrow[dr] & \\
H_p(K_{i-1}) \arrow[ur] \arrow[r] & H_p(K_i) \arrow[r] & H_p(K_{i+1})
\end{tikzcd}$$

As we just mentioned, the bottom-top diagonal map is the map α_i. However, it is not immediately clear how to define the top-bottom diagonal map. We would like to cycles of T_i into cycles of K_{i+1}. Unfortunately, if $\overline{\sigma} \in T_i$, then simplex $\overline{\sigma}$ on the same vertex set in K_{i+1} might not exist. But we know that there exists $\sigma \in K_{i-1}$ which creates $\overline{\sigma} \in T_i$. Since $K_{i-1} \subseteq K_{i+1}$, we have $\sigma \in K_{i+1}$. It means that there are some small simplices in K_{i+1} which create simplices in T_i. We will use these small simplices to find the corresponding cycles in K_{i+1}. More precisely, we will construct an auxiliary simplicial complex $\overline{K_{i+1}} \subseteq K_{i+1}$, $K_{i-1} \subseteq \overline{K_{i+1}}$, such that $\overline{K_{i+1}} \simeq T_i$ are homotopy equivalent, $\overline{K_{i+1}} \simeq T_i$. A reader not familiar with the concept of homotopy equivalence can consult [15] for the definition. However, we do not use the definition. Rather, we rely on the fact that homotopy equivalence induces an isomorphism between homology groups (see e.g. [15] for details). As a result, this homotopy equivalence allows us to define the necessary top-bottom map.

We will now start the process of constructing $\overline{K_i}$. This simplicial complex will need to possess certain nice geometric properties, which we will describe using three main lemmas. All three have a similar setup. We begin with a simplex $\overline{\sigma} \in T_i$. In the first lemma, we show that if a simplex $\overline{\sigma}$ with the same vertices exists in K_{i+1}, then for each simplex σ that creates $\overline{\sigma}$ there exists a big simplex with vertices from both σ and $\overline{\sigma}$ in K_{i+1}. In the second lemma, we show that if a simplex $\overline{\sigma}$ with the same vertices does not exist in K_{i+1}, then all simplices σ_i that create $\overline{\sigma}$ form one big simplex in K_{i+1}. The third lemma is a sort of union of the first two. We show that if a simplex $\overline{\sigma}$ with the same vertices does not exist in K_{i+1} but there is a face $\widehat{\sigma}$ of $\overline{\sigma}$ which exist in K_{i+1}, then all simplices σ_i which create $\overline{\sigma}$ together with simplex $\widehat{\sigma}$ form one big simplex in K_{i+1}. We will also employ several other lemmas, which are auxiliary lemmas to prove the main ones.

Lemma 3.1 (***Main lemma 1***)**.** Consider a simplex $\overline{\sigma} \in T_i$ such that $\overline{\sigma} \in K_{i+1}$. Consider any simplex $\sigma \in K_{i-1}$ which creates $\overline{\sigma}$. Let A be the union of vertices of $\overline{\sigma}$ and σ. Then A spans a simplex $\alpha \in K_{i+1}$.

Proof. By definition, a simplex α exists in K_{i+1} if all of its vertices lie inside the ball of radius 2^{i+1}. So we want to prove that there exists point O, such that $A \subset B_{2^{i+1}}(O)$. Let $O_{\overline{\sigma}}$ be the center of the ball with the minimal radius $r_{\overline{\sigma}}$ containing simplex $\overline{\sigma}$. Because $\overline{\sigma} \in K_{i+1}$ then $r_{\overline{\sigma}} < 2^{i+1}$. Let O_σ be the center of the ball with the minimal radius r_σ containing simplex σ. Because $\sigma \in K_{i-1}$, we have $r_\sigma < 2^{i-1}$.

Let's place new point $\widehat{O}$ at the point $O_{\overline{\sigma}}$ and start moving $\widehat{O}$ towards the point O_σ until the distance between $\widehat{O}$ and some point $\overline{C} \in \overline{\sigma}$ is equal to 2^{i+1} (Figure 3.1). If we never hit the distance 2^{i+1}, then $d(O_\sigma, \overline{C}) < 2^{i+1}$ for any point $\overline{C} \in \overline{\sigma}$ and $d(O_\sigma, C) < 2^{i-1}$ for any point $C \in \sigma$, so $A \subset B_{2^{i+1}}(O_\sigma)$. So, assume we hit the distance $d(\widehat{O}, \overline{C}) = 2^{i+1}$. Note that in this case, $\angle\overline{C}\widehat{O}O_\sigma$ is obtuse. Because σ creates $\overline{\sigma}$, there is $C \in \sigma$ such that C is a descendent of $\overline{C}$. Hence the distance $d(C, \overline{C}) < 2^{i+1}$, so $d(\overline{C}, O_\sigma) \leq d(\overline{C}, C) + d(C, O_\sigma) < 2^{i+1} + 2^{i-1}$. Now we can use the cosine theorem for triangle $\overline{C}\widehat{O}O_\sigma$: $d^2(\overline{C}, O_\sigma) = d^2(\overline{C}, \widehat{O}) + d^2(\widehat{O}, O_\sigma) - 2d(\overline{C}, \widehat{O})d(\widehat{O}, O_\sigma)\cos(\angle\overline{C}\widehat{O}O_\sigma)$ $> d^2(\overline{C}, \widehat{O}) + d^2(\widehat{O}, O_\sigma) = d^2(\widehat{O}, O_\sigma) + 2^{2i+2}$, so $d^2(\widehat{O}, O_\sigma) < d^2(\overline{C}, O_\sigma) - 2^{2i+2} < (2^{i+1} + 2^{i-1})^2 - 2^{2i+2} = 2^{2i+1} + 2^{2i-2} = 9 \cdot 2^{2i-2}$, so $d(\widehat{O}, O_\sigma) < 3 \cdot 2^{i-1}$. Then for any vertex $D \in \sigma$ $d(\widehat{O}, D) \leq d(\widehat{O}, O_\sigma) + d(O_\sigma, D) < 3 \cdot 2^{i-1} + 2^{i-1} = 2^{i+1}$. It means that $\sigma \in B_{2^{i+1}}(\widehat{O})$. Also, by construction $\overline{\sigma} \in B_{2^{i+1}}(\widehat{O})$. The above facts imply the existence of a point O such that both $\sigma \in B_{2^{i+1}}(O)$ and $\overline{\sigma} \in B_{2^{i+1}}(O)$, so α exists in K_{i+1}.

□

Lemma 3.2. Consider a finite set of points S. Let O_S be the center of the ball B_S of minimal radius containing S. Let γ be any hyperplane passing through O_S. This hyperplane divides the boundary of the ball into two hemispheres. For any of these hemispheres, there exists a point $C \in S$ that lies on this hemisphere.

Proof. We will use the proof by contradiction. Suppose all points on the boundary of B_S lie on

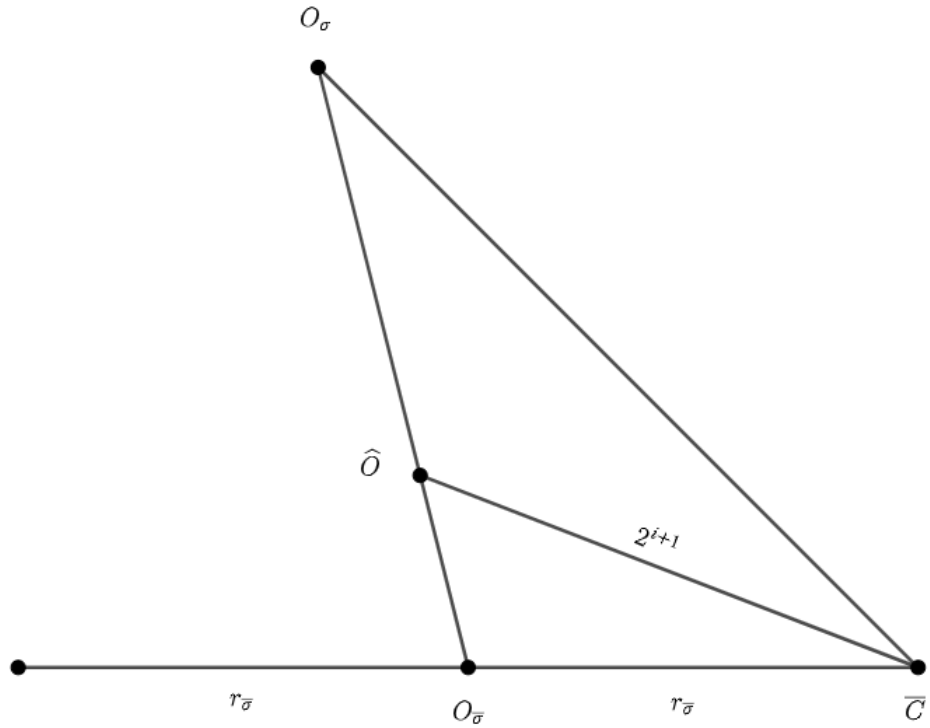

Figure 3.1: Supplementary figure for Lemma 3.1. The point $\widehat{O}$ can be chosen as O, i.e., $A \subset B_{2^{i+1}}(\widehat{O})$

one side of the hyperplane (Figure 3.2 (a)). Then we can move B_S perpendicularly to γ so none of the points from S will be on the boundary (Figure 3.2 (b)). But this implies that we can decrease the radius of the ball containing S, which contradicts the fact that B_S is the ball of minimal radius containing S.

$\square$

Lemma 3.3 (***Main lemma 2***)**.** Consider a simplex $\overline{\sigma} \in T_i$ such that $\overline{\sigma} \notin K_{i+1}$. Consider the family of all simplices $\sigma_j \in K_{i-1}$ which create $\overline{\sigma}$. Let A be the union of vertices of $\bigcup_j \sigma_j$. Then A spans a simplex $\alpha \in K_{i+1}$.

Proof. Similar to Lemma 3.1, we want to prove that there exists point O such that $A \subset B_{2^{i+1}}(O)$. Let $O_{\overline{\sigma}}$ be the center of the ball $B_{\overline{\sigma}}$ with the minimal radius $r_{\overline{\sigma}}$ containing simplex $\overline{\sigma}$. Since $\overline{\sigma} \notin K_{i+1}$ then $r_{\overline{\sigma}} > 2^{i+1}$. Consider any simplex $\sigma \in \bigcup_j \sigma_j$. Let O_σ be the center of the ball with the minimal radius r_σ containing simplex σ. Because $\sigma \in K_{i-1}$ we have $r_\sigma < 2^{i-1}$. Consider a hyperplane γ that passes through $O_{\overline{\sigma}}$ and is perpendicular to the line $O_{\overline{\sigma}}O_\sigma$. This hyperplane will divide the boundary of the ball $B_{\overline{\sigma}}$ into two hemispheres. By Lemma 3.2, there exists at

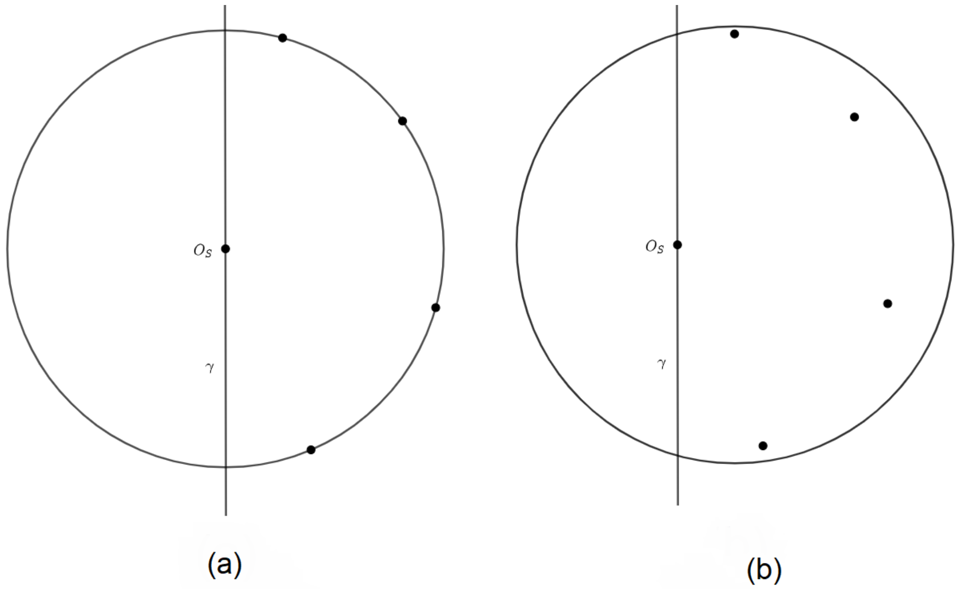

Figure 3.2: If all the points lie on the one side of the hyperplane **(a)**, then we can move the ball **(b)**, and decrease the radius

least one vertex $\overline{C}$ from $\overline{\sigma}$ that lies on the hemisphere and that lies on the opposite side from the point O_σ (Figure 3.3). Note that in this case, $\angle\overline{C}O_{\overline{\sigma}}O_\sigma \geq 90°$. Since $\overline{\sigma} \notin K_{i+1}$ then $d(O_{\overline{\sigma}}, \overline{C}) > 2^{i+1}$. Because σ creates $\overline{\sigma}$, there is $C \in \sigma$ such that C is a descendent of $\overline{C}$. Hence the distance $d(C, \overline{C}) < 2^{i+1}$, so $d(\overline{C}, O_\sigma) \leq d(\overline{C}, C) + d(C, O_\sigma) < 2^{i+1} + 2^{i-1}$. Now we can use the cosine theorem for the triangle $\overline{C}O_{\overline{\sigma}}O_\sigma$: $d^2(\overline{C}, O_\sigma) = d^2(\overline{C}, O_{\overline{\sigma}}) + d^2(O_{\overline{\sigma}}, O_\sigma) - 2d(\overline{C}, O_{\overline{\sigma}})d(O_{\overline{\sigma}}, O_\sigma)\cos(\angle\overline{C}O_{\overline{\sigma}}O_\sigma) > d^2(\overline{C}, O_{\overline{\sigma}}) + d^2(O_{\overline{\sigma}}, O_\sigma) > d^2(O_{\overline{\sigma}}, O_\sigma) + 2^{2i+2}$, so $d^2(O_{\overline{\sigma}}, O_\sigma) < d^2(\overline{C}, O_\sigma) - 2^{2i+2} < (2^{i+1} + 2^{i-1})^2 - 2^{2i+2} = 2^{2i+1} + 2^{2i-2} = 9 \cdot 2^{2i-2}$, so $d(O_{\overline{\sigma}}, O_\sigma) < 3 \cdot 2^{i-1}$. Then for any vertex $D \in \sigma$ $d(O_{\overline{\sigma}}, D) \leq d(O_{\overline{\sigma}}, O_\sigma) + d(O_\sigma, D) < 3 \cdot 2^{i-1} + 2^{i-1} = 2^{i+1}$. It means that $\sigma \in B_{2^{i+1}}(O_{\overline{\sigma}})$. The above is true for any $\sigma \in \bigcup_j \sigma_j$. This implies that $A \subset B_{2^{i+1}}(O_{\overline{\sigma}})$, hence α exists in K_{i+1}.

