A 3D Extension of the Sierpinski Relatives and an Analysis of their Persistent Homology

by

Joseph Geisz

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Masters of Science

Department of Applied Mathematics

2020

This thesis entitled:

A 3D Extension of the Sierpinski Relatives and an Analysis of their Persistent Homology

written by Joseph Geisz

has been approved for the Department of Applied Mathematics

_____

Dr. James Meiss

_____

Dr. Elizabeth Bradley

_____

Dr. Juan Restrepo

Date: _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Abstract

Geisz, Joseph (M.S., Applied Mathematics)

3D Extension of the Sierpinski Relatives and an Analysis of their Persistent Homology

Thesis directed by Dr. James Meiss

The Sierpinski relatives are a class of fractals that have been studied and characterized extensively. They serve as an excellent introductory example of self-similarity. While topological information about these fractals can be derived analytically, it has been demonstrated that "persistent homology", a concept in topological data analysis, can be applied to discrete point-set approximations of the fractals in order to reveal information about the underlying structure of the set. In this thesis we describe the Sierpinski relatives and some of the mathematical foundations of persistent homology, a fascinating and evolving field at the intersection of pure mathematics and computer science. We then propose a new set of fractals that are analogous to the Sierpinski relatives but are 3-dimensional. This set contains many more fractals than the classical 2D relatives and has more topological and geometric variation amongst its elements. The set contains elements with a variety of fractal dimensions, and topological properties that are not possible in $\mathbb{R}^2$. We then demonstrate the use of persistent homology techniques to analyze the structure of these fractals.

Acknowledgements

I would like to thank the many people who supported me academically and otherwise throughout the duration of my studies.

To Jim Meiss for mentoring me and for discussing fractals, topology, and beyond with me for hours at a time. To Juan Restrepo and Liz Bradley for agreeing to be on my thesis committee as well as showing me beauty in chaos. To Sujeet Bhat, Jem Corcoran, Jim Curry, Anne Dougherty, Nikki Sanderson, Brian Zaharatos, and all of the CU staff and students who have taught me about math and life.

To Mom and Dad for the thankless job of giving your kids a home where we are always welcome and can succeed and for guidance in every aspect of living. To Will, Anne, and Charlie for being with me through plagues, fires, and growing up.

To Madelyn for endless love, support, and patience. To the Colorado Crew team and to the many friends I have made in Boulder and abroad. To the many people with whom I have maintained friendships since before college.

To all my family and friends, everyone who has supported me, I guess there would be a few billion less numbered fractals without you.

Contents

List of Figures

Chapter 1

Background

1.1 Simplicial Homology

From dynamical system attractors to real world data, we often wish to classify sets according to their general "shape". Some of the most intuitive ways to classify sets such as these end up being the most difficult to describe mathematically. For example, how many "parts" or distinct components make up the whole set? Is it composed of a singular object or a collection of separate pieces? Does the object have a hole like a donut where a string could pass through the middle? Or does it have multiple holes? Is the object hollow like a balloon or solid all the way through like a rock? Does it have many balloon-like voids like a block of swiss cheese?

1.1.1 Betti Numbers

Topology allows us to quantify such descriptors with Betti numbers. The kth Betti number or β_k essentially gives the number of k-dimensional holes in an object. The 0th Betti number describes the number of connected components in a set, or how many pieces make up the whole object. The 1st Betti number describes the number of holes through an object. The 2nd Betti number is the number of voids within 3d objects, like the inside of a balloon. Figure 1.1 illustrates the concept of Betti numbers. The true definition of a Betti number extends to higher dimensional spaces, but a set can only have a non-zero kth Betti number in $\mathbb{R}^n$ if $k < n$. Despite the relatively intuitive explanation of these numbers in 3D, it takes many deep mathematical concepts to describe them rigorously and compute them. For more in-depth descriptions see [1] or [2].

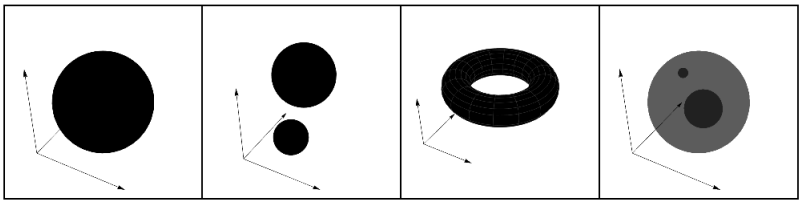


Figure 1.1: The first object (from left to right) has $\beta_0 = 1$ because it is only a single connected component. The second object has $\beta_0 = 2$ because it is composed of two disconnected components. The third, a solid torus, has $\beta_1 = 1$ because of the hole through the middle. The last, a ball with two balls removed from the interior, has $\beta_2 = 2$ because it contains two voids inside.

Most of Topology is not well suited for computers. Arbitrary topological spaces are very difficult or impossible to describe numerically. However simplicial complexes, a specific type of space, are easy to describe to a computer. Simplicial Homology, one of many homology theories, use simplicial complexes which "triangulate" a space to define homology groups and from these groups define Betti numbers. We explain the many ideas needed for this definition in the following section, and return to Betti numbers again when we have enough background to describe them rigorously.

1.1.2 Simplicial Complexes

Oriented Simplicies

Geometrically, an oriented k-simplex is the convex hull of k+1 geometrically independent points $\{x_0, x_1, ..., x_k\}$ in $\mathbb{R}^n$ with $k \leq n$. In $\mathbb{R}^3$ we can have 4 different kinds of oriented simplexes, shown in Figure 1.2. A 0-Simplex is a single point, a 1-Simplex is two points and the line between them, a 2-Simplex is 3 points and the triangle formed with the points as verticies, and a 3-Simplex is a 3D tetrahedron. Higher dimensional simplices cannot be visualized in 3 dimensions but can be thought of as extensions of these. We can denote a k-simplex as $\sigma^k = [x_0, x_1, ..., x_k]$, and can think of the simplex as purely an abstract or algebraic object independent of the geometry of the points.

The "faces" of a k-simplex are the (k-1)-simplices formed from taking the same points as the original simplex but removing one point. This means the faces of a 3-simplex or tetrahedron are the 4 triangles that make up its surface area, the faces of a triangle are the 3 line segments the perimeter is composed of, and the faces of a 1-simplex are the points at either end.

It is important to note that the order of the points listed in the simplex does matter, it defines

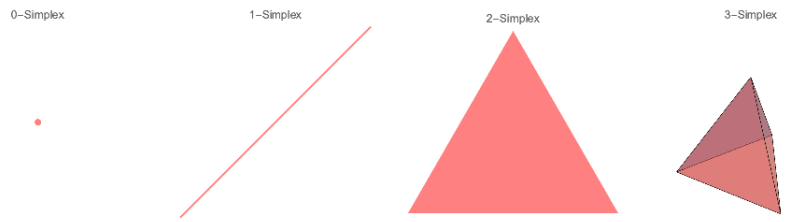


Figure 1.2: The first 4 types of simplices: a point, a line, a triangle, and a tetrahedron.

the orientation of the simplex. A simplex with the same points could represent the same simplex, or the "negative" simplex, one with the opposite orientation. In particular, given an ordering of the points in the simplex, any even permutation of the points will define a simplex with the same orientation as the original, and any odd permutation will give a simplex with negative orientation. The orientation is arbitrary but important to maintain for calculations of homology.

For example Figure 1.3 shows the tetrahedron represented by the simplicial complex $\sigma^3 = [x_0, x_1, x_2, x_3]$ We define the orientation of this simplex to be positive. Now consider a different simplex on the same points given as $\tilde{\sigma}^3 = [x_2, x_0, x_3, x_1]$. To get from the ordering of σ^3 to the ordering of $\tilde{\sigma}^3$ by switching two elements at a time we can perform the following:

$$\begin{aligned} &[x_0, x_1, x_2, x_3] \quad \text{switch 2 and 3} \\ &[x_0, x_1, x_3, x_2] \quad \text{switch 0 and 2} \\ &[x_2, x_1, x_3, x_0] \quad \text{switch 0 and 1} \\ &[x_2, x_0, x_3, x_1] \end{aligned}$$

Since this took 3 operations (an odd number) this is an odd permutation of the original ordering. Thus $\tilde{\sigma}^3$ has a negative orientation.

Simplicial Complexes

A simplicial complex is a set of oriented simplices with two required properties. The first is that a non-empty intersection of two simplices in the complex S must itself be a simplex in S, and the second is that any face of a simplex in S must also be in S. Since our points are in $\mathbb{R}^n$ a simplicial complex describes a subset of Euclidean space. We denote this region of space as $|S|$ and call it

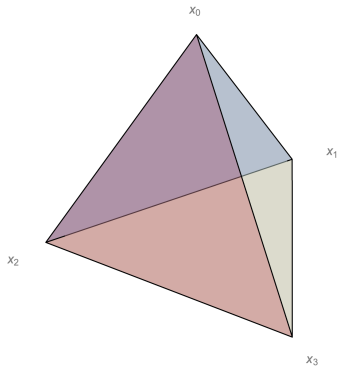


Figure 1.3: Example 3-Simplex

the "geometric realization of S". The kind of region that can be described by a simplicial complex is called a polytope or polyhedron.

We demonstrate the properties of a simplicial complex first with a true simplicial complex, and then with two sets of simplices which fail to have the required properties.

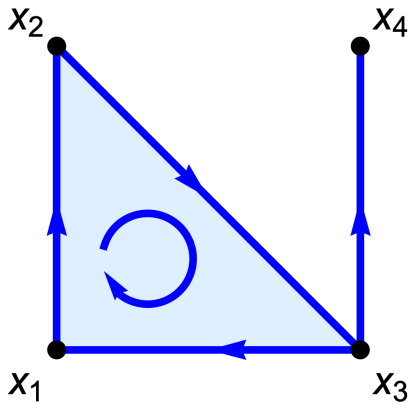


Figure 1.4: Example Simplicial Complex. Orientation is indicated with arrows.