□

Lemma 3.4. Let O be a point and S be a set of points such that for any $A_1, A_2 \in S$ we have $d(O, A_1) = d(O, A_2)$. Let O_S be the center of the ball B_S with the minimal radius r_S containing S. Suppose $A, B, C \in S$ are such that $d(A, O_S) = d(B, O_S) = r_S > d(C, O_S)$. Consider any point $\widehat{O}$ on the segment OO_S. Then $d(O, A) > d(\widehat{O}, A) = d(\widehat{O}, B) > d(\widehat{O}, C)$.

Proof. First, let us prove that $\angle OO_SA = \angle OO_SB = 90°$ and $\angle OO_SC > 90°$. Consider a hyperplane γ that passes through O_S and is perpendicular to OO_S. By Lemma 3.2, there is at least one point on each hemisphere. Suppose A lies on the one side and B on the other (Figure 3.4 (a)).

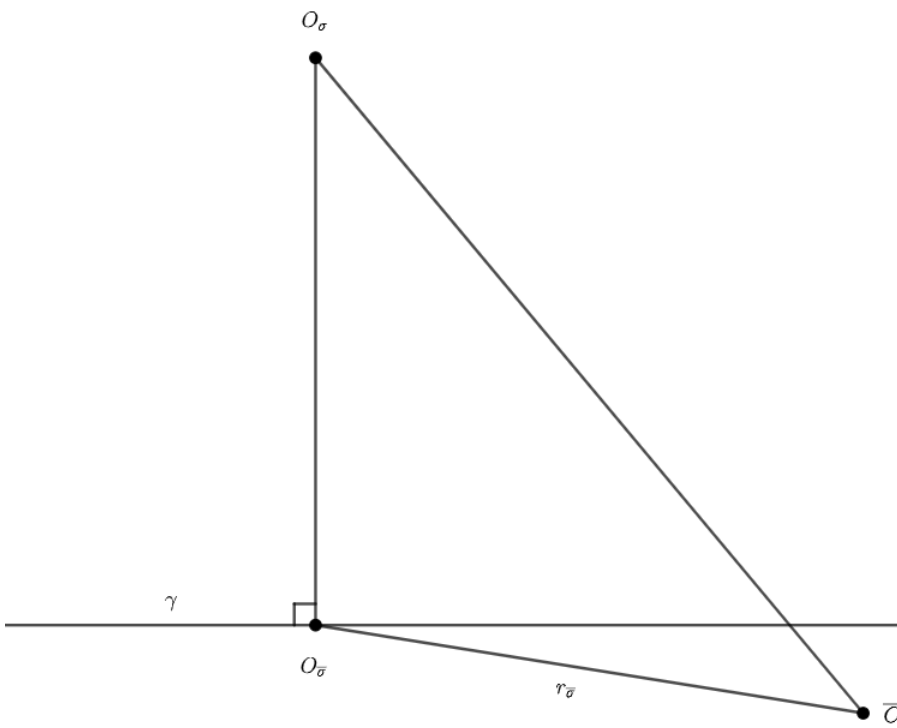

Figure 3.3: Supplementary figure for Lemma 3.3. The point $O_{\overline{\sigma}}$ can be chosen as O, i.e., $A \subset B_{2^{i+1}}(O_{\overline{\sigma}})$

Then $\angle OO_SA \leq 90°$ and $\angle OO_SB \geq 90°$. Consider triangles OO_SA and OO_SB: $OA = OB$, and $O_SA = O_SB$. So, $\Delta OO_SA = \Delta OO_SB$. But then $90° \geq \angle OO_SA = \angle OO_SB \geq 90°$, so $\angle OO_SA = \angle OO_SB = 90°$. Note that $OO_S^2 + O_SA^2 = OA^2$. Consider the triangle OO_SC. Note that $O_SC < r_S = O_SA$. Then $OO_S^2 + O_SA^2 = OA^2 = OC^2 = OO_S^2 + O_SC^2 - 2 \cdot OO_S \cdot O_SC \cdot \cos \angle OO_SC < OO_S^2 + O_SA^2 - 2 \cdot OO_S \cdot O_SC \cdot \cos \angle OO_SC$, so $\cos \angle OO_SC < 0$, and hence $\angle OO_SC > 90°$.

To finish the proof, let $d(\widehat{O}, O_S) = x$, then $\widehat{O}A = \widehat{O}B = x^2 + r_S^2$ and $\widehat{O}C = x^2 + O_SC^2 - 2 \cdot x \cdot O_S \cdot \cos \angle OO_SC$ (Figure 3.4 (b)). Let $f(x) = \widehat{O}A - \widehat{O}C = r_S^2 - O_SC^2 + 2 \cdot x \cdot O_S \cdot \cos \angle OO_SC$. This is a linear equation. When $\widehat{O} = O$, then $x = OO_S$ and $f(x) = 0$. When $\widehat{O} = O_S$, then $x = 0$ and $f(x) = O_SA - O_SC > 0$. Hence $f(x) > 0$ if $x \in (0, d(O, O_S))$, so $d(\widehat{O}, A) > d(\widehat{O}, C)$. Also, clearly $OA^2 = OO_S^2 + O_SA^2 > \widehat{O}O_S + O_SA^2 = \widehat{O}A$, so $d(O, A) > d(\widehat{O}, A) = d(\widehat{O}, B) > d(\widehat{O}, C)$.

□

Lemma 3.5 (***Main lemma 3***)**.** Consider a simplex $\overline{\sigma} \in T_i$ such that $\overline{\sigma} \notin K_{i+1}$ and such that there exists a face $\widehat{\sigma} \subset \overline{\sigma}$ with $\widehat{\sigma} \in K_{i+1}$. Consider the family of all simplices $\sigma_j \in K_{i-1}$ which creates $\overline{\sigma}$. Let A be the union of the vertices of $\bigcup_j \sigma_j$ together with the vertices of $\widehat{\sigma}$. Then A spans a simplex

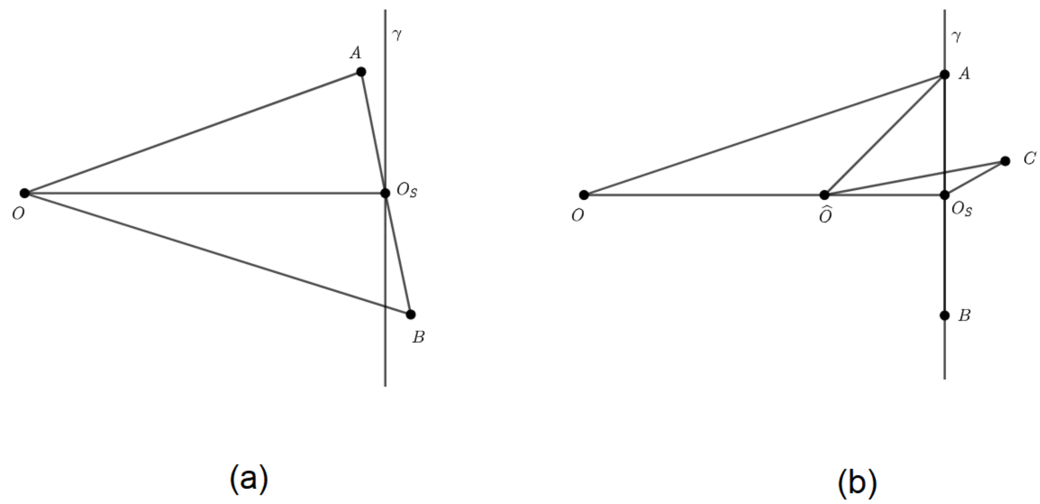

Figure 3.4: Supplementary figure for Lemma 3.4. The figure **(a)** is needed to prove that $\angle OO_SA = \angle OO_SA = 90°$. The figure **(b)** is needed for the rest of the proof

$\alpha \in K_{i+1}$.

Proof. Similarly to the previous main Lemmas, we want to prove that there exists a point O such that $A \subset B_{2^{i+1}}(O)$. Let $O_{\overline{\sigma}}$ be the center of the ball of minimal radius containing simplex $\overline{\sigma}$. By Lemma 3.3 we have $\bigcup_j \sigma_j \subset B_{2^{i+1}}(O_{\overline{\sigma}})$. Moreover, for any $\sigma \in \bigcup_j \sigma_j$, $d(O_{\overline{\sigma}}, O_\sigma) < 3 \cdot 2^{i-1}$, where O_σ is the center of the ball of minimal radius containing σ. Let S be the set of all points $\widehat{C} \in \widehat{\sigma}$ such that $d(\widehat{C}, O_{\overline{\sigma}}) = \max\limits_{D \in \widehat{\sigma}} d(D, O_{\overline{\sigma}})$ (it may consist of one point, but clearly, it is nonempty). Let O_S be the center of the ball B_S of minimal radius r_S containing S. Since S is the set of points that are vertices of $\widehat{\sigma} \in K_{i+1}$, we have $r_S < 2^{i+1}$. Let us take a point $\widehat{O} = O_{\overline{\sigma}}$ and start moving $\widehat{O}$ towards the point O_S until one of the following three things happens: (I) - for some $\sigma \in \bigcup_j \sigma_j$, $d(\widehat{O}, O_\sigma) = 3 \cdot 2^{i-1}$, where O_σ is the center of the ball with a minimal radius r_σ containing σ; (II) - there is at least one more vertex $E \in \widehat{\sigma}$, $E \notin S$ such that $d(E, \widehat{O}) = \max\limits_{D \in \widehat{\sigma}} d(D, \widehat{O})$; (III) - we reached the point O_S.

(I) Notice that $\angle O_S\widehat{O}O_\sigma \geq 90°$. Let γ be a hyperplane that passes through $\widehat{O}$ and is perpendicular to $\widehat{O}O_\sigma$. Let $\overline{\gamma}$ be a hyperplane that passes through O_S and is parallel to γ (Figure 3.5). Since $\angle O_S\widehat{O}O_\sigma \geq 90°$, we have O_S and O_σ lying on the opposite sides of γ. By Lemma 3.2, there is a point $\overline{C} \in S$ that lies on a hemisphere of B_S, and that also lies on the opposite side from O_σ relative to $\overline{\gamma}$. Clearly, $\overline{C}$ and O_σ lie on the opposite sides γ as well. Hence $\angle \overline{C}\widehat{O}O_\sigma \geq 90°$. Suppose $d(\overline{C}, \widehat{O}) \geq 2^{i+1}$. Let us use the cosine theorem for the triangle $\overline{C}\widehat{O}O_\sigma$:

$\overline{C}O_\sigma^2 = \overline{C}\widehat{O}^2 + \widehat{O}O_\sigma^2 - 2 \cdot \overline{C}\widehat{O} \cdot \widehat{O}O_\sigma \cdot \cos\angle\overline{C}\widehat{O}O_\sigma \geq (2^{i+1})^2 + (3 \cdot 2^{i-1})^2 = (2^{i+1} + 2^{i-1})^2$. So $\overline{C}O_\sigma \geq 2^{i+1} + 2^{i-1}$. Because σ creates $\overline{\sigma}$, there is $C \in \sigma$ such that C is descendent of $\overline{C}$. So the distance $d(C, \overline{C}) < 2^{i+1}$, so $d(\overline{C}, O_\sigma) \leq d(\overline{C}, C) + d(C, O_\sigma) < 2^{i+1} + 2^{i-1}$. We reached a contradiction. Hence $d(\overline{C}, \widehat{O}) < 2^{i+1}$. But $\overline{C}$ lies on the boundary of B_S. By Lemma 3.4 $d(\widehat{O}, \overline{C}) = \max_{D \in S} d(D, O_S) = \max_{D \in \widehat{\sigma}} d(D, O_S) < 2^{i+1}$. So $\widehat{\sigma} \subset B_{2^{i+1}}(\widehat{O})$. Also, we still have $\bigcup_j \sigma_j \subset B_{2^{i+1}}(\widehat{O})$. So we can pick $O = \widehat{O}$.

(II) In this scenario, we will update our set S by adding a point E to S, after which we update B_S and continue to move $\widehat{O}$ towards the new point O_S. By Lemma 3.4 we can see that as soon as any point $E \in \widehat{\sigma}$ becomes a point in S, it will never leave B_S. Because the number of vertices in $\widehat{\sigma}$ is finite, we will face scenario (II) finitely many times. So eventually, scenario (I) or (III) will happen.

(III) In this case, we can pick $O = O_S$. Indeed, notice that still for any $\sigma \in \bigcup_j \sigma_j$, $d(O_S, O_\sigma) < 3 \cdot 2^{i-1}$ where O_σ is the center of the ball of minimal radius containing σ. So, for any vertex $D \in \sigma$ $d(O_S, D) \leq d(O_S, O_\sigma) + d(O_\sigma, D) < 3 \cdot 2^{i-1} + 2^{i-1} = 2^{i+1}$, so $\sigma \in B_{2^{i+1}}(O_S)$. Let E be some vertex of $\widehat{\sigma}$ such that $d(E, \widehat{O}) = \max_{D \in \widehat{\sigma}} d(D, O_S)$. Since scenario (II) did not happen, then $E \in S$. So for any $D \in \widehat{\sigma}$, $d(D, O_S) \leq d(E, O_S) \leq r_S < 2^{i+1}$. Hence $A \subset B_{2^{i+1}}(O_S)$.

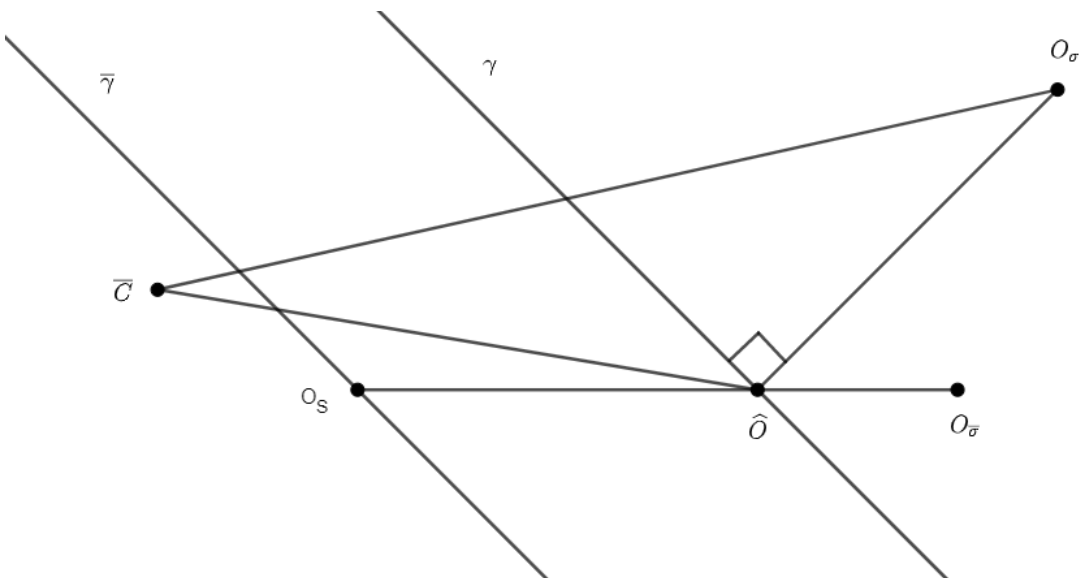

Figure 3.5: Supplementary figure for Lemma 3.5

□

Now we are ready to construct an auxiliary simplicial complex $\overline{K_i}$. Consider any simplex $\overline{\sigma} \in$

T_{i-1}. If $\overline{\sigma}$ exists in K_i, consider any simplex $\sigma \in K_{i-2}$ that creates $\overline{\sigma}$. By Lemma 3.1, there exists a simplex in K_i spanning the union of vertices of $\overline{\sigma}$ and σ. Let's this simplex $C_{\overline{\sigma},\sigma}$. Let $\overline{K_{i,\overline{\sigma}}} = \bigcup_{\sigma} C_{\overline{\sigma},\sigma}$. In other words, this is the union of big simplices where each big simplex has all parents and one set of descendants as vertices (Figure 3.6).

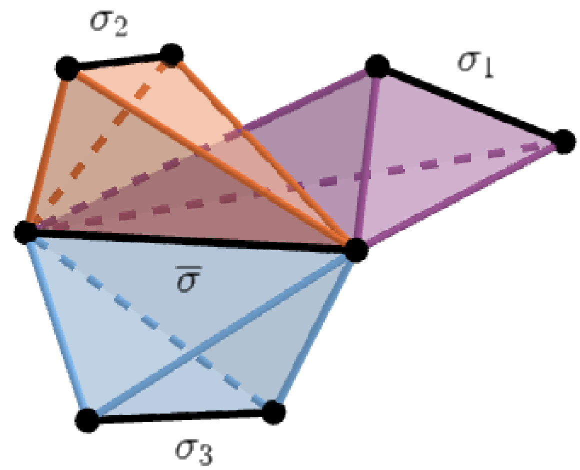

Figure 3.6: An example of $\overline{K_{i,\overline{\sigma}}}$, where we have 3 sets of descendants in K_{i-2} that create $\overline{\sigma}$

If, on the other hand, $\overline{\sigma}$ does not exist in K_i, consider all simplices $\sigma_j \in K_{i-2}$ that create $\overline{\sigma}$. By Lemma 3.3, there exists a simplex in K_i spanning the union of vertices of $\bigcup_j \sigma_j$. Moreover, by Lemma 3.5, for any $\widehat{\sigma} \subset \overline{\sigma}$, $\widehat{\sigma} \in K_i$, there exists a simplex in K_i spanning the union of vertices of $\bigcup_j \sigma_j \bigcup \widehat{\sigma}$. Let's call this simplex $C_{\overline{\sigma},\widehat{\sigma}}$. Let $\overline{K_{i,\overline{\sigma}}} = \bigcup_{\widehat{\sigma}} C_{\overline{\sigma},\widehat{\sigma}}$. In other words, this is the union of big simplices where each big simplex has all descendants and some subset of parents as vertices (Figure 3.7).

Define $\overline{K_i} = \bigcup_{\overline{\sigma} \in T_{i-1}} \overline{K_{i,\overline{\sigma}}}$. This complex $\overline{K_i}$ has a couple of important properties. First, note that $K_{i-2} \subset \overline{K_i}$. Indeed, recall that we have the simplicial map $h_{i-1} : K_{i-2} \longrightarrow T_{i-1}$, which maps vertices to their parents. Consider any simplex $\sigma \in K_{i-2}$. Then $h_{i-1}(\sigma) = \overline{\sigma}$ for some $\overline{\sigma} \in T_{i-1}$. It is easy to see that $\sigma \in \overline{K_{i,\overline{\sigma}}}$ by construction. Second, note that simplicial map $q_{i-1} : \overline{K_i} \longrightarrow T_{i-1}$, which maps vertices from $\overline{K_i}$ to the parents at level $i-1$, is well defined. In general, we do not have a simplicial map from K_i to T_{i-1} since some simplices do not exist in T_{i-1}, but in the case of $\overline{K_i}$, it is easy to see that any simplex from $\overline{K_{i,\overline{\sigma}}}$ is mapped to a face of $\overline{\sigma}$ which exists in T_{i-1} by construction. It turns out that this map q_{i-1} has an essential property.

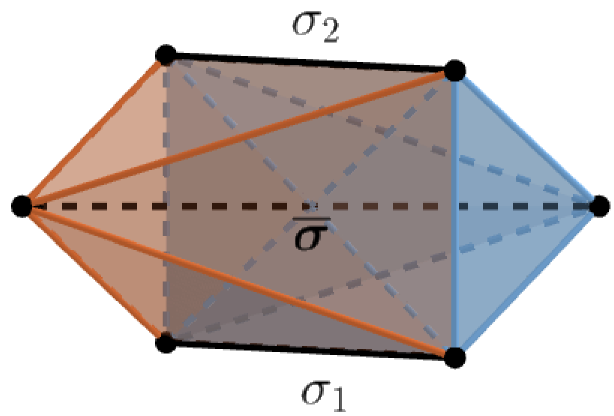

Figure 3.7: An example of $\overline{K_{i,\overline{\sigma}}}$, $\overline{\sigma} \notin T_{i-1}$ where we have two big simplices (blue and orange) built on one vertex from $\overline{\sigma}$ and all four descendants

Lemma 3.6. The simplicial map $q_{i-1} : \overline{K_i} \longrightarrow T_{i-1}$, which maps vertices from $\overline{K_i}$ to their parents on level $i-1$, is a homotopy equivalence.

Before proving this lemma, we need to introduce a new concept pertaining to simplicial complexes.

Definition 3.1. We call an edge $\{u, v\}$ of a simplicial complex $\triangle$ contractible if every simplex $\sigma \in \triangle$ satisfying $\{u\} \bigcup \sigma \in \triangle$ and $\{v\} \bigcup \sigma \in \triangle$ also satisfies $\{u, v\} \bigcup \sigma \in \triangle$.

If the edge $\{u, v\}$ is contractible, the contracted simplicial complex $\triangle/\{u, v\}$ is constructed as follows:

- We remove the vertices u and v from the vertex set S of $\triangle$ and add new vertex w
- A set $\tau \subseteq V \setminus \{u, v\}$ is a simplex of $\triangle/\{u, v\}$ if $w \notin \tau$ and $\tau \in \triangle$ or $w \in \tau$ and at least one of $\tau \setminus \{w\} \bigcup \{u\}$, $\tau \setminus \{w\} \bigcup \{v\}$ is a simplex of $\triangle$.

We are going to simplify notation and may identify the "new" vertex w with one of the old ones, u or v.

It is a well known fact that edge contraction is a homotopy equivalence (see [12] for details).

Theorem 3.2. Suppose an edge $\{u, v\}$ of a simplicial complex $\triangle$ is contractible. Then the simplicial map $f : \triangle \to \triangle/\{u, v\}$, which maps u and v into w and identity elsewhere, is a homotopy equivalence.

The idea behind the proof of Lemma 3.6 is the following. We will show that all edges in $\overline{K_i}$ between level $i-1$ parents and their descendants are contractible. Moreover, they will stay contractible after each such contraction. Ultimately, after all such contractions, we will get a simplicial complex on cover tree T_{i-1}. Since every contraction is a homotopy equivalence, the composition will also be a homotopy equivalence. We will divide the proof into smaller parts.

Lemma 3.7. Consider any vertex v which is not a parent itself in T_{i-1} along with its parent $\overline{v} \in T_{i-1}$. Edges $\{v, \overline{v}\}$ are contractible in $\overline{K_i}$. Moreover, they remain contractible after a series of contractions of any edges of the form $\{v, \overline{v}\}$ to $\overline{v}$.

Proof. Note that the second part of the lemma is not an obvious fact, and in general, it might not be true. For example in Figure 3.8 we can see that at first both edges $\{v_1, u_1\}$ and $\{v_2, u_2\}$ are contractible. However, after the first contraction $\{v_1, u_1\}$ to w, the second edge is not contractible anymore.

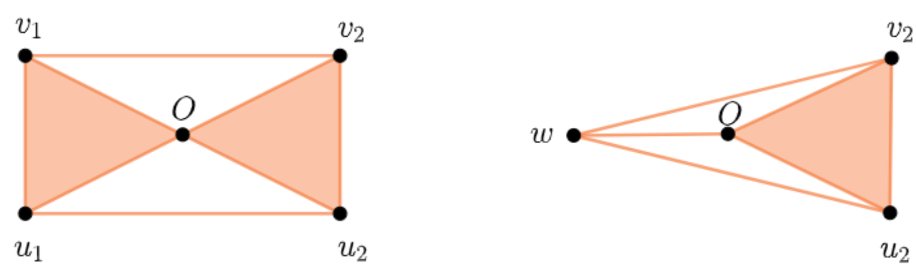

Figure 3.8: An example where edge becomes not contractible after other contraction

By definition, the edge $\{v, \overline{v}\}$ is contractible if for any simplex $\gamma \in \overline{K_i}^*$ satisfying $\{v\} \bigcup \gamma \in \overline{K_i}^*$ and $\{\overline{v}\} \bigcup \gamma \in \overline{K_i}^*$ also satisfies $\{v, \overline{v}\} \bigcup \gamma \in \overline{K_i}^*$, where $\overline{K_i}^*$ is the simplicial complex $\overline{K_i}$ after some edge contractions. Consider vertices of γ: $v_1^*, v_2^*, ..., v_k^*$. Since the simplex spanning $\{v_1^*, v_2^*, ..., v_k^*, v\}$ exists in $\overline{K_i}^*$, then simplex $\overline{\omega}$ spanning $\{\overline{v_1}, \overline{v_2}, ..., \overline{v_k}, \overline{v}\}$ exists in T_{i-1}, where $\overline{v_j}$ is the parent of v_j^*. Consider simplices $\{v_1^*, v_2^*, ..., v_k^*, v\}$ and $\{v_1^*, v_2^*, ..., v_k^*, \overline{v}\} \in \overline{K_i}^*$. Since they exist in $\overline{K_i}^*$, there are two simplices, $\{\widetilde{v_1}, \widetilde{v_2}, ..., \widetilde{v_k}, v\}$ and $\{\widehat{v_1}, \widehat{v_2}, ..., \widehat{v_k}, \overline{v}\} \in \overline{K_i}$, such that $\widehat{v_j}$ and $\widetilde{v_j}$ are the same as v_j^*

in the case if v_j^* is the original vertex of $\overline{K_i}$ which was not involved in contractions before, or they are some previous versions of the vertex v_j^* before contractions (might be the same vertices though). Consider two significant cases: (I) - simplex $\overline{\omega}$ does not exist in $\overline{K_i}$; (II) - simplex $\overline{\omega}$ exists in $\overline{K_i}$.

(I) Claim: both $\{\widetilde{v_1}, \widetilde{v_2}, ..., \widetilde{v_k}, v\}$ and $\{\widehat{v_1}, \widehat{v_2}, ..., \widehat{v_k}, \overline{v}\}$ belong to $\overline{K_{i,\overline{\omega}}}$. Consider any of these simplices, for example, $\{\widetilde{v_1}, \widetilde{v_2}, ..., \widetilde{v_k}, v\}$. Suppose it belongs to some other $\overline{K_{i,\widehat{\omega}}}$. Notice that it is possible only if $\overline{\omega} \subset \widehat{\omega}$ (all vertices $\overline{v_j}$ and $\overline{v}$ should be in $\widehat{\omega}$). Simplex $\{\widetilde{v_1}, \widetilde{v_2}, ..., \widetilde{v_k}, v\}$ consists of some parents $\{\widetilde{v_{j_1}}, \widetilde{v_{j_2}}, ..., \widetilde{v_{j_l}}\}$ and some descendants who are not parents $\{\widetilde{v_{j_{l+1}}}, \widetilde{v_{j_{l+2}}}, ..., \widetilde{v_{j_k}}, v\}$. The simplex on parents $\{\widetilde{v_{j_1}}, \widetilde{v_{j_2}}, ..., \widetilde{v_{j_l}}\}$ also exists in $\overline{K_{i,\overline{\omega}}}$. We only need to show that any of the descendants above also exist in $\overline{K_{i,\overline{\omega}}}$. Consider any of the above descendants, $u \in \overline{K_{i,\widetilde{\omega}}}$. It is part of the simplex that creates $\widetilde{\omega}$. A A face of this simplex will also create $\overline{\omega}$, and obviously, u will be part of this face since its parent is in $\overline{\omega}$. Hence $u \in \overline{K_{i,\overline{\omega}}}$. As we mentioned above, $\overline{K_{i,\overline{\omega}}}$ consists of big simplices where each big simplex is built on a face of $\overline{\omega}$ ($\{\widetilde{v_{j_1}}, \widetilde{v_{j_2}}, ..., \widetilde{v_{j_l}}\}$) and all of the descendants (we showed that each descendant of $\{\widetilde{v_{j_{l+1}}}, \widetilde{v_{j_{l+2}}}, ..., \widetilde{v_{j_k}}, v\}$ is in $\overline{K_{i,\overline{\omega}}}$). Hence, $\{\widetilde{v_1}, \widetilde{v_2}, ..., \widetilde{v_k}, v\} \in \overline{K_{i,\overline{\omega}}}$ (the same arguments hold for $\{\widehat{v_1}, \widehat{v_2}, ..., \widehat{v_k}, \overline{v}\}$).

Simplex $\{\widehat{v_1}, \widehat{v_2}, ..., \widehat{v_k}, \overline{v}\}$ is a part of a big simplex in $\overline{K_{i,\overline{\omega}}}$, and each big simplex contains all the descendants. v is one of the descendants of $\overline{K_{i,\overline{\omega}}}$. Therefore, $\{\widehat{v_1}, \widehat{v_2}, ..., \widehat{v_k}, \overline{v}, v\}$ is a simplex in $\overline{K_{i,\overline{\omega}}}$. Hence, $\{v_1^*, v_2^*, ..., v_k^*, \overline{v}, v\} = \{v, \overline{v}\} \bigcup \gamma \in \overline{K_i}^*$

(II) Unlike the above scenario, a similar claim might not be valid. Consider a simplex $\{\widetilde{v_1}, \widetilde{v_2}, ..., \widetilde{v_k}, v\}$. It belongs to some $\overline{K_{i,\widehat{\omega}}}$. Here we have two smaller cases: (a) - simplex $\widehat{\omega}$ does not exist in $\overline{K_i}$, (b) - simplex $\widehat{\omega}$ exists in $\overline{K_i}$.