First consider the simplicial complex S_1 below. Figure 1.4 shows the geometric realization of S_1 if $x_1 = (0,0), x_2 = (0,1), x_3 = (1,0), x_4 = (1,1)$.

$$S^1 = \{[x_1, x_2, x_3], [x_1, x_2], [x_2, x_3], [x_3, x_1], [x_3, x_4], [x_1], [x_2], [x_3], [x_4]\}$$

Note that if we take the intersection of any two simplices, we get a simplex that is still in the set, for example $[x_2, x_3] \cap [x_3, x_4] = [x_3] \in S^1$. Additionally any k-1 face of a k-simplex is also a simplex, for example all three edges of the triangle $[x_1, x_2, x_3]$ are also simplexes in S^1.

Now consider:

$$S^2 = \{[x_1, x_2], [x_2, x_3], [x_1], [x_3]\}$$

Note that $[x_1, x_2] \cap [x_2, x_3] = [x_2]$ is not in S^2; the set is not closed under intersection. Therefore S^2 is not a simplicial complex.

Then consider:

$$S^3 = \{[x_1, x_2, x_3], [x_1, x_2], [x_1], [x_2], [x_3]\}$$

Two faces of $[x_1, x_2, x_3]$ are not in the set, $[x_2, x_3]$ and $[x_3, x_1]$ so S^3 is also not a simplicial complex.

If we have a Topological Space X that is homeomorphic to a polytope $|S|$, we say X is a triangulated space and S is a triangulation of X. The triangulated space can have multiple different triangulations, however the homology groups for any triangulation are the same as the homology groups of X. Therefore finding the homology of some topological space can be reduced to calculating the homology groups of a simplicial complex, which is something that can be done computationally with well-studied algorithms. This is the foundation of most of computational topology [2]. However it is important to note that many sets, certain fractals for example, cannot be triangulated by a finite polytope.

1.1.3 Homology from Simplicial Complexes

Chain, Cycle, and Boundary Groups

We now discuss various groups which will lead us to the definition of a Homology Group. We introduce a non-trivial but relatively simple simplicial complex that we will use throughout the section as an example. We denote the complex as $\tilde{S}$ so as to avoid confusion with a general simplicial complex S. Figure 1.5 shows a geometric realization of $\tilde{S}$. Arrows on edges indicate the orientation of the 1-simplices and circular arrows indicate the orientation of the 2-simplices.

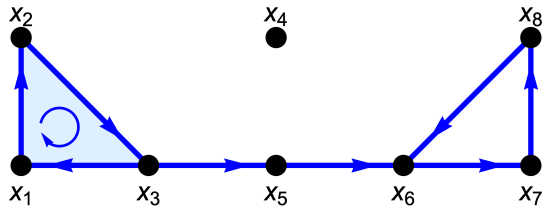


Figure 1.5: Example Simplicial Complex $\tilde{S}$

$$\tilde{S} = \{ \quad \sigma_1^0, \sigma_2^0, \sigma_3^0, \sigma_4^0, \sigma_5^0, \sigma_6^0, \sigma_7^0, \sigma_8^0, \quad \sigma_1^1, \sigma_2^1, \sigma_3^1, \sigma_4^1, \sigma_5^1, \sigma_6^1, \sigma_7^1, \sigma_8^1, \quad \sigma_1^2 \quad \}$$

where superscripts indicate the dimension of the simplex, subscripts are labels, and simplices are defined as follows:

$$\sigma_1^0 = [x_1], \quad \sigma_2^0 = [x_2], \quad \sigma_3^0 = [x_3], \quad ...$$
$$\sigma_1^1 = [x_1, x_2], \quad \sigma_2^1 = [x_2, x_3], \quad \sigma_3^1 = [x_3, x_1], \quad \sigma_4^1 = [x_3, x_5],$$
$$\sigma_5^1 = [x_5, x_6], \quad \sigma_6^1 = [x_6, x_7], \quad \sigma_7^1 = [x_7, x_8], \quad \sigma_8^1 = [x_8, x_6]$$
$$\sigma_1^2 = [x_1, x_2, x_3]$$

One can verify that this set has the properties required to be a simplicial complex. Intuitively we can see that there is one "hole" (the 2D kind, the 1st Betti number) in this complex since the edges on points x_6, x_7 and x_8 make a loop. We can also see that despite the similar loop of points on x_1, x_2 and x_3 there is not a second hole here because the full triangle is filled in/there is a 2-simplex on those points as well as the edges. We need a way to differentiate these two loops to ensure one counts as a hole while the other does not.

We denote the "k^{th}-chain group" of a simplicial complex S as C_k. This is the free abelian group (see appendix 4.1) with the collection of k-simplices from S as a basis. The inverses are then the simplices with opposite orientation to the original set. A word in C_k is a k-chain, a formal sum of a finite number of oriented k-simplices. We generally denote a k-chain as

$$c_k = \sum_i a_i \sigma_i^k$$

so we write the k^{th} chain group of a simplicial complex S as

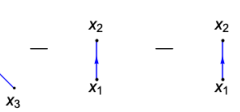
$$C_k = \{c_k = \sum_i a_i \sigma_i^k \mid \sigma_i^k \in S\} \tag{1.1}$$

We consider $a_i \in \mathbb{Z}$ but this definition can be extended to any abelian group like the reals.

As an example consider the group C_1 from our example complex $\tilde{S}$. The basis for this group is $\{\sigma_1^1, \sigma_2^1, \sigma_3^1, \sigma_4^1, \sigma_5^1, \sigma_6^1, \sigma_7^1, \sigma_8^1\}$. Thus any 1-chain, a formal sum of "edges" of the complex is in C_1,

we write for example:

$$\begin{aligned}\sigma_2^1 - 2\sigma_1^1 &= [x_2, x_3] - 2[x_1, x_2] \\ &= [x_2, x_3] - [x_1, x_2] - [x_1, x_2] \\ &= [x_2, x_3] + [x_2, x_1] + [x_2, x_1] \in C_1\end{aligned}$$

“ = ” 

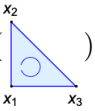
Arbitrary collections of simplices can be described as words in the chain group, for example the loop that we have said intuitively forms a hole can be described as $\sigma_6^1 + \sigma_7^1 + \sigma_8^1$, and the loop around the triangle as $\sigma_1^1 + \sigma_2^1 + \sigma_3^1$. However they could also be described as $-\sigma_6^1 - \sigma_7^1 - \sigma_8^1$ and $-\sigma_1^1 - \sigma_2^1 - \sigma_3^1$. Because we will be discussing them later we will define these 1-chains as: $l_1 = \sigma_1^1 + \sigma_2^1 + \sigma_3^1$ and $l_2 = \sigma_6^1 + \sigma_7^1 + \sigma_8^1$.

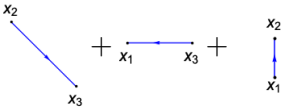
We then define the Boundary Operator, a function between chain groups. We say the boundary operator $\partial_k : C_k \to C_{k-1}$ maps a k-simplex onto a sum of the (k-1)-simplices in its boundary. It is the alternating sum of the simplex's faces. We write the operator as:

$$\partial_k(\sigma^k) = \sum_{i=0}^{k} (-1)^i [x_0, ..., \hat{x}_i, ..., x_k] \tag{1.2}$$

where $[x_0, ..., \hat{x}_i, ..., x_k]$ is the (k-1)-simplex/face resulting from deleting the vertex x_i. From our example take the boundary operator of the only 2-simplex we have:

$$\begin{aligned}\partial_2(\sigma_1^2) &= [x_2, x_3] - [x_1, x_3] + [x_1, x_2] \\ &= [x_2, x_3] + [x_3, x_1] + [x_1, x_2] \\ &= \sigma_2^1 + \sigma_3^1 + \sigma_1^1\end{aligned}$$

$\partial_2($ 

$)$ “ = ” 

Note that this sum of edges, $\sigma_2^1 + \sigma_3^1 + \sigma_1^1$ is the previously defined l_1; the “loop” surrounding the 2-simplex.

Note three important properties of the boundary operator:

1. $\partial_k(\sum_i a_i \sigma_i^k) = \sum_i a_i \partial_k(\sigma_i^k)$

2. $\partial_k(\partial_{k+1}(c_k)) = 0$

3. $\partial_0(\sigma^0) = 0$ for all zero-simplexes

Property 1 is linearity: the boundary operator is linear. Property 2 says that the boundary of any boundary is zero. This is important in the following definitions of the cycle and boundary groups and a proof can be seen in [3]. Property three is a convention, that the boundary operator of a 0-simplex or single point is 0.

As an example supporting property (2), we take the boundary of the boundary of the 2 simplex from before:

$$\begin{aligned}
\partial_1(\partial_2(\sigma_1^2)) &= \partial_1(l_1) = \partial_1(\sigma_2^1 + \sigma_3^1 + \sigma_1^1) \\
&= \partial_1([x_2, x_3] + [x_3, x_1] + [x_1, x_2]) \\
&= \partial_1([x_2, x_3]) + \partial_1([x_3, x_1]) + \partial_1([x_1, x_2]) \qquad \text{(by linearity)} \\
&= ([x_3] - [x_2]) + ([x_1] - [x_3]) + ([x_2] - [x_1]) \\
&= [x_3] - [x_3] + [x_1] - [x_1] + [x_2] - [x_2] \\
&= 0
\end{aligned}$$

Now we define the k$^{\text{th}}$ cycle group Z_k to be

$$Z_k = \{c_k \in C_k \mid \partial_k(c_k) = 0\}. \tag{1.3}$$

This is a subgroup of C_k; it is the group consisting of k-chains that map to zero under the boundary operator. In other words it is the kernel of ∂_k. Any k-chains in this set are called k-cycles.

We look at some examples from Z_1 of $\tilde{S}$. We showed previously that the chain $\sigma_2^1 + \sigma_3^1 + \sigma_1^1$ is in the kernel of ∂_1. Additionally the other "loop" mentioned before, $\sigma_6^1 + \sigma_7^1 + \sigma_8^1$ can easily be shown to have 0 boundary. Therefore both l_1 and l_2 are in Z_1. For 1-simplices this group consists essentially of the loops on the edges in the complex.

The cycle group has allowed us to identify the loops in our example complex $\tilde{S}$, but as mentioned there is a topological difference between the right and left loops, the left is filled in with a 2-simplex and thus not a "hole" while the left loop is empty. To distinguish these we introduce B_k, the k$^{\text{th}}$ boundary group. This is the group of k-chains which bound a k+1 chain, i.e. the image of ∂_{k+1}. Formally this becomes

$$B_k = \{\ c_k \in C_k \mid \partial_{k+1}(c_{k+1}) = c_k \text{ for some } c_{k+1} \in C_{k+1}\ \} \tag{1.4}$$

Note that by property (2) every boundary is a cycle so we can say that:

$$B_k \subseteq Z_k \subseteq C_k$$

For our two 1-cycles or loops (the left loop: $l_1 = \sigma_1^1 + \sigma_2^1 + \sigma_3^1$ and the right loop: $l_2 = \sigma_6^1 + \sigma_7^1 + \sigma_8^1$) we can then distinguish the two, as l_1 will be in B_k as the boundary of a 2-simplex but l_2 will not as there is no simplex in C_2, no triangle, which has this cycle as a boundary.

In summary, we have described 3 groups built from a general simplicial complex S for each dimension k:

C_k = the k$^{\text{th}}$ chain group: the free group with the k-simplices in S as the generating set. Elements are called k-chains.

Z_k = the k$^{\text{th}}$ cycle group: the subgroup of C_k consisting of only k-chains with zero boundary. Elements are called k-cycles.

B_k = the k$^{\text{th}}$ boundary group: the subgroup of Z_k consisting of only k-cycles that are also the boundary of some (k+1)-chain. Elements are boundaries or boundary-cycles.

We see from our example that the boundary group gives us the difference between cycles that bound holes and cycles that are "filled in".

The Homology Groups and Betti Numbers

We can now define H_k, the k$^{\text{th}}$ Homology group. It is the quotient group of Z_k and B_k, we write

$$H_k = Z_k / B_k. \tag{1.5}$$

Elements of the group are equivalence classes of k-cycles that do not bound any k+1 chain. Two k-cycles $z_k^1, z_k^2 \in Z_k$ are in the same equivalence class if $z_k^1 - z_k^2 \in B_k$. We denote the equivalence class that contains z_k as $[z_k]$.

This implies that any cycle that is itself in the boundary group is in the equivalence class with the 0 element. Therefore our filled in loop l_1 in a way "vanishes" in the homology group: $[0] = [l_1]$. However, l_2 does not as it does not belong to B_k. So l_1 and l_2 are fundamentally different in the Homology Group. We briefly noted earlier that both l_2 and $-l_2$ can equivalently describe the right loop. We see then that $l_2 - l_2 = 0 \in B_k$ so both are in the same equivalence class that corresponds to a single loop around the hole. $2l_2$ will be in a different equivalence class and for our simple

example we can show that we will have:

$$H_1 = \{[0]\} \cup \{\ [nl_2] \mid n \in \mathbb{Z}^+\}$$

We also now formally define the k$^{\text{th}}$ Betti number as

$$\beta_k = \text{rank}(H_k) \tag{1.6}$$

This gives us what we wanted from the beginning, a way to say how many k-dimensional holes an object has, at least for simplicial complexes. Then if a topological space can be triangulated by some simplicial complex, we can define the Betti number of the space as the Betti number of the triangulation, as the homology groups are the same.

The rank of H_1 for $\tilde{S}$ is 1, for our example complex $\tilde{S}$. This is the 1$^{\text{st}}$ Betti number of $\tilde{S}$. We could do similar analysis to show that:

$$H_0 = \{[0]\} \cup \{\ [n\sigma_1^0 + m\sigma_4^0] \mid n, m \in \mathbb{Z}^+\ \}$$

Every 0-simplex will be in Z_0 because of property 3 of the boundary operator, all points have zero boundary. Then points which are in the same connected component, or equivalently are the boundary of some 1-chain, will be put in the same equivalence class. Essentially every 0-simplex that is in the large connected component will be put in the same equivalence class as σ_1^0, and σ_4^0 will have its own equivalence class, and then any integer combination of these two classes will be their own class in the homology group, giving this Homology group rank= 2. This is what we would expect, as the complex has two disconnected components.

Thus this simplicial complex has $\beta_0 = 2$ and $\beta_1 = 1$. Any higher Betti numbers are zero, as the homology groups are trivial in $\mathbb{R}^2$ for $k \geq 3$

This general theory allows us to find the Betti numbers for arbitrary simplicial complexes by finding homology groups and identifying their ranks. However for any complex larger than our example, this computation would be extremely long and complicated. In the next section we show how the rank of homology groups can be identified computationally using known matrix algorithms.

$$
\begin{array}{c}
\begin{array}{cccccccc} \sigma_1^1 & \sigma_2^1 & \sigma_3^1 & \sigma_4^1 & \sigma_5^1 & \sigma_6^1 & \sigma_7^1 & \sigma_8^1 \end{array} \\
D_1 = \begin{bmatrix}
-1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1
\end{bmatrix} \begin{array}{c} \sigma_1^0 \\ \sigma_2^0 \\ \sigma_3^0 \\ \sigma_4^0 \\ \sigma_5^0 \\ \sigma_6^0 \\ \sigma_7^0 \\ \sigma_8^0 \end{array}
\end{array}
$$

$$
\begin{array}{c}
\begin{array}{cccccccc} \sigma_1^1 & \sigma_2^1 & \sigma_3^1 & \sigma_4^1 & \sigma_5^1 & \sigma_6^1 & \sigma_7^1 & \sigma_8^1 \end{array} \\
D_2^T = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \begin{array}{c} \sigma_1^2 \end{array}
\end{array}
$$

Figure 1.6: Boundary Matrices of $\tilde{S}$. Note the transpose on D_2.

1.2 Computing Topology

1.2.1 Computing Betti Numbers: Smith Normal Form

We introduced simplicial homology with the motivation that we can use algorithms to find homology groups and Betti numbers. Therefore we show that we can reduce the computation of homology groups and Betti numbers to linear algebra and matricies, which is easily implemented computationally.

We represent our boundary operators $\partial_k : C_k \to C_{k-1}$ as matricies D_k. Using the notation of [4] if our complex S has m_k k-simplices and m_{k-1} (k-1)-simplices in it, then D_k is an $m_{k-1} \times m_k$ matrix. Since the boundary of each k-simplex is a sum of (k-1) simpicies, the entry $[D_k]_{i,j}$ is ± 1 if the i^{th} (k-1)-simplex is in the boundary of the j^{th} k-simplex and 0 if it is not.

In Figure 1.6 we show the boundary matrices from our example $\tilde{S}$. As an example, we see that $\partial_1(\sigma_2^1) = \sigma_3^0 - \sigma_2^0$ so entry $(3, 2)$ is 1 and $(2, 2)$ is -1, the respective coefficients in the boundary sum.

The Boundary Matrices D_k and D_{k+1} completely characterize the homology group H_k, and computing the Smith Normal Form of these matrices tells us all we wish to know. The Smith Normal Form, originally described in [5], is found essentially by performing Gaussian Elimination, except we want to maintain integer entries of the matrix, so no division is allowed. We then reduce D_k to D'_k in the block matrix form below.

$$D'_k = \begin{bmatrix} B_k & 0 \\ 0 & 0 \end{bmatrix}$$

B_k is an $l_k \times l_k$ diagonal matrix with $l_k \leq m_k$ and non-zero diagonal entries. The rank of the cycle group Z_k is $m_k - l_k$ or the number of zero columns in the matrix D'_k. The rank of the Boundary group is the rank of matrix D_{k+1} or the number of non-zero columns in D'_{k+1} which is l_{k+1}. The rank of a quotient group is the rank of the original group minus the rank of the subgroup so we have

$$\beta_k = \text{rank}(H_k) = \text{rank}(Z_k) - \text{rank}(B_k) = m_k - l_k - l_{k+1}. \tag{1.7}$$

The diagonal entries of B_k, if greater than 1, give us the torsion coefficients of the homology group H_k. However in cases where we are not interested in torsion coefficients, we can simply perform Gaussian Elimination on D_k instead of computing the Smith Normal Form. This will give us the same number of non-zero diagonal elements. This is advantageous because requiring integer elements of the matrix often results in extremely large numbers during the calculation which can cause computational issues.

1.2.2 Simplicial Complexes from Finite Data

Since in most contexts we are only given a discrete set of points to approximate a topological space, we wish to construct a simplicial complex from the set of points which approximates the underlying space. There are many ways to do this, but we show two of the most straight-forward ways to construct the complex. Generally we have some parameter ϵ that represents the scale at which we are looking at our data. To be more precise we wish to approximate the ϵ-neighborhood of the original topological space with the ϵ-neighborhood of our points, as this takes into account the discrete nature of our point sample. Figure 1.7 illustrates generally what we wish to accomplish. For simplicity we restrict our discussion to $\mathbb{R}^n$ with the L_2 metric but the complexes can be defined using different metrics.

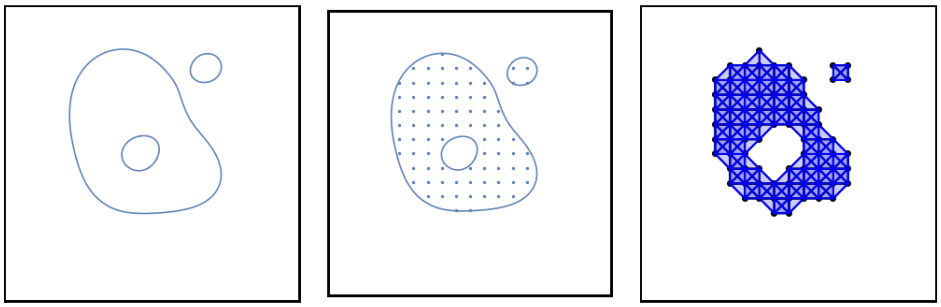


Figure 1.7: We have a sample of points from a topological space, and wish to construct a simplicial complex that approximates the space.