(a): Simplex $\{\widetilde{v_1}, \widetilde{v_2}, ..., \widetilde{v_k}, v\}$ consists of some parents $\{\widetilde{v_{j_1}}, \widetilde{v_{j_2}}, ..., \widetilde{v_{j_l}}\}$ and some descendants which are not parents $\{\widetilde{v_{j_{l+1}}}, \widetilde{v_{j_{l+2}}}, ..., \widetilde{v_{j_k}}, v\}$. If we add one more parent $\overline{v}$, we will get simplex $\{\widetilde{v_{j_1}}, \widetilde{v_{j_2}}, ..., \widetilde{v_{j_l}}, \overline{v}\}$ which also belongs to $\overline{K_{i,\widehat{\omega}}}$ since $\{\widetilde{v_{j_1}}, \widetilde{v_{j_2}}, ..., \widetilde{v_{j_l}}, \overline{v}\}$ is a face of $\overline{\omega}$, which exists in $\overline{K_i}$. Since each big simplex in $\overline{K_{i,\widehat{\omega}}}$ contains all the descendants, there exists a simplex on $\{\widetilde{v_{j_1}}, \widetilde{v_{j_2}}, ..., \widetilde{v_{j_l}}, \overline{v}\} \bigcup \{\widetilde{v_{j_{l+1}}}, \widetilde{v_{j_{l+2}}}, ..., \widetilde{v_{j_k}}, v\} = \{\widetilde{v_1}, \widetilde{v_2}, ..., \widetilde{v_k}, v, \overline{v}\}$. Therefore $\{v_1^*, v_2^*, ..., v_k^*, \overline{v}, v\} = \{v, \overline{v}\} \bigcup \gamma \in \overline{K_i}^*$

(b): Simplex $\{\widetilde{v_1}, \widetilde{v_2}, ..., \widetilde{v_k}, v\}$ is a part of a big simplex in $\overline{K_{i,\widehat{\omega}}}$. Each simplex in $\overline{K_{i,\widehat{\omega}}}$ consists of some descendants and all the parents. Hence, we can add the vertex $\overline{v}$, which is a parent to the above simplex. Thus, $\{\widetilde{v_1}, \widetilde{v_2}, ..., \widetilde{v_k}, v, \overline{v}\} \in \overline{K_{i,\widehat{\omega}}}$. Therefore, $\{v_1^*, v_2^*, ..., v_k^*, \overline{v}, v\} = \{v, \overline{v}\} \bigcup \gamma \in \overline{K_i}^*$. □

Now we can prove Lemma 3.6.

Proof. Using Lemma 3.7, we know we can contract all the edges connecting descendants and parents. Also, Theorem 3.2 gives us that the simplicial map that maps vertices into the contracted vertices is a homotopy equivalence. The last step is to notice that the final complex we get after all contractions is the complex T_{i-1} itself. Indeed, the only vertices we get in the end are parents on level $i-1$. And clearly, each $\overline{K_{i,\sigma}}$ contracts into σ itself. Therefore, map $q_{i-1} : \overline{K_i} \longrightarrow T_{i-1}$ is a homotopy equivalence. $\square$

Now we have all the necessary ingredients to prove Theorem 3.1. We will consider each diagram from the theorem in separate lemmas. We will only consider the case when $k = 0$. The proof remains is essentially the same for the case when $k > 0$.

Lemma 3.8. There exist maps $\alpha_{i-1} : H_p(K_{i-1}) \longrightarrow H_p(T_i)$ and $\beta_{i-1} : H_p(T_{i-1}) \longrightarrow H_p(K_i)$ such that the following diagram commutes:

$$\begin{tikzcd}
H_p(T_{i-1}) \arrow[r, "g_i^*"] \arrow[dr, "\beta_i"'] & H_p(T_i) \arrow[r, "g_{i+1}^*"] & H_p(T_{i+1}) \\
& H_p(K_i) \arrow[ur, "\alpha_{i+1}"'] &
\end{tikzcd}$$

Proof. Recall that we already defined the map α_{i+1} as the map induced by the simplicial map h_{i+1} described earlier. By Lemma 3.6, the map $q_{i-1} : \overline{K_i} \longrightarrow T_{i-1}$ is a homotopy equivalence. We also know that $\overline{K_i} \subset K_i$, so we can define the inclusion map $p_i : \overline{K_i} \longrightarrow K_i$. Now, consider the following diagram:

$$\begin{tikzcd}
T_{i-1} \arrow[r, "g_i"] & T_i \arrow[r, "g_{i+1}"] & T_{i+1} \\
& \overline{K_i} \arrow[ul, "q_{i-1}"] \arrow[d, "p_i"'] & \\
& K_i \arrow[uur, "h_{i+1}"'] &
\end{tikzcd}$$

It is easy to see that the diagram above commutes. Indeed, both $q_{i-1} \circ g_i \circ g_{i+1}$ and $p_i \circ h_{i+1}$ map vertices from $\overline{K_i}$ to their parents on level $i+1$. Hence, the following induced diagram between homology groups commutes as well:

$$\begin{tikzcd}
H_p(T_{i-1}) \arrow[r, "g_i^*"] & H_p(T_i) \arrow[r, "g_{i+1}^*"] & H_p(T_{i+1}) \\
 & H_k(\overline{K_i}) \arrow[ul, "q_{i-1}^*"'] \arrow[d, "p_i^*"'] & \\
 & H_k(K_i) \arrow[uur, "\alpha_{i+1}^*"'] &
\end{tikzcd}$$

But since q_{i-1} is a homotopy equivalence, then q_{i-1}^* is an isomorphism. So the following diagram commutes as well:

$$\begin{tikzcd}
H_p(T_{i-1}) \arrow[r, "g_i^*"] \arrow[dr, "(q_{i-1}^*)^{-1}"] & H_p(T_i) \arrow[r, "g_{i+1}^*"] & H_p(T_{i+1}) \\
 & H_p(\overline{K_i}) \arrow[d, "p_i^*"'] & \\
 & H_p(K_i) \arrow[uur, "\alpha_{i+1}^*"'] &
\end{tikzcd}$$

Finally, we can define $\beta_i = (q_{i-1}^*)^{-1} \circ p_i^*$ □

Lemma 3.9. The same maps $\alpha_{i-1} : H_p(K_{i-1}) \longrightarrow H_p(T_i)$ and $\beta_{i-1} : H_p(T_{i-1}) \longrightarrow H_p(K_i)$ as in Lemma 3.8 make the following diagram commute:

$$
\begin{tikzcd}
& H_p(T_i) \arrow[dr, "\beta_{i+1}"] & \\
H_p(K_{i-1}) \arrow[ur, "\alpha_i"] \arrow[r, "f_i^*"] & H_p(K_i) \arrow[r, "f_{i+1}^*"] & H_p(K_{i+1})
\end{tikzcd}
$$

Proof. Similar to the previous lemma, consider the following diagram:

$$
\begin{tikzcd}
& T_i & \\
& & \overline{K_{i+1}} \arrow[ul, "q_i"'] \arrow[d, "p_{i+1}"] \\
K_{i-1} \arrow[uur, "h_i"] \arrow[urr, "r_{i+1}"] \arrow[r, "f_i"] & K_i \arrow[r, "f_{i+1}"] & K_{i+1}
\end{tikzcd}
$$

where the map $r_{i+1} : K_{i-1} \longrightarrow \overline{K_{i+1}}$ is the inclusion map, since we know that $K_{i-1} \subset \overline{K_{i+1}}$. The diagram above commutes. Indeed, all of the $r_{i+1}, p_{i+1}, f_i, f_{i+1}$ are just inclusion maps. And both $r_{i+1} \circ q_i$ and h_i map vertices from K_{i-1} into their parents on level i. So the induced map between homology groups commutes as well:

$$
\begin{tikzcd}
& H_p(T_i) & \\
& & H_p(\overline{K_{i+1}}) \arrow[ul, "q_i^*"'] \arrow[d, "p_{i+1}^*"] \\
H_p(K_{i-1}) \arrow[uur, "\alpha_i"] \arrow[urr, "r_{i+1}^*"] \arrow[r, "f_i^*"] & H_p(K_i \arrow[r, "f_{i+1}^*"] & H_p(K_{i+1})
\end{tikzcd}
$$

But again, q_i^* is an isomorphism, so the following diagram commutes as well:

$$
\begin{tikzcd}
 & H_p(T_i) \arrow[dr, "(q_i^*)^{-1}"] & \\
 & & H_p(\overline{K_{i+1}}) \arrow[d, "p_{i+1}^*"] \\
H_p(K_{i-1}) \arrow[uur, "\alpha_i"] \arrow[urr, "r_{i+1}^*"] \arrow[r, "f_i^*"] & H_p(K_i) \arrow[r, "f_{i+1}^*"] & H_p(K_{i+1})
\end{tikzcd}
$$

Finally, recall that $\beta_{i+1} = (q_i^*)^{-1} \circ p_{i+1}^*$, which finishes the proof of the lemma and the Theorem 3.1.

$\square$

CHAPTER 4
THEOREM IN CASE OF NPC SPACE

Theorem 3.1 from the previous chapter is proved in the case when our data set is a finite subset of a Euclidean space. It is interesting to look deeper and see if this theorem holds when the points lie in other kind of spaces.

Feeling optimistically, we can try to consider the case when the ambient space is an arbitrary metric space. But, as we will now show, the theorem does not hold in such a case. In particular, we will build an example for which the statement of Lemma 3.9 does not hold. Consider a circle with the circumference of 18 and the intrinsic metric d: for any two points x, y, $d(x, y)$ is the length of the shortest arc between them. Consider a point set as shown in Figure 4.1.

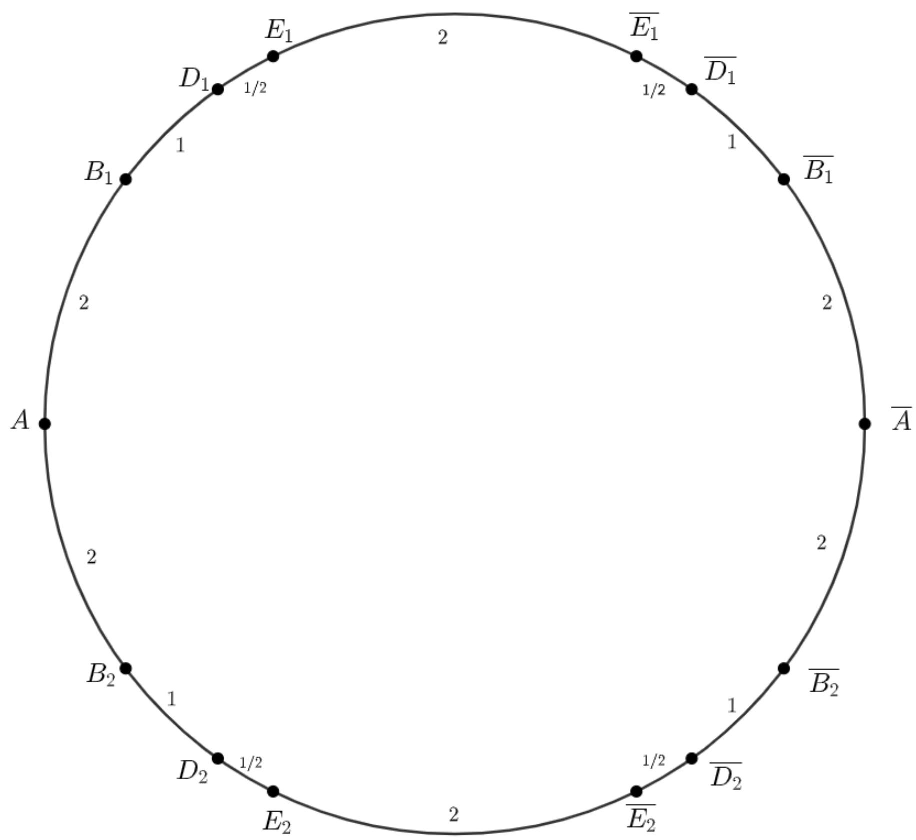

Figure 4.1: Setup of points on a circle for which the statement of the Main theorem does not hold

Let us construct a cover tree for this set. C_j for $j \geq 4$ will consist of only one point A. C_3 will have two points now: A and $\overline{A}$. $\overline{A}$ will be a child of A. Indeed, we can see that $d(A, \overline{A}) = 9$, and $2^3 < 9 < 2^4$. The sets C_2 and C_1 will be the same as the set C_3, and they will have only two points. Set C_0 will have four new points: B_1 and B_2 as children of A and $\overline{B_1}$ and $\overline{B_2}$ as children of $\overline{A}$. Set C_{-1} will have four more new points: D_1 as a child of B_1, D_2 as a child of B_2, $\overline{D_1}$ as a child of $\overline{B_1}$, and $\overline{D_2}$ as a child of $\overline{B_2}$. And finally, the set C_{-2} will be the same as the original point set: E_1 is a child of D_1, E_2 is a child of D_2, $\overline{E_1}$ is a child of $\overline{D_1}$, and $\overline{E_2}$ is a child of $\overline{D_2}$. Let us look at the simplicial complex T_1 on a set C_1. As we showed, it consists of two points. In order to understand if these points are connected with an edge, we need to look at the simplicial complex K_0 (Figure 4.2). We can observe that E_1 and $\overline{E_1}$ are connected with an edge, but E_1 is a descendant of A and $\overline{E_1}$ is a descendant of $\overline{A}$. This implies that A and $\overline{A}$ are connected in T_1.

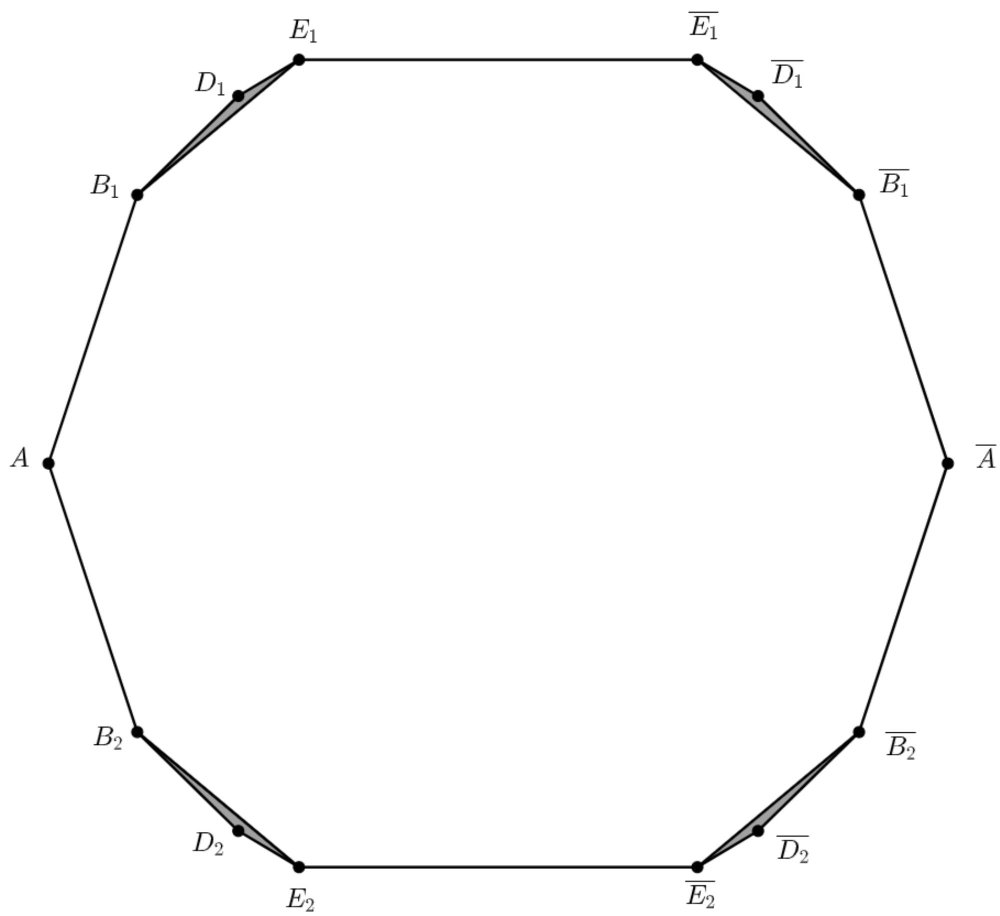

Figure 4.2: Simplicial complex K_0. The loop is present

Also, we can see a non-trivial homology class in dimension one present in $H_1(K_0)$, so the question is, does this homology class live in K_2? In order to check it, we have computed persistent homology using a C++ library called mayysus 2 [17]. The computation revealed that the homology class of interest was born at radius 1 (as we observe in Figure 4.2) and died at radius 4.5. When creating K_2, we look at the intersection of balls of radius $2^2 = 4 < 4.5$. Hence, this homology class exists in $H_1(K_2)$. We can also describe an intuitive idea behind the existence of this homology class in $H_1(K_2)$. If we consider three points A, B, and C on the circle, they will divide the circle into three arcs as shown in Figure 4.3. The radius of the ball of minimum radius containing these three points will be $\min(\frac{a+b}{2}, \frac{b+c}{2}, \frac{a+c}{2})$. In order to cover representatives of our nontrivial homology class with triangles, we need to cover the interior of the corresponding cycles. So, there will be a triangle that contains the center of the circle. However, for such a triangle, all three $a+b$, $a+c$, and $c+b$ are greater or equal to half of the circumference. Hence, $\min(\frac{a+b}{2}, \frac{b+c}{2}, \frac{a+c}{2}) \geq 4.5$. This implies that our homology class lives in $H_1(K_0)$ and $H_1(K_2)$. However, there is no non-trivial homology classes in $H_1(T_1)$ since it is just an edge. This is a contradiction to Lemma 3.9.