Cech Complex

All of our complexes will begin with all of the points in the data set which approximates our topological space as 0-simplexes. Then we will add higher-order simplices based on some closeness condition on sets of points.

Informally the Cech Complex grows n-dimensional balls around each simplex in Euclidean space $\mathbb{R}^n$ and when the balls overlap, the simplex composed of the points at the center of each sphere is added to the complex. A more formal description follows, see [6] for more details.

Given our data set of N points $D = \{x_i\}_{i=1}^N \subset \mathbb{R}^n$ and a parameter ϵ we construct a simplicial complex $C(\epsilon)$, called the Cech complex. Define $B_i(\epsilon)$ as the closed ball of radius ϵ centered at point x_i for $i = 1, 2, 3, ..., N$. For any set of $k+1 \leq N$ points $\{x_{(j)}\}_{j=1}^{k+1} \subset D$ the k-simplex $\sigma^k = [x_{(1)}, x_{(2)}, ..., x_{(k+1)}]$ is in $C(\epsilon)$ if and only if

$$\bigcap_{j=1}^{k+1} B_{(j)}(\epsilon) \neq \emptyset \tag{1.8}$$

Figure 1.8 Illustrates the process of creating the Cech Complex. This complex is desirable as the simplicial complex is guaranteed to have the same homology as the union of ϵ-balls used in its definition, equivalently the ϵ-neighborhood of the point set, due to the nerve theorem, see [4].

Vietoris-Rips complex

The Vietoris-Rips or Rips complex is very similar to the Cech Complex, but instead of looking for overlapping balls we simply measure the distance between each pair of points and add a simplex when all pairwise points are within a distance ϵ of each other.

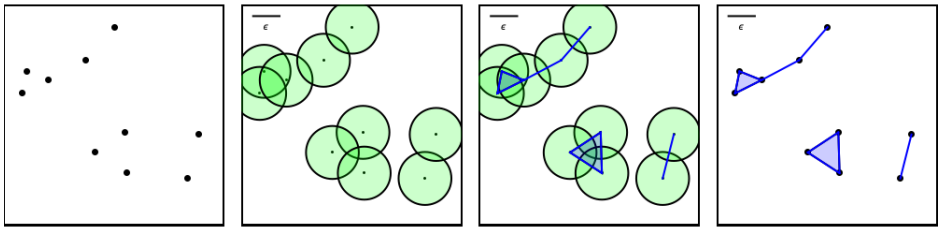


Figure 1.8: **Building a Cech Complex:** The first frame shows the original data set. The second shows the balls of radius ϵ centered at each point. Then for each set of balls with a non-empty intersection we add a simplex in the third frame. The final frame shows the resulting simplicial complex.

More formally, given a data set $D = \{x_i\}_{i=1}^N \subset \mathbb{R}^n$ and a parameter ϵ we construct the Rips complex $R(\epsilon)$. Instead of considering the intersection of balls we simply say that for any set of $k+1 \leq N$ points $\{x_{(j)}\}_{j=1}^{k+1} \subset D$ the k-simplex $\sigma^k = [x_{(1)}, x_{(2)}, ..., x_{(k+1)}]$ is in $R(\epsilon)$ if and only if

$$d(x_{(i)}, x_{(j)}) \leq \epsilon \quad \text{for all } i, j \in \{1, 2, 3, ..., k+1\} \tag{1.9}$$

This requirement is only subtly different from the Cech Complex. When comparing the Cech and Vietoris-Rips complexes we often consider the ϵ of the Cech complex to be half that of the Vietoris-Rips Complex, because two points which connect as a 1-simplex in the Cech Complex at ϵ_1 will connect at $2\epsilon_1$ in the Vietoris-Rips Complex on the same set of points.

In fact in [7] it has been shown that we can bound the Cech complex using the following lemma.

Lemma: For any $\epsilon > 0$ we have that $R(\epsilon) \subseteq C(\epsilon\sqrt{2}) \subseteq R(\epsilon\sqrt{2})$

It is however possible to have a set of points that are pair-wise within a distance ϵ of each other despite having no intersection of the balls of radius $\epsilon/2$ centered at each one. An example is shown in Figure 1.9. The Cech complex has better guarantees on accurate topology, see [8], however the Rips complex is much less computationally expensive and still gives good results.

Alpha Complex

The alpha complex has many computational advantages over the Cech and Rips complex especially for point sets in $\mathbb{R}^2$ and $\mathbb{R}^3$. See [9] for a summary and [10] for algorithms in the 3D case. Essentially

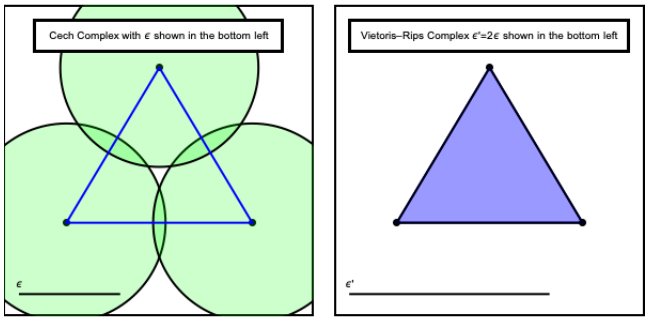


Figure 1.9: Cech complex vs Vietoris-Rips complex on the same 3 points. Note the 2-simplex (triangle) filled in on the left but not the right, and the lack of a 3-way intersetion of balls on the left.

the alpha complex uses the same concept of ϵ-balls as the Cech complex, but takes their union with the Voronoi cells of the point at their center. The Voronoi cell of point x_j is the region of space in which the distance to the point x_j is less than the distance to any other point in the data set. If V_j is the Voronoi cell of point x_j then the definition is nearly the same as the Cech complex but condition (1.8) for a simplex to be in the alpha complex $A(\epsilon)$ becomes

$$\bigcap_{j=1}^{k+1} B_{(j)}(\epsilon) \cup V_j \neq \emptyset \tag{1.10}$$

Assuming general position of the data points, this guarantees that the complex will never have a simplex of dimension greater than the space in which the points are embedded, and will have fewer k-simplices than the Cech complex even for k less than the dimension. In fact $\mathrm{A}(\epsilon) \subseteq C(\epsilon)$, but still gives the topology of the ϵ-neighborhood of the points.

There are many other ways to construct simplicial complexes from finite data sets, see [11] for a description of the witness complex, or [12] for a description of cubical complexes which use "elementary cubes" as the basis for homology rather than simplices. Some more complicated constructions down-sample the data by not using all of the data points in the original 0-simplex set, which can be very advantageous for computation. These three examples, the Cech, Rips, and alpha complexes however give a good intuition as to how complexes are constructed using some closeness parameter, generally called ϵ. Considering how the simpical complex changes as a function of this parameter leads us to the notion of persistence discussed in the following section.

1.3 Persistent Homology

1.3.1 Filtrations and Persistence

Now for any type of simplicial complex we define with a distance parameter ϵ, we can imagine "growing" the complex as we let ϵ range from 0 to infinity. We begin to think of this parameter as a sort of time variable. Observing the way the homology of the complexes change over time can give us an idea of which holes are important and which can be considered noise. Persistent homology is developed in [13] and [4].

Given a simplicial complex K, which for our purposes we assume to be finite, we assign each simplex a "birth time". This can be thought of as a function from K into the reals $b : K \to \mathbb{R}$ where for a simplex σ in the complex we have a birth-time $b(\sigma)$. We require that if σ_1 is a face of σ_2 then $b(\sigma_1) \leq b(\sigma_2)$. This requirement ensures that the set

$$K(\epsilon) = \{\sigma \in K | b(\sigma) \leq \epsilon\} \tag{1.11}$$

is itself a simplicial complex. This can also be thought of as the preimage of the interval $(-\infty, \epsilon]$, in other words $K(\epsilon) = b^{-1}((-\infty, \epsilon])$. Because of our assumption that K is finite we know there is a minimum and a maximum birthtime. Therefore for any ϵ which is lower than the minimum birth-time we know that $K(\epsilon)$ is the empty set and for any ϵ which is greater than or equal to the maximum birth-time that $K(\epsilon)$ is K. With a finite, discrete set of birth-times we also know that there are a finite and discrete number of unique complexes, and they are all sub-complexes of K. We can then index them so that we have the following.

$$\emptyset = K_0 \subseteq K_1 \subseteq K_2 \subseteq ... \subseteq K_{n-1} \subseteq K_n = K \tag{1.12}$$

This set of simplicial complexes $\{K_i\}_{i=0}^n$ each a sub-complex of the next, is called a filtration. As discussed before we can calculate the Betti number of each one of these simplicial complexes individually. However, if we consider the way in which the complexes grow, we can glean more information than simply just how many holes exist at each step. Holes are born and die at different times/ ϵ values, and we can take into account those that persist for longer intervals in the filtration.

Each complex K_i will have homology groups $H_p(K_i)$ which contain equivalence classes corresponding to the p-dimensional holes of each complex. For $0 \leq i \leq j \leq n$ we have that $K_i \subseteq K_j$ and thus we also have an inclusion map between the spaces. This gives us the induced homomorphisms

$f_p^{i,j} : H_p(K_i) \to H_p(K_j)$ for each dimension p. These homomorphisms map the equivalence classes representing holes in K_i to equivalence classes representing holes in K_j so that we can identify holes that exist in both subcomplexes. We define the image of these homomorphisms to be the persistent homology groups, $H_p^{i,j} = \text{im} f_p^{i,j}$ and the ranks of these groups are the persistent Betti numbers,

$$\beta_p^{i,j} = \text{rank}(H_p^{i,j}). \tag{1.13}$$

The persistent Betti numbers are distinct but related to the standard Betti numbers. $\beta_p(K_i)$ and $\beta_p(K_j)$ are the number of p-dimensional holes in the complexes K_i and K_j respectively. $\beta_p^{i,j}$ Is the number of p-dimensional holes that exist in both K_i and K_j. Therefore the persistent Betti number must be less than or equal to both of the standard Betti numbers.