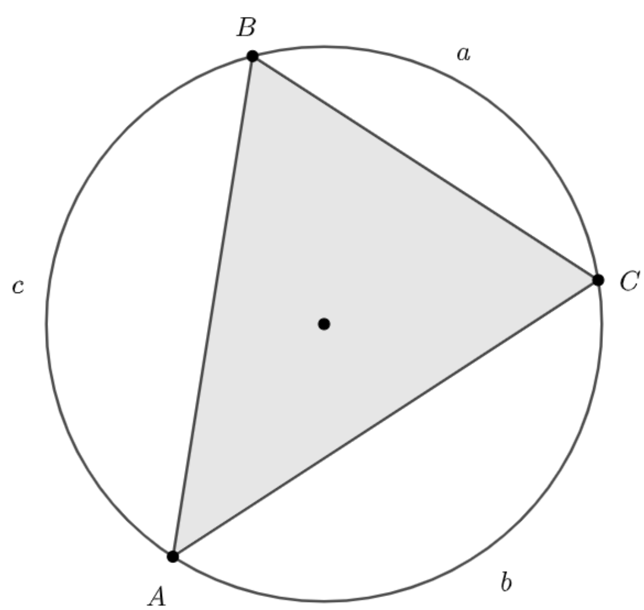

Figure 4.3: Supplementary figure which is helping to see why the loop is still present in K_2

It should be noted that the circle with the intrinsic metric is a space of positive curvature. So, the counterexample above tells us that our theorem does not work in general spaces of positive curvature. It is then reasonable to ask if it holds true in spaces of nonpositive curvature. It turns out that the situation is much better in the case of spaces of nonpositive curvature. We will show that the Theorem 3.1 holds in the case of a $CAT(0)$ space, also known as Hadamard space.

Let us start with definitions. We will closely follow [2] and suggest the reader consult this text for more details. Let (X, d) be a metric space. The **geodesic map** is an isometric map ρ between a convex subset $I \subseteq \mathbb{R}$ and X, $\rho : I \to X$. The map ρ is called a **geodesic segment, ray**, or **line** if I is a closed interval, half-line, or line respectively. A **geodesic metric space** is a metric space in which every two points are connected with a geodesic segment. A comparison triangle for a triple $(a, b, c) \in X^3$ is a triple $(\overline{a}, \overline{b}, \overline{c})$ of points in a Euclidean plane $\mathbb{R}^2$, such that $d(a, b) = \overline{d}(\overline{a}, \overline{b})$, $d(a, c) = \overline{d}(\overline{a}, \overline{c})$ and $d(b, c) = \overline{d}(\overline{b}, \overline{c})$, where $\overline{d}$ is a usual metric in Euclidean space. A geodesic triple (a, b, c) is said to satisfy **CAT(0) inequality**, if for any point p on any geodesic segment (a, c), and for any point q on any geodesic segment (b, c) the following inequality holds: $d(p, q) \leq \overline{d}(\overline{p}, \overline{q})$, where $\overline{p}$ is a point on a segment $(\overline{a}, \overline{c})$ and $\overline{q}$ is a point on a segment $(\overline{b}, \overline{c})$ in a comparison triangle for (a, b, c), such that $d(a, p) = \overline{d}(\overline{a}, \overline{p})$ and $d(b, q) = \overline{d}(\overline{b}, \overline{q})$. If X is a geodesic space such that all of its triples satisfy $CAT(0)$ inequality, then X is called an **Hadamard space**.

Let us introduce some useful nomenclature that we will employ a little later in this work. If A, B, C are points in Hadamard space (X, d), then define angle $\angle ABC$ as the angle $\angle\overline{ABC}$ of a comparison triangle for a triple (A, B, C). Also, if $\angle ABC < 90°$, then we are going to say, that "*by the cosine theorem*" $d(A, C)^2 < d(A, B)^2 + d(B, C)^2$. Although there is no such thing as the "cosine theorem" in Hadamard space, we can use the cosine theorem for a comparison triangle for a triple (A, B, C), which, together with the fact that the length of sides of the comparison triangle is the same as the length of the sides of the original one, will give us the inequality above. Similarly, we will use the "cosine theorem" if $\angle ABC \geq 90°$.

In order to prove Theorem 3.1 in the case of an Hadamard space, it suffices to prove the main geometric lemmas that we used in the proof for a Euclidean space, which are Lemma 3.1, Lemma 3.3 and Lemma 3.5. After proving these lemmas, an argument identical to the one in the case of a Euclidean space shows that Theorem 3.1 also holds in the case of an Hadamard space.

Let us start by recalling which geometric properties of Euclidean space were essential for us to prove these lemmas. Multiple arguments in our proofs involved moving one point towards another along the straight line. Therefore, we need to show that an analogous argument can be made in the case of an Hadamard space. Fortunately, we have the following theorem [2]:

Theorem 4.1. Every two points in an Hadamard space are connected by a unique geodesic

Also, we employed heavily the notion of the ball of minimal radius containing some finite set. As it turns out, such a ball is well defined in an Hadamard space (see [2] for details).

Theorem 4.2. In an Hadamard space, every bounded set has a unique closed ball of minimal radius containing this set.

We can now consider what changes need to be made to the original proofs to adapt them to the case of an Hadamard space. Suppose we are moving one point (let us say A) toward another (B) until the distance to some third point (C) increases to some number a. In a Euclidean space, we know some facts: if we continue to move this point, then the distance to C will keep increasing, and $\angle CAB \geq 90°$. We need to prove analogous results in an Hadamard space.

Lemma 4.1. Consider a geodesic triangle $\triangle(A, B, C)$ in an Hadamard space (X, d). Let D be any point on the geodesic segment (B, C). Then $d(A, D) < \max(d(A, C), d(A, B))$.

Proof. Consider a comparison triangle $\overline{\triangle}(\overline{A}, \overline{B}, \overline{C})$ for the triangle $\triangle(A, B, C)$. Let $\overline{D}$ be a point on the segment $\overline{BC}$ such that $\overline{d}(\overline{D}, \overline{B}) = d(D, B)$ (Figure 4.4). First, notice that the fact we are trying to prove is obvious in the case of a Euclidean space. Indeed, $\angle\overline{ADC} + \angle\overline{ADB} = 180°$. Then at least one of these angles is greater or equal than $90°$. Without loss of generality, assume that $\angle\overline{ADC} > 90°$. Then in the triangle $\overline{\triangle}(\overline{A}, \overline{D}, \overline{C})$ the side $\overline{AC}$ is the largest, so $\overline{d}(\overline{A}, \overline{C}) > \overline{d}(\overline{A}, \overline{D})$. Hence $\overline{d}(\overline{A}, \overline{D}) < \max(\overline{d}(\overline{A}, \overline{C}), \overline{d}(\overline{A}, \overline{B}))$. But since every geodesic triangle satisfies $CAT(0)$ inequality, we have $d(A, D) \leq \overline{d}(\overline{A}, \overline{D})$. So, $d(A, D) \leq \overline{d}(\overline{A}, \overline{D}) < \max(\overline{d}(\overline{A}, \overline{C}), \overline{d}(\overline{A}, \overline{B})) = \max(d(A, C), d(A, B))$. □

Lemma 4.2. Consider a geodesic line γ in an Hadamard space (X, d) and any point $A \notin \gamma$. For any $l \in \mathbb{R}$, there are at most two points B, $C \in \gamma$, such that $d(A, B) = d(A, C) = l$

Proof. Suppose there are three points A, B, $C \in \gamma$, such that $d(A) = d(B) = d(C) = \gamma$. Without loss of generality, suppose D lies between A and C. By Lemma 4.1 $d(A, D) < \max(d(A, B), d(A, C)) = l$. Contradiction. □

Lemma 4.3. Consider a geodesic triangle $\triangle(A, B, C)$ in an Hadamard space (X, d), and let $\overline{\triangle}(\overline{A}, \overline{B}, \overline{C})$ be a comparison triangle for the triangle $\triangle(A, B, C)$. Suppose $\angle\overline{ABC} < 90°$. Then

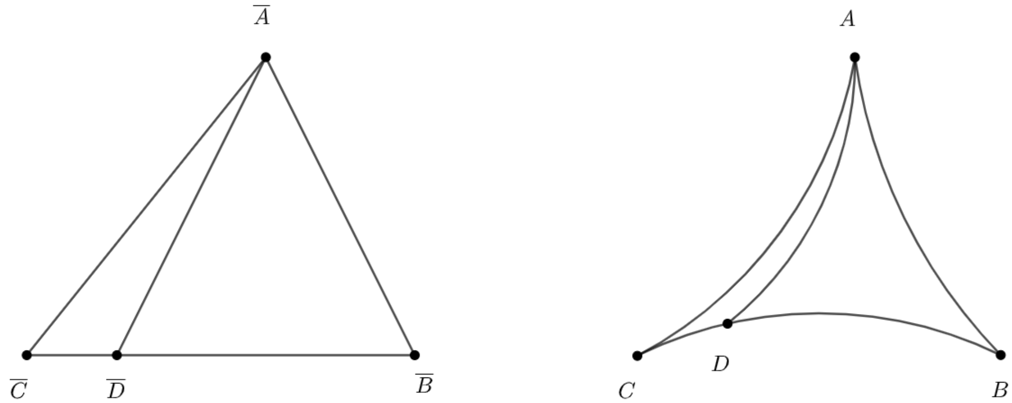

Figure 4.4: The length of a geodesic AD in an Hadamard space is less than the length of segment $\overline{AD}$ in comparison triangle

there exists $\epsilon_1 > 0$ such that for any $0 < \epsilon_2 \le \epsilon_1$ and a point D on the geodesic segment BC satisfying $d(B, D) < \epsilon_2$, the distance $d(A, B) > d(A, D)$. In other words, if we are moving a point D starting from B towards C, then the distance to A, $d(A, D)$, is decreasing.

Proof. Let $\overline{F}$ be a point on the ray $\overline{BC}$ such that $\angle\overline{AFB} = 90°$. Then for any point $\overline{D} \in \overline{FB}$, $\overline{d}(\overline{D}, A) < \overline{d}(\overline{B}, A)$. We can always take such a point $\overline{D}$ to lie on the segment $\overline{BC}$ (even though the point $\overline{F}$ may lie farther from $\overline{B}$ than the point $\overline{C}$). Let D be a point on a geodesic segment BC such that $d(B, D) = \overline{d}(\overline{B}, \overline{D})$. Since every geodesic triangle satisfies $CAT(0)$ inequality, then $d(A, D) \le \overline{d}(\overline{A}, \overline{D}) < \overline{d}(\overline{B}, A) = d(A, B)$. Hence, ϵ_1 exists and $epsilon_1 = \min(\overline{d}(\overline{B}, \overline{F}), \overline{d}(\overline{B}, \overline{C}))$. $\square$

The following corollary is immediate.

Corollary 4.1. Consider a geodesic triangle $\triangle(A, B, C)$ in an Hadamard space (X, d), and let $\overline{\triangle}(\overline{A}, \overline{B}, \overline{C})$ be a comparison triangle for the triangle $\triangle(A, B, C)$. Suppose there exists $\epsilon_1 > 0$ such that for any $0 < \epsilon_2 < \epsilon_1$ and a point D on a geodesic segment BC satisfying $d(B, D) < \epsilon_2$, the distance $d(A, B) < d(A, D)$. In other words, if we are moving point D starting from B towards C, then the distance to A, $d(A, D)$, is increasing. Then $\angle\overline{ABC} \ge 90°$.

Now we are ready to prove Lemma 3.1 in the case of an Hadamard space. The main idea and approach of the proof remain the same as in the the original proof.

Lemma 4.4 (***Main lemma 1***)**.** Consider a simplex $\overline{\sigma} \in T_i$ such that $\overline{\sigma} \in K_{i+1}$. Take any simplex

$\sigma \in K_{i-1}$ that "creates" simplex $\overline{\sigma}$. Let A be the union of vertices of $\overline{\sigma}$ and σ. Then A spans a simplex $\alpha \in K_{i+1}$.

Proof. The proof is very similar to the original one. We want to prove that there exists a point O such that $A \subset B_{2^{i+1}}(O)$. Let $O_{\overline{\sigma}}$ be the center of the ball with the minimal radius $r_{\overline{\sigma}}$ containing simplex $\overline{\sigma}$. Because $\overline{\sigma} \in K_{i+1}$ then $r_{\overline{\sigma}} < 2^{i+1}$. Let O_σ be the center of the ball with the minimal radius r_σ containing a simplex σ. Because $\sigma \in K_{i-1}$ then $r_\sigma < 2^{i-1}$. Let us place a point $\widehat{O}$ at the point $O_{\overline{\sigma}}$ and start moving $\widehat{O}$ towards the point O_σ until the distance between $\widehat{O}$ and any point $\overline{C} \in \overline{\sigma}$ is equal to 2^{i+1} (Figure 4.5). If we never hit the distance 2^{i+1} then $d(O_\sigma, \overline{C}) < 2^{i+1}$ for any point $\overline{C} \in \overline{\sigma}$ and $d(O_\sigma, C) < 2^{i-1}$ for any point $C \in \sigma$, so $A \subset B_{2^{i+1}}(O_\sigma)$. Suppose we hit the distance $d(\widehat{O}, \overline{C}) = 2^{i+1}$. We claim that for any point D on the geodesic ray $\widehat{O}O_\sigma$ we have $d(\widehat{O}, \overline{C}) < d(D, \overline{C})$. Suppose, by contradiction, it is not the case. Then, since the ray is infinitely long, there exists a point $F \neq \widehat{O}$ on a geodesic ray $\widehat{O}O_\sigma$, such that $d(F, \overline{C}) = d(\widehat{O}, \overline{C}) = 2^{i+1}$. However, similarly, if we consider the other ray of the same geodesic line starting from $O_{\overline{\sigma}}$, there will be a point $\overline{F}$ on this ray, such that $d(\overline{F}, \overline{C}) = d(\widehat{O}, \overline{C}) = 2^{i+1}$, since $d(O_{\overline{\sigma}}, \overline{C}) < 2^{i+1}$. But then we have three points on a geodesic line F, $\overline{F}$, and $\widehat{O}$, such that $d(\overline{C}, F) = d(\overline{C}, \overline{F}) = d(\overline{C}, \widehat{O}) = 2^{i+1}$, which contradicts Lemma 4.2. Suppose $d(\widehat{O}, O_\sigma) \geq 3 \cdot 2^{i-1}$. Since for any point D on the geodesic ray $\widehat{O}O_\sigma$ we have $d(\widehat{O}, \overline{C}) < d(D, \overline{C})$, we also have $\angle \overline{C}\widehat{O}O_\sigma \geq 90^\circ$ by Corollary 4.1. Then by the cosine theorem for the comparison triangle of the triangle $\triangle(\overline{C}, \widehat{O}, O_\sigma)$ we get $d(\overline{C}, O_\sigma)^2 \geq d(\overline{C}, \widehat{O})^2 + d(\widehat{O}, O_\sigma)^2 \geq (2^{i+1})^2 + (3 \cdot 2^{i-1})^2 = 25 \cdot (2^{i-1})^2$, so $d(\overline{C}, O_\sigma) \geq 5 \cdot 2^{i-1}$. But because σ creates $\overline{\sigma}$, there is $C \in \sigma$ such that C is a descendant of $\overline{C}$. So the distance $d(C, \overline{C}) < 2^{i+1}$, hence $d(\overline{C}, O_\sigma) \leq d(\overline{C}, C) + d(C, O_\sigma) < 5 \cdot 2^{i-1}$. Contradiction. Hence, $d(\widehat{O}, O_\sigma) < 3 \cdot 2^{i-1}$. Then for any vertex $G \in \sigma$ $d(\widehat{O}, G) \leq d(\widehat{O}, O_\sigma) + d(O_\sigma, G) < 3 \cdot 2^{i-1} + 2^{i-1} = 2^{i+1}$. This means that $\sigma \in B_{2^{i+1}}(\widehat{O})$. Also, by construction, $\overline{\sigma} \in B_{2^{i+1}}(\widehat{O})$. The above facts imply the existence of a point O such that both $\sigma \in B_{2^{i+1}}(O)$ and $\overline{\sigma} \in B_{2^{i+1}}(O)$, so α exists in K_{i+1}.

□

In the original proof of Lemma 3.3, we used Lemma 3.2 which employs hyperplanes. Unfortunately, such an approach does not extend to Hadamard spaces. However, similar arguments with some adjustments allow us to prove the result for Hadamard spaces as well.

Lemma 4.5 (***Main lemma 2***)**.** Consider a simplex $\overline{\sigma} \in T_i$ such that $\overline{\sigma} \notin K_{i+1}$. Consider also the family of all simplices $\sigma_j \in K_{i-1}$ that "create" simplex $\overline{\sigma}$. Let A be the union of vertices of $\bigcup_j \sigma_j$. Then A spans a simplex $\alpha \in K_{i+1}$.

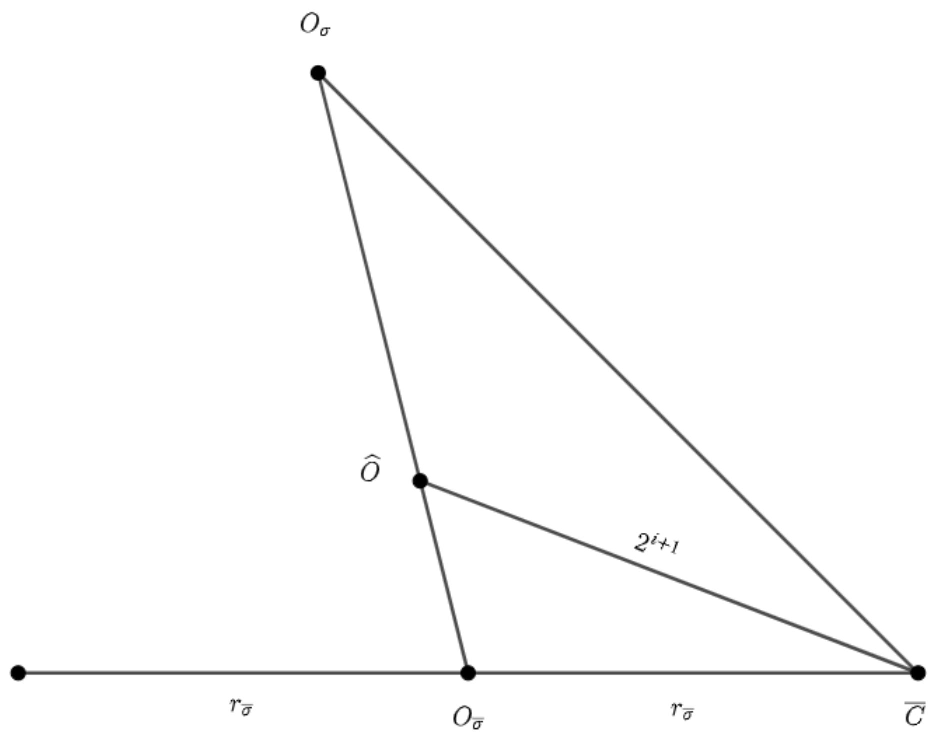

Figure 4.5: Supplementary figure for Lemma 4.5. The point $\widehat{O}$ can be chosen as O, i.e., $A \subset B_{2^{i+1}}(\widehat{O})$

Proof. Let $O_{\overline{\sigma}}$ be the center of a ball $B_{\overline{\sigma}}$ with the minimal radius $r_{\overline{\sigma}}$ containing simplex $\overline{\sigma}$. Since $\overline{\sigma} \notin K_{i+1}$ then $r_{\overline{\sigma}} > 2^{i+1}$. Consider any simplex $\sigma \in \bigcup_j \sigma_j$. Let O_σ be the center of the ball with the minimal radius r_σ containing simplex σ. Because $\sigma \in K_{i-1}$, we have $r_\sigma < 2^{i-1}$. Similarly to the original proof, we want to show that $d(O_{\overline{\sigma}}, O_\sigma) < 3 \cdot 2^{i-1}$. By contradiction, suppose that $d(O_{\overline{\sigma}}, O_\sigma) \geq 3 \cdot 2^{i-1}$. Consider any point $\overline{C} \in \overline{\sigma}$ which lies on the boundary of $B_{\overline{\sigma}}$, i.e., $d(\overline{C}, O_{\overline{\sigma}}) = r_{\overline{\sigma}} > 2^{i+1}$. Suppose $\angle \overline{C} O_{\overline{\sigma}} O_\sigma \geq 90°$. Then by the cosine theorem for a comparison triangle of the triangle $\triangle(\overline{C}, O_{\overline{\sigma}}, O_\sigma)$ we have $d(\overline{C}, O_\sigma)^2 \geq d(\overline{C}, O_{\overline{\sigma}})^2 + d(O_{\overline{\sigma}}, O_\sigma)^2 \geq (2^{i+1})^2 + (3 \cdot 2^{i-1})^2 = 25 \cdot (2^{i-1})^2$, so $d(\overline{C}, O_\sigma) \geq 5 \cdot 2^{i-1}$. But, as was already shown, $d(\overline{C}, O_\sigma) < 5 \cdot 2^{i-1}$. This leads to a contradiction. Then, for any such $\overline{C}$, $\angle \overline{C} O_{\overline{\sigma}} O_\sigma < 90°$. But by Lemma 4.3, if we start moving the point $O_{\overline{\sigma}}$ towards the point O_σ, the distance to the point $\overline{C}$ decreases. This is true for any point $\overline{C}$ on a boundary $B_{\overline{\sigma}}$. Hence, we can slightly move the point $O_{\overline{\sigma}}$ towards the point O_σ, so that the distance $d(\overline{C}, O_{\overline{\sigma}}) < r_{\overline{\sigma}}$ for any point $\overline{C}$ on a boundary of $B_{\overline{\sigma}}$ and still having $d(\overline{D}, O_{\overline{\sigma}}) < r_{\overline{\sigma}}$ for any point $\overline{D} \in \overline{\sigma}$ inside the $B_{\overline{\sigma}}$. This contradicts the fact that $O_{\overline{\sigma}}$ is the center of a ball $B_{\overline{\sigma}}$ with the minimal radius $r_{\overline{\sigma}}$ containing simplex $\overline{\sigma}$. Therefore, $d(O_{\overline{\sigma}}, O_\sigma) < 3 \cdot 2^{i-1}$. Then for any vertex

$D \in \sigma$ $d(O_{\overline{\sigma}}, D) \leq d(O_{\overline{\sigma}}, O_\sigma) + d(O_\sigma, D) < 3 \cdot 2^{i-1} + 2^{i-1} = 2^{i+1}$. This means that $\sigma \in B_{2^{i+1}}(O_{\overline{\sigma}})$. The above is true for any $\sigma \in \bigcup_j \sigma_j$. This implies that $A \subset B_{2^{i+1}}(O_{\overline{\sigma}})$, hence α exists in K_{i+1}. $\square$

In Lemma 3.5, hyperplanes and parallel translation of hyperplanes were the keys to the proof. Unfortunately, as we already discussed, we cannot use hyperplanes in an Hadamard space. While we were not able to find a workaround for the original proof, we did find a different approach to the proof. We still need to employ an idea similar to the one in presented in Lemma 3.4, in the sense that we should be able to move a point in some direction and decrease the distance to a certain set of points.

Lemma 4.6. Let A, B_1, B_2, ... , B_n be a set of points in an Hadamard space such that $d(A, B_i) = d(A, B_j)$ for all i, j. Let O be the center of a ball with the minimal radius r containing points $\bigcup_j B_j$. Then there exists $\epsilon_1 > 0$ such that for any $0 < \epsilon_2 < \epsilon_1$ and a point C on the geodesic segment AO satisfying $d(A, C) < \epsilon_2$ the distance $d(B_j, A) > d(B_j, C)$ for any j. In other words, if we are moving the point C starting from A towards O, then the distance to B_j, $d(B_j, C)$ is decreasing.

Proof. Consider any B_i. Suppose $\angle B_i AO \geq 90°$. Then $B_i O$ is the largest side of the triangle $\triangle(A, B_i, O)$. But then $d(A, B_i) < d(B_i, O)$. Hence A is closer to B_j than O for all j, so O cannot be the center of the ball of minimal radius containing points $\bigcup_j B_j$. Contradiction. Hence, $\angle B_i AO < 90°$ for all i. After that, Lemma 4.3 finishes the proof. $\square$

The new proof is quite technical, but the idea behind it can be roughly described as follows. Recalling the statement of the lemma, we see that we may prove it if we show that the vertices of interest belong to an appropriate ball. We will define a function on the ambient space whose minimizer coincides as the center of such a ball. To will prove the latter claim by contradiction, assuming that the minimizer is not the center of the sought ball. We will create a finite set of points with a certain "nice property", and we will show that the convex hull of this set (i.e. the smallest convex set containing this set) satisfies this property as well. To prove the latter, we show in a separate lemma that a convex set of a finite set of points can be constructed in a concrete way. In a Euclidean space, the convex hull of a finite set of points will be closed, but that might not be the case in an Hadamard space. Hence, we will show in another lemma that the closure of the convex hull is convex. Also, we will prove in a separate lemma that the center of the minimal ball containing the original set is contained in this closed convex hull. Thus, this center will have the "nice property" as well. Combined with other properties of the center of the minimal ball, it will give us the contradiction to the original point being a minimizer.

We will also need to prove that our function actually attains its minimum. In order to do this, we will introduce the notion of a convex function in an Hadamard space and state some properties of these functions (see [2] for details).

Definition 4.1. Let X be a geodesic space. A function $f : X \to \mathbb{R}$ is said to be convex if its restriction to any constant-speed geodesic $\gamma(t) : \mathbb{R} \to X$ is convex. That is, $f \circ \gamma$ is a convex function, $f(\gamma(tx + (1-t)y)) \leq tf(\gamma(x)) + (1-t)f(\gamma(y))$, $0 \leq t \leq 1$, $x, y \in \mathbb{R}$.

Definition 4.2. Let X be a geodesic space and $\lambda > 0$. A function $f : X \to \mathbb{R}$ is λ-convex if for any unit-speed geodesic $\gamma \in X$, the function $t \to f(\gamma(t)) - \lambda t^2$ is convex.

A function that is λ-convex for some $\lambda > 0$ is called strongly convex.

Theorem 4.3. A continuous function $f : X \to \mathbb{R}$ is λ-convex if and only if for any $x, y \in X$ and z the midpoint between x and y satisfies

$$f(z) \leq \frac{f(x) + f(y)}{2} - \frac{\lambda}{4} d(x, y)^2.$$

Theorem 4.4. For every point p in an Hadamard space, the function $d_p^2(x) = d(p, x)^2$ is 1-convex.

Theorem 4.5. Let X be a complete space with a strictly intrinsic metric, and let $f : X \to \mathbb{R}$ be a continuous strongly convex function bounded from below. Then f has a unique minimum point

We now have all the tools we need to prove the lemma about existence of a minimizer for a function that we will employ later. For convenience, we will slightly abuse notation and write $\bigcup_j A_j$ instead of $\bigcup_j \{A_j\}$ to denote the union of singletons. We shall also omit the (finite) index set for j when it is not used.

Lemma 4.7. Let S be a convex closed set. Let $\bigcup_j B_j$ be a finite set of points. Let $g(x) = \max_{B \in \bigcup_j B_j} d(x, B)$. There exists a point $D \in S$ such that $g(D) = \inf_{C \in S} g(C)$

Proof. We will prove that $g^2(x)$ is 1-convex in S. According to Theorem 4.3, we want to show that $g(z)^2 \leq \dfrac{g(x)^2 + g(y)^2}{2} - \dfrac{d(x,y)^2}{4}$ where $x, y \in S$ and z is a midpoint between x and y. Obviously, since S is convex, then $z \in S$. Note that $g(z) = d(z, B)$ for some $B \in \bigcup_j B_j$. Also, Notice that

$g(x) \geq d(x,B)$ and $g(y) \geq d(y,B)$. By Theorem 4.4 function d_B^2 is 1-convex. So

$$g(z)^2 = d(z,B)^2 \leq \frac{d(x,B)^2 + d(y,B)^2}{2} - \frac{d(x,y)^2}{4} \leq \frac{g(x)^2 + g(y)^2}{2} - \frac{d(x,y)^2}{4},$$

which proves that $g(x)^2$ is 1-convex. Hence, according to Theorem 4.5, there is point in S, which minimize function $g(x)^2$. Clearly, the same point minimizes $g(x)$ which finishes the proof. $\square$

Let us now consider the topic of the convex hull of a finite set of points in an Hadamard space.