The complexes we have discussed, Rips, Alpha, Cech, as well any other complex defined similarly using a distance parameter effectively defines the birth time function b. The birth time for any simplex in one of these complexes is the minimum epsilon at which the complex contains the simplex. For example, the birth-time of an edge in the Rips complex is the distance between the two points.

The full complex K for the Rips complex is the complete complex on the set of points that we begin from. If our point set has m points, $K = \lim_{\epsilon \to \infty} R(\epsilon)$ is an m-simplex with every one of its faces, and their faces, and so on. Once ϵ is greater than the diameter of the point-set, every edge will be connected and therefore every possible triangle, tetrahedron, and so on. The same is true for the Cech complex, the only difference is in intermediate filtration values.

The full complex K for the alpha-complex is the Delaunay Triangulation of the point set. This has far fewer simplexes than the full m-simplex of the m points, and assuming general position of the points will never give a simplex of dimension higher than that of the embedding space. This is highly advantageous for computations, as persistent homology algorithms scale poorly with number of simplexes.

1.3.2 Visualizing Persistence

Being able to visualize how the homology of a filtration changes is important, especially for non-trivial examples with complex topology. To illustrate some of the ways to visualize the persistent homology of a simplicial complex built from a point set we introduce the example in Figure 1.10.

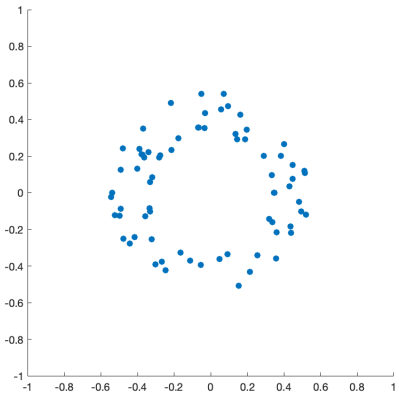


Figure 1.10: We have a sample of 75 points from annulus and wish to analyze the topology

Intuitively we can see these points have been sampled from an annulus. We would expect an annulus to have 1 component and 1 hole. Figure 1.11 shows the balls of various radii growing around each point.

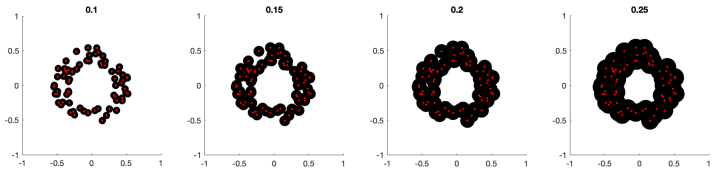


Figure 1.11: We grow balls of epsilon around each point. The collection of these balls will have the same homology as the Cech complex, approximated closely by the Rips Complex.

We calculate a filtration using the Rips Complex on these points. Thus when our filtration parameter is zero, we will have a large number of disconnected components (in our case 75). Then as the parameter grows, more and more components will become connected, until eventually we expect a single connected component. As for the number of holes, initially with distinct points we expect to have no holes at all. Then as the components connect we expect some holes to form but then die quickly due to the noisiness of the data. Then at some point we would expect all holes to fill in except the large one in the middle. This large hole once it is born should last a long time compared to the other smaller, noisier holes at the beginning of the filtration. We use

Persistence Diagrams and Barcodes to visualize the exact calculated persistent homology to connect our intuition with quantitative results.

Barcodes

Barcodes, see [7], are perhaps the most intuitive way to visualize persistence. Throughout the filtration, topological features will be born and die, and a barcode plots an interval for each topological feature, beginning at the birth time of the feature and dying at the death time. Figure 1.12 shows the 0 and 1 dimensional Barcodes of our annulus data. These were created using Javaplex, see [14].

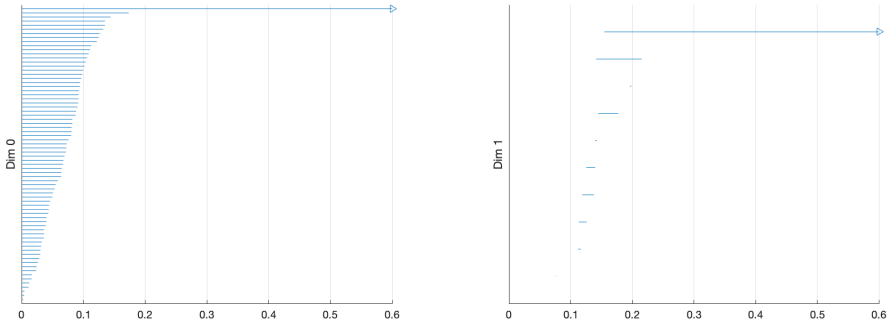


Figure 1.12: Barcodes from the Rips Complex constructed from the annulus-sampled point set

In the 0-dimensional barcode, there are 75 lines, each beginning at 0. These each correspond to one of the points in our original point-set. then as epsilon grows, edges form and components begin to combine, so some of the intervals end as they are absorbed into other components. By about $\epsilon = 0.2$ all of the components have combined into one single large component. This final large component lasts the entirety of the filtration. This supports our intuition as well as our hypothesis that the points come from a single annulus.

In the 1-dimensional barcode, there are initially no holes. Between 0.1 and 0.2 some holes form but generally die shortly after forming. These short intervals correspond to our supposed "noisy" holes. Then after 0.2 only one of the holes lasts, but it lasts the rest of the filtration. (At least past $\epsilon = 0.6$. It eventually also closes.) This corresponds to the large hole in the middle of the annulus. Its long interval quantitatively supports this hole belongs to the "true" underlying topological space.

Given the set of barcode intervals, we can easily compute Betti numbers and persistent Betti

numbers. The pth Betti number of the complex $\beta_p(K(\epsilon))$ is the number of intervals that intersect with a vertical line at ϵ. The persistent Betti number $\beta_p^{\lambda,\epsilon}(K)$ is the number of intervals that intersect with both a vertical line at λ as well as a vertical line at ϵ. These topological features "persist" throughout the interval. [7]

Persistence Diagrams

Persistence Diagrams, see [15], encode the same information as the Barcodes but slightly differently. Each topological feature is represented as a point in the plane, with x coordinate as the birth time and y coordinate as the death time. Features that persist beyond the end of the filtration are represented as triangles along the final filtration value. Figure 1.13 shows the Persistence diagrams for our example data.

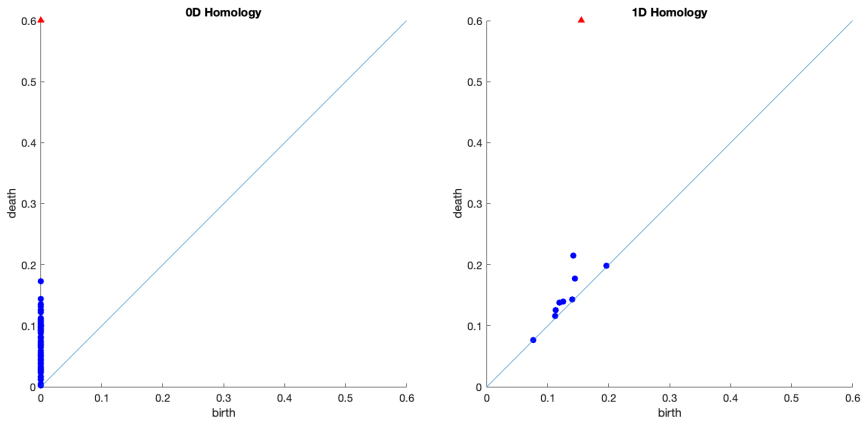


Figure 1.13: Persistence Diagrams from the Rips Complex constructed from the annulus-sampled point set

Of course a feature's death time must always come after its birth, so all points will be above the line $y = x$. The length of an interval in the barcode corresponds to the closeness of a point to this line. We can see that in both the 0 and 1 dimensional persistence diagrams there are many features that are close to this line, and a single feature that exists far from the diagonal, corresponding to the singular connected component and hole in the underlying annulus.

Betti numbers and persistent Betti numbers can also be calculated from a persistence diagram. The persistent Betti number $\beta_p^{\lambda,\epsilon}(K)$ is the number of points which exist above and to the left of a point (λ, ϵ). This naturally gives rise to a "persistent rank function" which is a function of

two variables which returns the persistent Betti number. This rank function has proved useful for defining statistics to compare filtrations [16]. The classical Betti number is the rank function exactly on the diagonal.

1.4 Fractals and Symmetry

1.4.1 Scaling Properties and Homology of Fractals

Persistent homology can provide powerful computational results for finding the topological features of simple spaces approximated by point sets. But many topological spaces have features that are not so easily deduced from data. In particular, fractal sets will have topological features that are much less intuitive. Consider the Sierpinski gasket in Figure 1.14, which will be discussed in detail later.

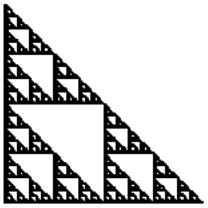


Figure 1.14: The Sierpinski Gasket

What are the Betti numbers of the Sierpinski gasket? This question is much more difficult to answer than for an annulus. It can be shown that the gasket is indeed a single connected component, so $\beta_0 = 1$. However, the number of holes the object has is infinite, no matter how close you look, there will be another hole by definition. The 1st Betti number for this fractal is therefore not well-defined.