Lemma 4.8. Let $S = \{A_1, A_2, ..., A_n\}$ be a set of points in an Hadamard space. Let $S_1, S_2, \dots$ be a sequence of sets constructed as follows. Take $S_1 = S$. For all pairs of points $x, y \in S_i$ consider a geodesic segment connecting these two points and take union of all such geodesic segments to construct S_{i+1}. Then $coS = \bigcup_j S_j$ is the convex hull of set S.

Proof. First, let us prove that coS is a convex set. Consider any two points $x, y \in S$. There exist i_1 and i_2, such that $x \in S_{i_1}$ and $y \in S_{i_2}$. Without loss of generality, suppose $i_1 \geq i_2$. Hence, both $x, y \in S_{i_1}$. Then, by construction, the geodesic segment connecting x and y must be in S_{i_1+1}. So, this geodesic segment belongs to coS. Therefore, coS is a convex set.

Now, let us prove that coS is the smallest convex set containing S. Let $\widehat{S}$ be the convex hull of S. We will prove that for any j, $S_j \subseteq \widehat{S}$. If this is true, then $coS = \bigcup_j S_j \subseteq \widehat{S}$. But since $\widehat{S}$ is a minimal convex set, then $coS = \widehat{S}$. Let us prove the above fact by induction. Clearly, $S_1 \subseteq \widehat{S}$. Suppose $S_i \subseteq \widehat{S}$. Then for any two points $x, y \in S_i$, the geodesic segment connecting these two points belong to $\widehat{S}$. Hence, $S_{i+1} \subseteq \widehat{S}$. Therefore $coS = \widehat{S}$, so coS is the convex hull of S. $\square$

Lemma 4.9. Let (X, d) be an Hadamard space and $S \subset X$. Then $\overline{coS}$ is a convex set, where $\overline{coS}$ is a closure of coS

Proof. We will first prove that a geodesic segment connecting a point on the boundary of coS, i.e. in the set $\overline{coS} \setminus coS$, with any point from coS belongs to $\overline{coS}$. Then we will prove the same statement for two points on the boundary of coS. Let A be a point on the boundary of coS, and $B \in coS$. Suppose the geodesic segment γ connecting A and B is not contained in $\overline{coS}$. Then there exist $D \in \gamma$ and $\epsilon > 0$ such that $d(D, \overline{coS}) > \epsilon$. Since A belongs to the boundary of coS, there exists $C \in coS$ such that $d(A,C) < \epsilon$. Lets consider a comparison triangle $\triangle(\overline{A}, \overline{B}, \overline{C})$ for the triangle $\triangle(A,B,C)$. Let $\overline{D}$ be a point on segment $\overline{AB}$, such that $\overline{d}(\overline{A}, \overline{D}) = d(A,D)$. It is easy to see that distance from $\overline{D}$ to the segment $\overline{BC}$ is less then $\overline{d}(\overline{A}, \overline{C}) < \epsilon$. So, there is point $\overline{E} \in \overline{BC}$ such that $\overline{d}(\overline{D}, \overline{E}) < \epsilon$.

Let E be a point on the geodesic segment BC such that $d(C, E) = \overline{d}(\overline{C}, \overline{E})$ (Figure 4.6). Then by $CAT(0)$ inequality $d(D, E) < \overline{d}(\overline{D}, \overline{E}) < \epsilon$. But $E \in \overline{coS}$ since the geodesic segment connecting B and C belongs to $\overline{coS}$. This contradicts the fact that $d(D, \overline{coS}) > \epsilon$. Thus, we established that a geodesic segment connecting a point on the boundary of coS and a point from coS belongs to $\overline{coS}$. Using this fact, the proof for the case when two points lie on the boundary of coS follows the same argument.

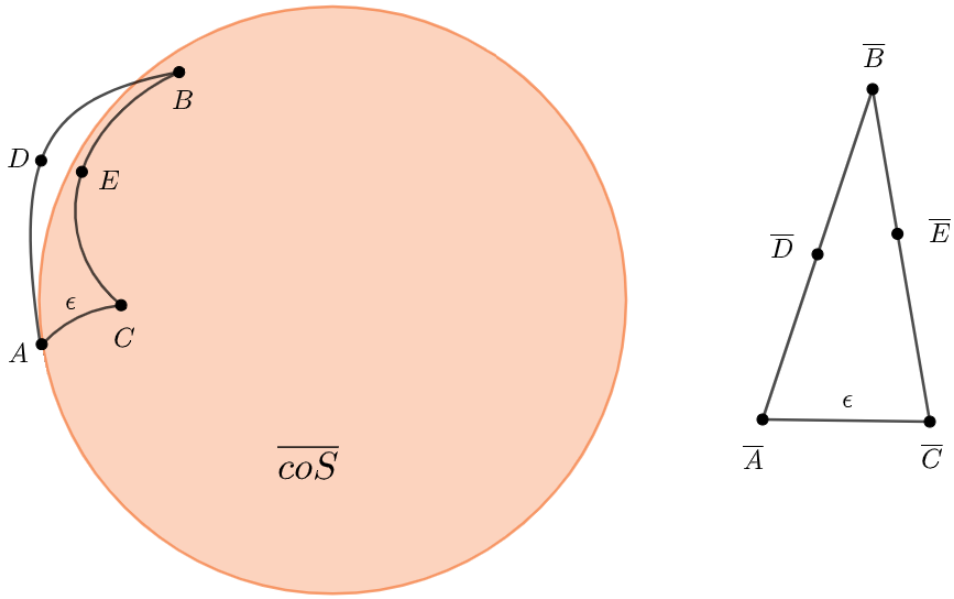

Figure 4.6: Supplementary figure for Lemma 4.9. The figure depicts the case when $B \in coS$. The picture for the case when B is on the boundary of coS can be obtained by simply moving the point B to the boundary.

$\square$

Lemma 4.10. Let $S = \{A_1, A_2, ..., A_n\}$ be a set of points in an Hadamard space. Let O be the center of the ball of minimal radius containing S. Then $O \in \overline{coS}$, where $\overline{coS}$ is a closure of coS.

Proof. By contradiction, suppose that $O \notin \overline{coS}$. Let $D \in \overline{coS}$ be such that $d(O, D) = d(O, \overline{coS}) > 0$. Such a point D exists according to Lemma 4.7. Since D is not the center of minimal ball containing S, then there is $A \in S$, such that $d(A, D) > d(A, O)$. Suppose $\angle ODA \geq 90°$. But then, OA is a greatest side in $\triangle(A, O, D)$, hence $d(O, A) > d(A, D)$. Contradiction. Therefore, $\angle ODA < 90°$. Consider geodesic segment γ connecting points A and D. By Lemma 4.9, $\overline{coS}$ is a convex set, so $\gamma \subseteq \overline{coS}$. Since $\angle ODA < 90°$, by Lemma 4.3, if we start moving point D towards A, then its distance to point O decreases. So, there exist point $E \in \gamma$ such that $d(O, E) < d(O, D)$ (Figure 4.7). But this contradicts the fact that $d(O, D) = d(O, \overline{coS})$. Hence, $O \in \overline{coS}$.

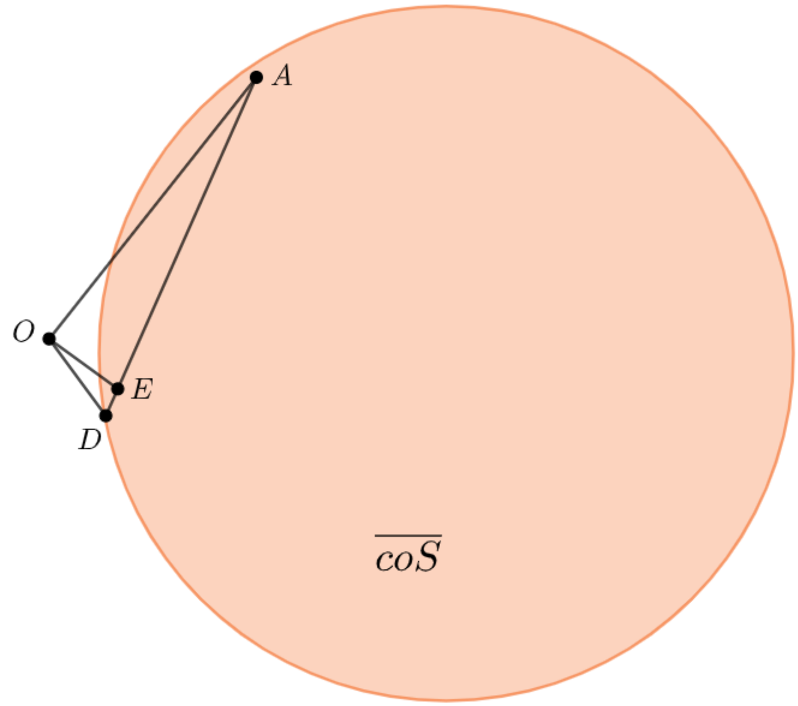

Figure 4.7: Supplementary figure for Lemma 4.10 which shows that $O \in \overline{coS}$

□

Now we have all the necessary ingredients to start proving Lemma 3.5 in the case of an Hadamard space.

Lemma 4.11 (***Main lemma 3***)**.** Consider a simplex $\overline{\sigma} \in T_i$ such that $\overline{\sigma} \notin K_{i+1}$ such that there exists a face $\widehat{\sigma} \subset \overline{\sigma}$ with $\widehat{\sigma} \in K_{i+1}$. Consider the family of all simplices $\sigma_j \in K_{i-1}$ that "creates" simplex $\overline{\sigma}$. Let A be the union of vertices in $\bigcup_j \sigma_j$ together with vertices of $\widehat{\sigma}$. Then A spans a simplex $\alpha \in K_{i+1}$.

Proof. For each j, let O_j be the center of the ball with the minimal radius r_j containing all vertices of simplex σ_j. Define a function $f : X \to \mathbb{R}$ as $f(x) = \max_{O \in \bigcup_j O_j} d(x, O)$. Let set $S_O = \{x | f(x) \leq 3 \cdot 2^{i-1}\}$. Firstly, note that S_O is not empty. Indeed, it follows from the proof of Lemma 4.5. Secondly, note that S_O is an intersection of closed balls. Hence S_O is a closed convex set.

Let $\bigcup_j \widehat{C_j}$ be the union of all vertices of simplex $\widehat{\sigma}$. Define a function $g : X \to \mathbb{R}$ as $g(x) = \max_{\widehat{C} \in \bigcup_j \widehat{C_j}} d(x, \widehat{C})$. By Lemma 4.7, there exists a point $D \in S_O$, such that $g(D) = \min_{C \in S_O} g(C)$. If $g(D) \leq 2^{i+1}$, then the ball of radius 2^{i+1} with center at D contains all vertices from simplex $\widehat{\sigma}$, and contains all vertices from simplex σ_j for any j, since the distance to O_j is less or equal to $3 \cdot 2^{i-1}$ and $r_j \leq 2^{i-1}$. So α exists in K_{i+1}.

Suppose $g(D) > 2^{i+1}$. First, let's prove that D lies on the boundary of S_O, i.e. there exists O_j, such that $d(D, O_j) = 3 \cdot 2^{i-1}$, or $f(D) = 3 \cdot 2^{i-1}$. Suppose $f(d) < 3 \cdot 2^{i-1} - \epsilon_1$ for some $\epsilon_1 > 0$. Let $P_{\widehat{C}} \subset \bigcup_j \widehat{C_j}$ be a set of points such that $g(D) = d(D, \widehat{C})$ for any $\widehat{C} \in P_{\widehat{C}}$. Let $\widehat{O}$ be the center of the ball of minimal radius containing $P_{\widehat{C}}$. Then, by Lemma 4.6, if we start moving point D towards $\widehat{O}$, the distance to all points from $P_{\widehat{C}}$ will be decreasing. For all points $\widehat{C} \in \bigcup_j \widehat{C_j}$, $\widehat{C} \notin P_{\widehat{C}}$ we have $d(D, \widehat{C}) > g(D)$. So we can choose $\epsilon_2 > 0$ such that $d(D, \widehat{C}) - \epsilon_2 > g(D)$ for all such $\widehat{C}$. Then we can move point D towards point $\widehat{O}$ by a small enough ϵ to D' (which satisfies conditions of Lemma 4.6 above and is less than ϵ_1 and ϵ_2), so the distance to all points from $P_{\widehat{C}}$ decreases, the distance to all other points from $\bigcup_j \widehat{C_j}$ still less than $g(D)$, and still $f(D') < 3 \cdot 2^{i-1}$. This means that $g(D') < g(D)$ and $D' \in S_O$, which contradicts the fact that $g(D) = \min_{C \in S_O} g(C)$. So $f(D) = 3 \cdot 2^{i-1}$.

Let $P_O \subseteq \bigcup_j O_j$ be a set of points such that $d(D, O) = 3 \cdot 2^{i-1}$ for all $O \in P_O$. Consider any $O \in P_O$ and any $\widehat{C} \in P_{\widehat{C}}$. If $\angle \widehat{C} DO \geq 90°$, then by the cosine theorem for comparison triangle of $\triangle(\widehat{C}, D, O)$ we have $d(\widehat{C}, O) \geq 5 \cdot 2^{i-1}$. But this is impossible, since there is a child of $\widehat{C}$, C, such that $d(O, C) \leq 2^{i-1}$ and $d(\widehat{C}, C) < 2^{i+1}$. Then $d(\widehat{C}, O) \leq d(\widehat{C}, C) + d(C, O) < 2^{i-1} + 2^{i+1} = 5 \cdot 2^{i-1}$. Hence, $\angle \widehat{C} DO < 90°$ for all $O \in P_O$ and $\widehat{C} \in P_{\widehat{C}}$.

Consider again any $O \in P_O$ and any $\widehat{C} \in P_{\widehat{C}}$. By Lemma 4.3 there exist $\epsilon_{\widehat{C},O}$ such that for any B on geodesic segment $D, \widehat{C}$ and $d(B, D) < \epsilon_{\widehat{C},O}$, $d(B, O) < d(D, O)$. We can choose $\epsilon_{\widehat{C}}$ small enough such that $\epsilon_{\widehat{C}} < \min_{O \in P_O} \epsilon_{\widehat{C},O}$ and if B is a point on the geodesic segment $D, \widehat{C}$ satisfying $d(D, B) = \epsilon_{\widehat{C}}$, then still $d(B, \widetilde{O}) < 3 \cdot 2^{i-1}$, where $\widetilde{O}$ is any point from $\bigcup_j O_j \setminus P_O$. Then $f(B) < 3 \cdot 2^{i-1}$. Let $\epsilon = \min_{\widehat{C} \in P_{\widehat{C}}} \epsilon_{\widehat{C}}$. For any $\widehat{C_j} \in P_{\widehat{C}}$ take the point $\widehat{C}_{j,\epsilon}$ on the geodesic segment $\widehat{C_j} D$ such that $d(D, \widehat{C_{j,\epsilon}}) = \epsilon$ (Figure 4.8. Note that $f(\widehat{C_{j,\epsilon}}) < 3 \cdot 2^{i-1}$.