In order to gain useful information about fractal sets from finite data, we draw inspiration from the box-counting or Minkowski dimension. The box-counting dimension is defined as

$$\dim(S) = \lim_{\epsilon \to 0} \frac{\log(N(\epsilon))}{\log(1/\epsilon)} \tag{1.14}$$

where $N(\epsilon)$ is the number of boxes of side-length ϵ required to cover the set. In the case where this limit does not exist we replace the limit with the limsup. This is derived from the assumption

that there is power-law scaling $N(1/n) \propto n^d$ for manifolds, see [17]. The growth rate d is always an integer for manifolds, but with fractal sets we often find non-integer dimensions.

In [18], Robins defines the "disconnectedness index" or γ_0 analogously to the box-counting dimension as

$$\gamma_0(S) = \lim_{\epsilon \to 0} \frac{\log(C(\epsilon))}{\log(1/\epsilon)} = \lim_{\epsilon \to 0} \frac{\log(\beta_0^{0,\epsilon}(S))}{\log(1/\epsilon)} \tag{1.15}$$

with the function $C(\epsilon)$ being the number of components that the ϵ neighborhood of a fractal has. Persistent Betti $\beta_0^{0,\epsilon}(S)$ is the number of 0-dimensional holes that exist in the fractal and persist into the ϵ neighborhood of the set S. Our definition of persistent betti numbers assumed a simplicial complex and filtration, but they can be defined over ϵ neighborhoods just as we defined them over a filtration.

This index can be similarly defined for any arbitrary dimension, which only slightly changes the formula. For $\beta_i^{0,\epsilon}(S)$ being the number of persistent i-dimensional holes in the ϵ-neighborhood of a set S, we define

$$\gamma_i(S) = \lim_{\epsilon \to 0} \frac{\log(\beta_i^{0,\epsilon}(S))}{\log(1/\epsilon)} \tag{1.16}$$

to be the ith persistent Betti growth rate. Robins calls the growth rate for $i = 1$ the "loopiness" index, as it measures how the loops or holes scale within the fractal. The index for $i = 2$ she calls the "holiness" index as it measures how the number of bubble-like voids a fractal has scales with ϵ, reminiscent of holey swiss-cheese.[4]

It is important that the indices are defined using persistent Betti numbers, because it is possible that short-lived holes form and disappear as ϵ grows. We only want to count holes in the ϵ neighborhoods of the fractals that correspond to holes in the fractal itself.

This is only analytically computable in idealized situations. In practice, as with the box counting dimension we wish to be able to calculate approximate growth rates from approximations of the fractal. This is where persistent homology becomes useful. Using a finite point-set which approximates our fractal we create a filtration such as the Rips or Alpha complex which is well suited for approximating the homology of a set's ϵ neighborhood. We calculate a value ρ, our cutoff value, which is the minimum value that we believe our filtration reasonably approximates the fractal's ϵ neighborhood. ρ depends on the noisiness and density of our point-set. We can then compute the

approximate growth-rate as

$$\gamma_i(S) \approx \lim_{\epsilon \to \rho} \frac{\log(\beta_i^{\rho,\epsilon}(S))}{\log(1/\epsilon)} \tag{1.17}$$

using the persistent betti numbers of the filtration. This is the method we use to analyze our fractals in the following sections, see [4] for examples in 2D.

1.4.2 The Sierpinski gasket and its relatives

The Sierpinski gasket is a classical example of a 2D fractal, often used to introduce the idea of self-similarity [19]. The intuitive way to imagine forming the Sierpinski gasket is to begin with a unit square, divide it into 4 equal quarters, remove the top right quadrant, shrink the resulting shape uniformly by a factor of two, and replace the 3 remaining quarters with the smaller version of the whole. Repeating the shrinking and replacing over and over again limits to the Sierpinski gasket. It is easy to see the self-similar structure of this set: one third of the overall shape is exactly similar to the entirety of the fractal.

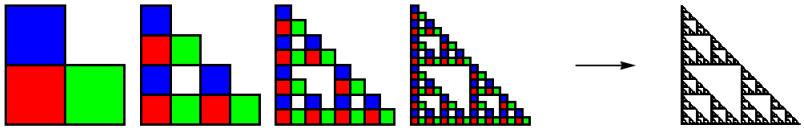


Figure 1.15: Creating the Sierpinski gasket iteratively

The Sierpinski relatives are a class of fractals created similarly to the Sierpinski gasket, however at the "shrinking" step we introduce an additional rotation/flip for each of the 3 non-empty quadrants. Figure 1.16 shows some examples. Each rotation/flip is an element of the symmetry group of the square which has 8 elements. The symmetry group of the square is discussed in appendix 4.2. With 3 squares there are $8^3 = 512$ ways to choose the symmetries, but excluding identical or reflected fractals there are 232 unique fractals.

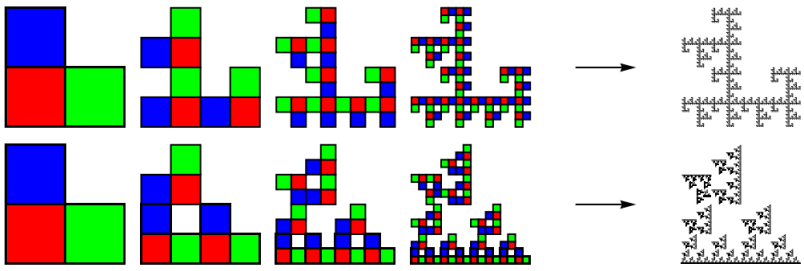


Figure 1.16: Some examples of forming Sierpinski relatives

A second way to describe the Sierpinski gasket and its relatives is with an Iterated Function System or IFS. An IFS, in general, is a finite set of contraction mappings on a complete metric space. The contraction mapping theorem shows that in a complete metric space an IFS has a unique, non-empty fixed set S, see [19]. This gives us a more mathematical way of describing the

Sierpinski relatives. For the original gasket we have the following 3 contraction mappings from $\mathbb{R}^2$ to $\mathbb{R}^2$:

$$f_1(\vec{x}) = \frac{1}{2}\vec{x}$$
$$f_2(\vec{x}) = \frac{1}{2}\vec{x} + \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}$$
$$f_3(\vec{x}) = \frac{1}{2}\vec{x} + \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}$$

The Serpinski gasket is then the unique fixed set $S \subset \mathbb{R}^2$ such that

$$S = \cup_{i=1}^{3} f_i(S) = f_1(S) \cup f_2(S) \cup f_3(S) \tag{1.18}$$

Each Serpinski Relative can be described by equation (1.18) as the attractor of an Iterated Function System consisting of 3 affine contraction mappings but with different functions f_i.

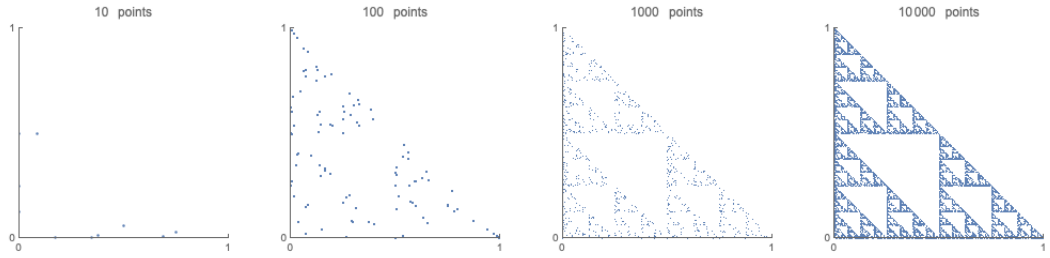


Figure 1.17: Approximating the Sierpinski gasket with the Chaos Game

The IFS description of fractals gives us a powerful way to computationally approximate the attracting set. Often called the "chaos game" we begin by choosing a random point in the unit square x_0. We then iterate this point to find $x_{k+1} = f_i(x_k)$ where f_i is a randomly chosen contraction mapping from the Function System. For our purposes we choose a random function uniformly and iid, but it is common to assign different probabilities for each mapping. If the initial point x_0 is in the fixed set of the IFS, the iterates will stay within and fill in the fixed set. This creates a dynamical map with the fixed set of the IFS as an attractor, see [20]. Implementing this computationally, we find the orbit of a randomly chosen point and by ignoring the first few iterates we obtain a

point-cloud approximation of the attractor. The more iterates we take, the more densely we fill the attractor, allowing us to balance precision and computational efficiency when we wish to analyze the properties of these sets.

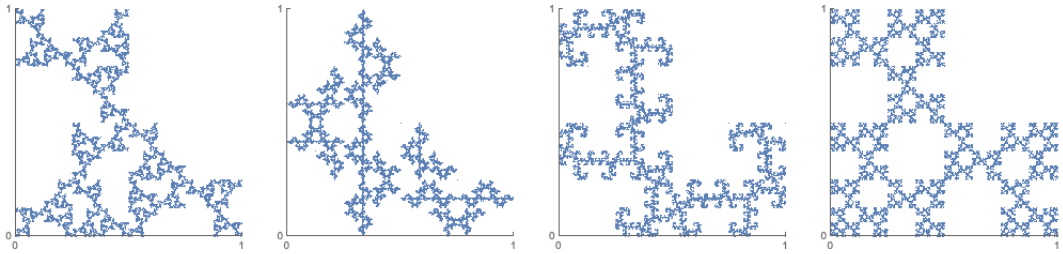


Figure 1.18: Some Sierpinski relatives created with 20000 points and the chaos game

Chapter 2

3D Sierpinski relatives and Analysis

2.1 3D extension of the Sierpinski relatives

We wish to create a set of 3D fractals analogous to the Sierpinski relatives. Instead of beginning with the 2D unit square we generate our fractals by subdividing the unit cube into 8 smaller cubes. It is not obvious which of the 8 cubes, and how many, we should remove in 3 dimensions so to be as general as possible we allow our set of fractals to include any choice of the 8 sub-cubes to be removed. After we remove n cubes with $0 < n < 8$, the analogy continues. Just as in the 2D case we shrink our new shape by a factor of 2, and replicate it in each of the $8 - n$ remaining subcubes. Each of these $8 - n$ cubes will be rotated, flipped, and/or inverted in some way at each iteration. Repeating this shrinking, rotating, and replicating process gives us a class of fractals which intuitively extend the Sierpinski relatives into 3D.

The choice of flipping/rotating/inverting for each sub-cube is associated with one of the elements of the symmetry group of the cube, just as the Sierpinski relatives were associated with elements of the symmetry group of the square. The cube is "dual" to the octahedron, so the symmetry group of the cube is referred to as the octahedral symmetry group, which has 48 elements including symmetries that are not oreintation-preserving. See appendix 4.2.

We introduce a reference cube with labeled verticies for convenience in Figure 2.1. Each sub-cube/eighth will be referred to by the numbers on the vertex incident to the cube. We assume point 3 is at the origin, point 1 is at (0,0,1), and point 4 is at (1,0,0).

As an example of a fractal from this set, Figure 2.2 shows the first few iterations and attracting

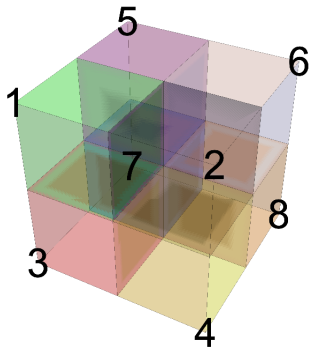


Figure 2.1: Reference cube, 3 is the origin, 1 is on the z axis, 4 is on the x axis

set of a particular fractal. We remove cubes 2, 5, 6, and 8. We color cube 1 green, 3 red, 4 yellow, and 7 blue. We shrink this shape down, and replicate it in each sub square. In analogy to the Serpinski gasket we choose our symmetries as the identity. This gives us a 3D Serpinski gasket with 2D gaskets on each plane.

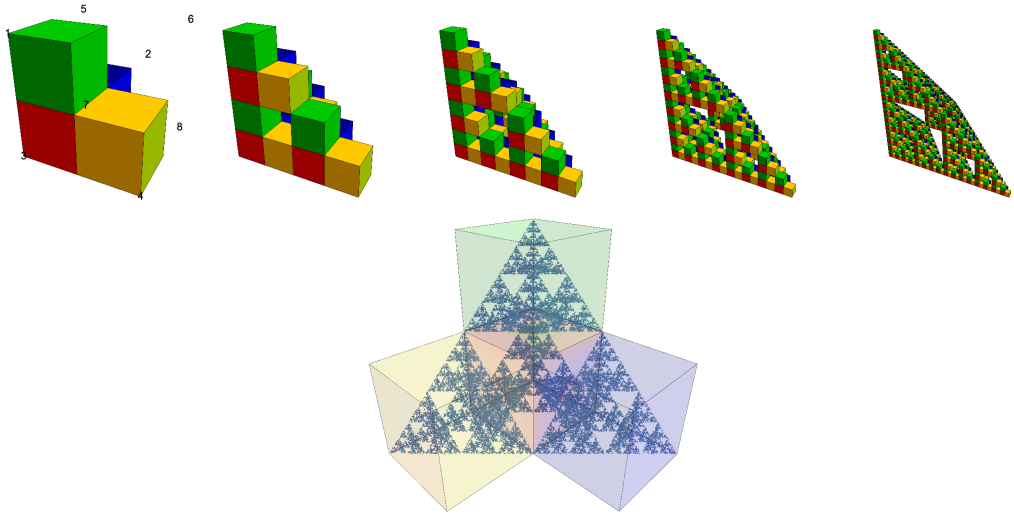


Figure 2.2: 3D fractal $(0, -1, 0, 0, -1, -1, 0, -1)$, iterations and attractor

We would like to know how many unique fractals are in this set. There are 8 initial cubes to choose from. Each one can be either chosen as empty, or if non-empty can be associated with any of the 48 elements of the octahedral symmetry group. There are thus 49 choices for each of the 8 cubes so we have $49^8 - 1 = 33,232,930,569,600$ or over 33 trillion possible attractors. 1 must be subtracted to avoid counting the "empty" fractal where we choose all cubes to be removed. This leads us to a notation to specify any attractor in our set. We specify a fractal from our set

using an 8-tuple with elements from the set $\{-1, 0, 1, 2, 3, ..., 47\}$. If the 1st number in the tuple is -1, then sub-cube 1 is removed, otherwise the number refers to which symmetry group element sub-cube 1 is associated with for iteration. Thus each number in the tuple indicates what we do to each corresponding sub-cube, whether it be to remove, rotate, flip, and/or invert. Our previous example would thus be indicated as fractal $(0, -1, 0, 0, -1, -1, 0, -1)$ as d_0 is the identity element of the octahedral symmetry group.

Of course, similar to the 2D case there will be many identical attractors within this set, and many that only differ by a rotation. Since the attractor itself can be rotated in any of the 48 ways associated with the cube and another distinct combination of symmetries can be chosen to create this rotated attractor we can immediately divide this number by 48, so there are less than $(49^8 - 1)/48 = 692, 352, 720, 200$ unique attractors, up to symmetry.

There are other repeat attractors in our set, even eliminating these rotational duplicates. A full analysis of how many there are is out of the scope of this thesis, but would be an interesting problem likely involving group theory and combinatorics. There are also attractors that are not fractal, for example any choice where no sub-cube is chosen to be removed gives an attractor that is the entire cube. Any choice where all but one sub-cube is removed gives a single fixed point. Even eliminating repeats and "boring" attractors leaves us with an overwhelming number of possibilities, and thus a full analysis of the topology and properties of the elements our new set of 3D fractals is not currently feasible. Thus we perform whatever general analysis we can and then search for interesting fractals among the sea of possibilities which will give us an idea of the types of interesting shapes we can find.

2.1.1 Fractal Dimension

We investigate the box-counting dimension of the Sierpinski relatives as defined in equation (1.14). Thinking of the way we iterated smaller and smaller squares to create the 2D Sierpinski relatives gives us an easy manner of calculating their box-counting Dimensions. At the first step we cover our fractal with 3 squares, each of side length 1/2. Thus $N(1/2) = 3$. At each consecutive step, the three previous squares become 3 more squares each having side length 1/2 of the previous step's length. Thus we can say at step k that $N(1/2^k) = 3^k$. Epsilon goes to zero as k goes to infinity so we calculate

$$\dim(S) = \lim_{\epsilon \to 0} \frac{\log(N(\epsilon))}{\log(1/\epsilon)} = \lim_{k \to \infty} \frac{\log(3^k)}{\log(2^k)} = \frac{\log(3)}{\log(2)} \approx 1.58496.$$

The analogy to 3D can be used in the same way for the 3D Sierpinski relatives. Say we remove $8 - m$ sub-cubes out of 8, leaving us with m sub-cubes. Then at step one, the side length of each cube is 1/2, and we need m cubes to cover the attractor. At each next step the side length of the cube is halved as before, and each of the cubes will result in m more than the previous step. Thus for m sub-cubes at step k we have that $N(1/2^k) = m^k$. Then we have

$$\dim(S) = \lim_{\epsilon \to 0} \frac{\log(N(\epsilon))}{\log(1/\epsilon)} = \lim_{k \to \infty} \frac{\log(m^k)}{\log(2^k)} = \frac{\log(m)}{\log(2)} \tag{2.1}$$

The table in Figure 2.3 shows the Box counting dimension for $m = 1, 2, ...8$. There are interesting implications of these numbers. We discuss each m along-side examples in the following section.

2.1.2 Subsets

It is interesting to examine each subset of our set of attractors formed by keeping m sub-cubes separately. We discuss briefly the various cases of m alongside a selection of 4 attractors from each subset. The color of each point in the figures is determined by its physical location in space to help visualize in 3D; the RBG color value is exactly the x, y, z coordinate in space.

The following table summarizes the different subsets. The "number" column lists how many possible attractors there are in this subset. Keep in mind this is out of the ≈ 33 trillion fractals. This is calculated as $\binom{8}{m}48^m$, as we need to choose m cubes to keep, and then a symmetry operation for each of the m cubes. Dividing these numbers by 48 to eliminate symmetrically-equivalent attractors gives a tighter bound on the number of truly unique attractors, but the percentages will remain the same and thus give a better idea of the magnitude of these subsets.

As previously mentioned, the attractor for any IFS with a single contraction mapping will be a fixed point, so it is clear that the dimension when $m = 1$ is zero. See figure 2.4. There are 384 of these fixed point attractors. This is about $1.155 * 10^{-9}$ percent of the 33 trillion total attractors, a very small amount of the total. In fact nearly all of the attractors will be in the also trivial $m = 8$ case discussed later.

For $m = 2$ the dimension is exactly 1, see figure 2.5. We do find attractors that are simply lines in 3d space, but interestingly there are a wide variety of shapes and patterns that emerge in

m	$\dim_m(S)$	number	$\approx$ percentage
1	0	384	$1.1554 * 10^{-9}$
2	1	64,512	$1.9412 * 10^{-7}$
3	1.58496	6,193,152	0.000018635
4	2	371,589,120	0.0011181
5	2.32193	14,269,022,208	0.042936
6	2.58496	342,456,532,992	1.0304
7	2.80735	4,696,546,738,176	14.132
8	3	28,179,280,429,056	84.793

Figure 2.3: box-counting dimensions for a 3D Sierpinski Relative Fractal from m sub-cubes, and how many of the fractals are in each subset

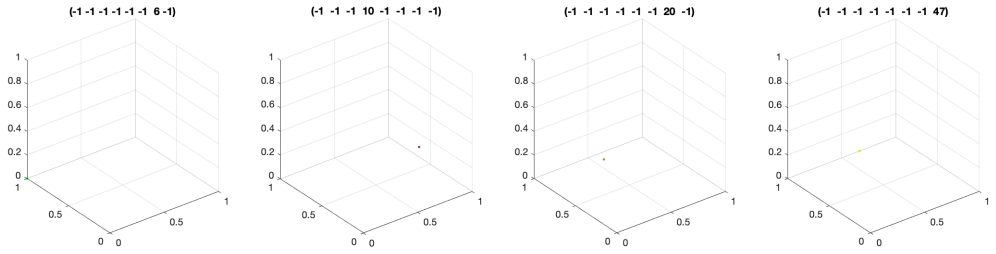


Figure 2.4: 3D Sierpinski relatives with 1 cube

this case, despite the integer fractal dimension. These fractals are great examples to show us that integer box-counting dimension does not necessarily imply a non-fractal set.