Now, lets construct the convex hull S of $\bigcup_j \widehat{C_{j,\epsilon}}$ as we did in Lemma 4.8. It is important to note that at any step of constructing this convex hull, any point B that we add has the following property: $f(B) < 3 \cdot 2^{i-1}$. It follows from Lemma 4.1 and the fact that for all original points $\widehat{C_{j,\epsilon}}$, $f(\widehat{C_{j,\epsilon}}) < 3 \cdot 2^{i-1}$. By Lemma 4.10, the center $\widehat{O_\epsilon}$ of the ball with the minimal radius r_ϵ containing

$\bigcup \widehat{C_{j,\epsilon}}$ belongs to S. Hence $f(\widehat{O_\epsilon}) < 3 \cdot 2^{i-1}$, so $\widehat{O_\epsilon} \in S_O$.

The last step is to check that $g(\widehat{O_\epsilon}) < g(D)$. Recall that $g(x) = \max_{\widehat{C} \in \bigcup_j \widehat{C_j}} d(x, \widehat{C})$. First, consider any $\widehat{C_j} \in P_{\widehat{C}}$. Since $f(D) = 3 \cdot 2^{i-1} > f(\widehat{O_\epsilon})$, points D and $\widehat{O_\epsilon}$ are different. Since $d(D, \widehat{C_{j,\epsilon}}) = \epsilon$, we have $r_\epsilon < \epsilon$. Then by triangle inequality $d(\widehat{O_\epsilon}, \widehat{C_j}) \leq d(\widehat{O_\epsilon}, \widehat{C_{j,\epsilon}}) + d(\widehat{C_{j,\epsilon}}, \widehat{C_j}) = r_\epsilon + d(\widehat{C_j}, D) - \epsilon < d(D, \widehat{C_j}) = g(D)$. Now consider any $\widehat{C_j} \notin P_{\widehat{C}}$. Clearly, $d(D, \widehat{C_j}) < g(D)$. The claim is that we still can choose our ϵ small enough so $d(\widehat{O_\epsilon}, \widehat{C_j}) < g(D)$. Indeed, by triangle inequality $d(\widehat{O_\epsilon}, \widehat{C_j}) \leq d(\widehat{O_\epsilon}, D) + d(D, \widehat{C_j}) \leq r_\epsilon + \epsilon + d(D, \widehat{C_j}) < d(D, \widehat{C_j}) + 2\epsilon$. Clearly, we can choose ϵ small enough, such that $d(D, \widehat{C_j}) + 2\epsilon < g(D)$. Hence, $d(\widehat{O_\epsilon}, \widehat{C_j}) < g(D)$ for all $\widehat{C_j} \notin P_{\widehat{C}}$ and for all $\widehat{C_j} \in P_{\widehat{C}}$. Consequently, $g(\widehat{O_\epsilon}) < g(D)$. But this contradicts the fact that $g(D) = \min_{C \in S_O} g(C)$ since $g(\widehat{O_\epsilon}) < g(D)$ and $\widehat{O_\epsilon} \in S_O$. Therefore, $g(D)$ cannot be greater than 2^{i+1}, which finishes the proof.

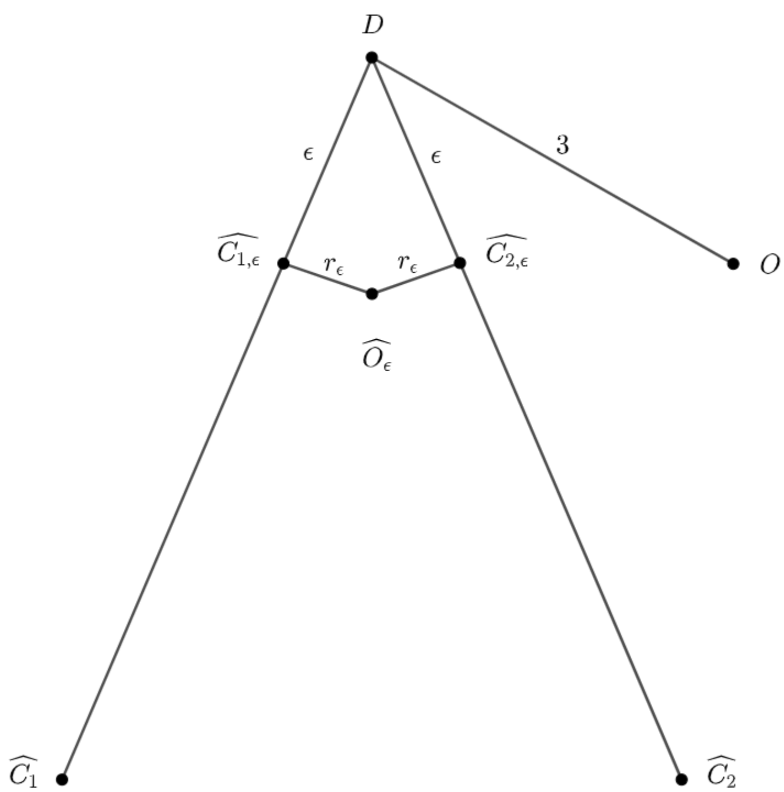

Figure 4.8: Supplementary figure for Lemma 4.11. The general idea is to show that $\widehat{O_\epsilon}$ can be picked instead of D

$\square$

CHAPTER 5
COMPUTATION AND ALGORITHM

Our theoretical result establishes an interleaving between persistence modules obtained via cover tree refinements and the ones obtained using the usual Čech construction. However, the simplicial complexes constructed from the cover refinements are also Čech-like and not Rips-like, in a sense that our condition for inclusion in the complex is checked for the whole simplex, not just its edges. Consequently, as is true for the usual Čech complexes, the required computations may be too costly. Note however that our cover tree based approach also allows for some sort of a Vietoris-Rips construction: instead of adding a simplex to the simplicial complex T_i at level i of the cover tree when there is a simplex on its descendants in K_{i-1}, we can add a simplex to T_i as soon as all of its edges showed up. It is easy to see that there still going to be well a defined simplicial map that maps children to their parents $g_i : \widehat{T_{i-1}} \to \widehat{T_i}$, where $\widehat{T_i}$ is a simplicial complex on a cover tree which is built using the Vietoris-Rips construction. And, as we said in Chapter 2.1.1, construction of simplicial complexes using the Vietoris-Rips construction instead of the Čech one is significantly computationally cheaper. But our theoretical result holds only for Čech construction. Hence, it would be nice to have some sort of a connection between Čech and Vietoris-Rips constructions. We already know that $\check{C}ech(r) \subset$ *Vietoris-Rips*(r). It turns out there is also an inclusion in the opposite direction (see e.g. [11]).

Lemma 5.1 (Vietoris-Rips Lemma)**.**
Let S be a finite set of points in a Euclidean space and let $r \geq 0$. Then we have *Vietoris-Rips*$(r) \subseteq \check{C}ech(\sqrt{2}r)$.

For the case of Hadamard spaces, and any other metric space, we can use a little bit more rough estimate: *Vietoris-Rips*$(r) \subseteq \check{C}ech(2r)$. The proof is straightforward, since if we know that the distance between any pair of the points is less than or equal to $2r$, then all of these points lie inside a closed ball of radius $2r$.

Thus, we can construct the following commutative diagram:

$$
\begin{tikzcd}
\cdots \arrow[r] & H_p(K_{i-1}) \arrow[r] \arrow[d] & H_p(K_i) \arrow[r] \arrow[d] & H_p(K_{i+1}) \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & H_p(\widehat{K_{i-1}}) \arrow[r] \arrow[ur] & H_p(\widehat{K_i}) \arrow[r] \arrow[ur] & H_p(\widehat{K_{i+1}}) \arrow[r] & \cdots
\end{tikzcd}
$$

Here, $K_i = \check{C}ech(2^i)$, $\widehat{K_i} = \textit{Vietoris-Rips}(2^i)$, and maps between homology groups are induced by simplicial inclusion maps. If D_K is a persistence diagram (at the logarithmic scale) of the persistent module obtained using the Čech construction and $\widehat{D_K}$ is a diagram (at the logarithmic scale) obtained suing the Vietoris-Rips construction, then $W_\infty(D_K, \widehat{D_K}) \leq 1$. This follows from the result in [3] regarding strongly interleaved persistence modules

Exactly the same thing happens when using the cover tree approach. It is easy to check that since $\check{C}ech(2^i) \subseteq \textit{Vietoris-Rips}(2^i)$, then $T_i \subseteq \widehat{T}_i$. Similarly, since $\textit{Vietoris-Rips}(2^i) \subseteq \check{C}ech(2^{i+1})$, then $\widehat{T}_i \subseteq T_{i+1}$. We also have the following commutative diagram:

$$
\begin{tikzcd}
\cdots \arrow[r] & H_p(T_{i-1}) \arrow[r] \arrow[d] & H_p(T_i) \arrow[r] \arrow[d] & H_p(T_{i+1}) \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & H_p(\widehat{T_{i-1}}) \arrow[r] \arrow[ur] & H_p(\widehat{T}_i) \arrow[r] \arrow[ur] & H_p(\widehat{T_{i+1}}) \arrow[r] & \cdots
\end{tikzcd}
$$

Consequently, the distance between persistence diagrams (at the logarithmic scale) obtained using the two construction is also bounded by 1. Together with Theorem 3.1, we get $W_\infty(\widehat{D_T}, \widehat{D_K}) \leq W_\infty(\widehat{D_T}, D_T) + W_\infty(D_T, D_K) + W_\infty(D_K, \widehat{D_K}) \leq 5$.

The above discussion suggests that in practice we can use the Vietoris-Rips construction, which significantly simplifies calculations.

There is another important computational aspect of the construction of simplicial complex $\widehat{T}_i$ that we need to mention. In order to build this simplicial complex, we need to figure out which edges should be added to this complex. Our criterion tells us that we add an edge at level i between two points $\overline{u}, \overline{v} \in C_i$ if and only if there are two descendants u, v of $\overline{u}, \overline{v}$ respectively such that $d(u,v) \leq 2^{i-1}$. Let S_u be the set of all descendants of $\overline{u}$ and S_v be the set of all descendants of $\overline{v}$. We want to understand if the distance between these two sets is greater than 2^{i-1}. In order to determine this, we can modify the algorithm for finding nearest neighbors from [1].

Algorithm 1 __Distance between sets__ **(cover tree T, level i, point $\overline{u}$, point $\overline{v}$)**

1. Set $Q_{u,i} = \overline{u}$, $Q_{v,i} = \overline{v}$.
2. for j from i down do $-\infty$:
   (a) Set $Q_u = \{\text{Children}(q) : q \in Q_{u,j}\}$, $Q_v = \{\text{Children}(q) : q \in Q_{v,j}\}$.
   (b) Form set $Q_{u,j-1} = \{q \in Q_u : \exists p \in Q_v, d(p,q) \leq d(Q_u, Q_v) + 2^j\}$, $Q_{v,j-1} = \{q \in Q_v : \exists p \in Q_u, d(p,q) \leq d(Q_u, Q_v) + 2^j\}$.
3. return $\text{argmin}_{(p,q)\in(Q_{u,-\infty}, Q_{v,-\infty})} d(p,q)$.

Unfortunately, analysis of the running time of the algorithm above is highly nontrivial. The issue is that the existing analysis of the algorithm for finding nearest neighbors using the cover tree does not extend to the case of finding the closest pair of points. Our intuition suggests that the running time of the algorithm above is likely to be dominated (at least in most cases) by the running time of computing persistent homology (or cohomology). Therefore, we are confident that in practice the speed-up when computing persistent homology using the cover tree approach will be very significant.

CHAPTER 6
CONCLUSIONS AND FUTURE WORK

We have shown that it is possible to leverage an existing efficient data structure called a cover tree to construct a family of simplicial complexes connected by simplicial maps representing cover refinements of the unknown topological space underlying a given data set. This family may provide a substantial speed-up when computing persistent homology and cohomology, and can be viewed, in a sense, as augmenting the choice of landmarks for the vertex set at each level. We also show that our nontrivial procedure for deciding which simplices should be included in the simplicial complexes allowed us to prove the interleaving between the persistence modules obtained using our novel methodology and the usual Čech complexes constructed over the full data set. When combined with the interleaving between Čech and Vietoris-Rips constructions, our result yields a well-quantified level of the coarseness of our approach. When such a coarse persistent (co)homology estimate is appropriate, which may be the case in many practical applications with large data sets, our cover tree approach can significantly reduce the computational cost.

The novel cover tree approach to computing persistent (co)homology presents several directions for future work. Firstly, we believe that the result also holds for the cover tree base slightly less than 2. However, the proof will likely have to follow a completely different route, since none of our main geometrical lemmas will be true in this case.

Secondly, even though we understand that the computational cost of our cover tree approach is dominated by the computational cost incurred due to computing persistent homology or cohomology, it is still important to conduct a detailed analysis of its computational complexity.

Another direction of research that is important (and interesting) to pursue is trying to understand whether our results can be extended (in some way) to general spaces of non-positive curvature, where one can employ comparison triangles only locally, unlike in the Hadamard space, where one can appeal to this property globally.

We should also mention that persistent (co)homology computations can be done not only for the whole data set but also locally, thus capturing important local structures. It might be useful to investigate if our cover tree approach can be modified to allow for computations of local (co)homology.

BIBLIOGRAPHY

[1] A. Beygelzimer, S. Kakade, and J. Langford. Cover trees for nearest neighbor. In *Proceedings of the 23rd international conference on Machine learning*, pages 97–104, 2006.

[2] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33. American Mathematical Society, 2022.

[3] F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas, and S. Y. Oudot. Proximity of Persistence Modules and Their Diagrams. In *Proceedings of the Twenty-Fifth Annual Symposium on Computational Geometry*, SCG '09, pages 237–246, 2009.

[4] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.

[5] V. De Silva and G. E. Carlsson. Topological estimation using witness complexes. In *PBG*, pages 157–166, 2004.

[6] V. De Silva and R. Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7(1):339–358, 2007.

[7] A. Dirafzoon and E. Lobaton. Topological mapping of unknown environments using an unlocalized robotic swarm. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5545–5551. IEEE, 2013.

[8] A. Dirafzoon and E. Lobaton. Topological mapping of unknown environments using an unlocalized robotic swarm. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5545–5551, 2013.

[9] A. N. Duman and H. Pirim. Gene coexpression network comparison via persistent homology. *International journal of genomics*, 2018, 2018.

[10] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 454–463, 2000.

[11] H. Edelsbrunner and J. Harer. *Computational topology: an introduction.* American Mathematical Soc., 2010.

[12] R. Ehrenborg and G. Hetyei. The topology of the independence complex. *European Journal of Combinatorics*, 27(6):906–923, 2006.

[13] K. Emmett, D. Rosenbloom, P. Camara, and R. Rabadan. Parametric inference using persistence diagrams: A case study in population genetics. *arXiv preprint arXiv:1406.4582*, 2014.

[14] R. Ghrist and A. Muhammad. Coverage and hole-detection in sensor networks via homology. In *IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, 2005.* Pages 254–260. IEEE, 2005.

[15] A. Hatcher. *Algebraic topology.* 2005.

[16] S. Mandal. *Applications of Persistent Homology and Cycles.* The Ohio State University, 2020.

[17] D. Morozov. Welcome to Dionysus' documentation! `www.mrzv.org/software/dionysus/`.

[18] M. Nicolau, A. J. Levine, and G. Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17):7265–7270, 2011.

[19] N. Singh, H. D. Couture, J. Marron, C. Perou, and M. Niethammer. Topological descriptors of histology images. In *International Workshop on Machine Learning in Medical Imaging*, pages 231–239. Springer, 2014.