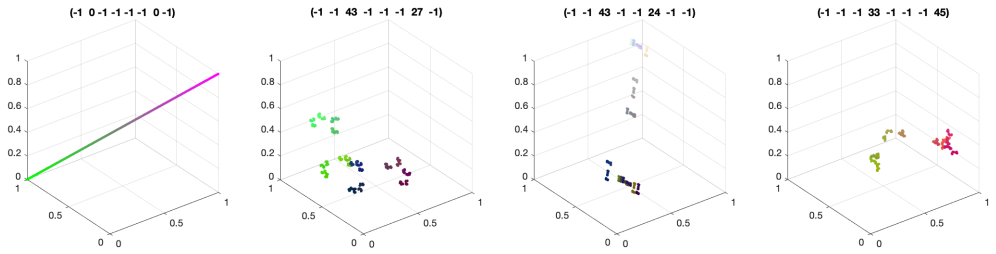


Figure 2.5: 3D Sierpinski relatives with 2 cubes

Figure 2.6 shows examples when $m = 3$. The dimension is exactly the same as the 2D relatives, despite being embedded in 3D space. In fact we can find fractals in this set that are simply the

2D relatives on any of the faces of the unit cube. Yet despite being less than 2 dimensional, these fractals also have a wide range of shapes and patterns.

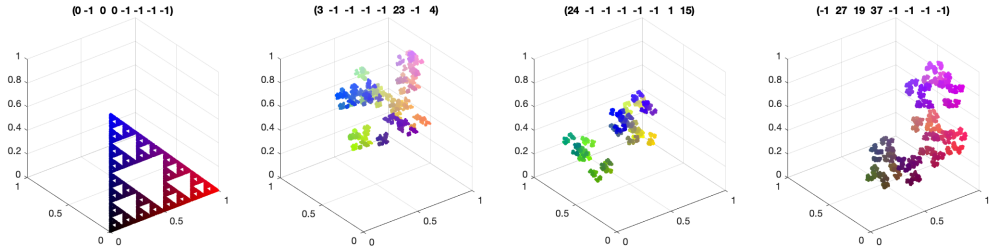


Figure 2.6: 3D Sierpinski relatives with 3 cubes

For $m = 4$ the dimension is exactly 2. As with $m = 2$ this case shows us that there can be complex fractal shapes that have integer box-counting dimension, see figure 2.7.

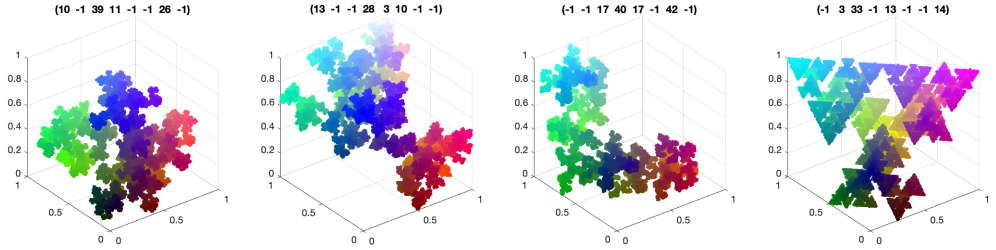


Figure 2.7: 3D Sierpinski relatives with 4 cubes

For $m = 5, 6, 7$ with examples shown in figure 2.8 the dimensions are increasing between 2 and 3. As could be expected, the fractals become more and more "dense" in visualizations as the dimension increases, closer to filling a region of 3D space. When $m = 7$, only one cube is removed of the 8 and we find interesting sponge-like structures.

For $m = 8$ nothing is removed so the attractor becomes the entire cube, no matter which symmetries are assigned to each sub-cube. Plots of these attractors are omitted. Making up almost 85% of the total number of attractors, this is a very large and boring subset. It is interesting to note that if we eliminate these attractors from consideration, as well as any symmetric attractors, we have only 105,284,377,928 remaining attractors, a slightly more manageable amount.

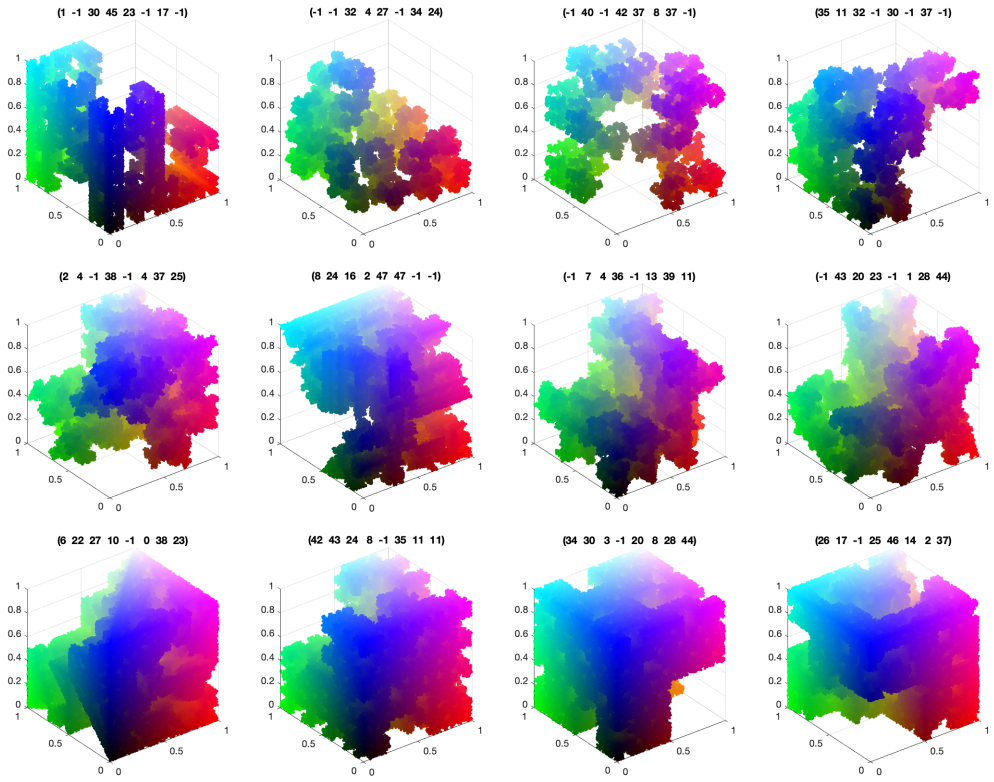


Figure 2.8: 3D Sierpinski relatives with 5, 6, and 7 cubes

2.1.3 Sierpinski Tetrahedron

The Sierpinski Tetrahedron, or fractal (-1, 0, 0, -1, 0, -1, -1, 0), is shown in Figure 2.9. This Fractal has been studied previously, and due to its symmetry allows us to calculate exact values of the Betti growth rates.

Since it is constructed by leaving 4 of the 8 sub-cubes, the fractal dimension of this set is 2. Like the 2D Sierpinski gasket, this set is fully connected, therefore we know it has a disconnectedness index of 0.

The most interesting value for this fractal is the "loopiness" index or γ_1. Because we know the set is connected, we know that each of the largest triangular holes on the faces exist in the true fractal's topology. Thus we know that there will be an ϵ neighborhood of the fractal that is homeomorphic to the second cubical iteration shown in figure 2.10, and more importantly that any holes in this cubical complex correspond to true holes that exist in the fractal set. This is important because we know that not only is the Betti number of the epsilon neighborhood the same as the

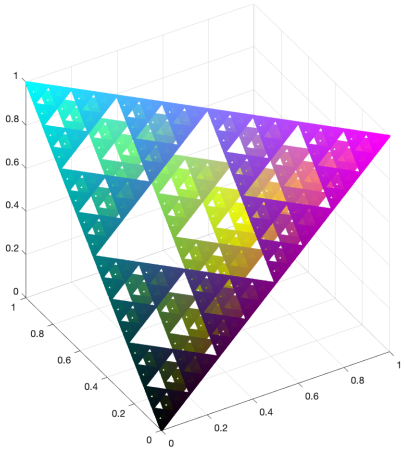


Figure 2.9: The Sierpinski Tetrahedron

Betti number of the cubical complex, but also that their respective persistent Betti numbers are the same. This logic holds for each smaller triangular hole, so we can calculate the true γ_1 index of the fractal in a similar fashion to the box counting dimension if we know β_1 of each iterative cubical complex.

The first iteration, which has only the main triangular holes exposed, can be shown to be homeomorphic to a solid torus with three holes, thus having $\beta_1 = 3$. We conjecture that at each iteration, these three holes remain, but each of the 4 subcubes gives rise to 3 new smaller holes. Thus for $\epsilon = (\frac{1}{2})^n$ we have that $\beta_1 = 3\sum_{i=1}^{n} 4^{i-1} = 4^n - 1$. We then calculate the growth factor:

$$\gamma_1 = \lim_{\epsilon \to 0} \frac{\log\left(\beta_1^{0,\epsilon}\right)}{\log\left(1/\epsilon\right)} = \lim_{n \to \infty} \frac{\log\left(4^n - 1\right)}{\log\left(2^n\right)} = 2$$

Using "Perseus", a persitent homology program [21] which can analyze cubical complexes of this kind, we verify our Betti numbers for the first 7 iterations.

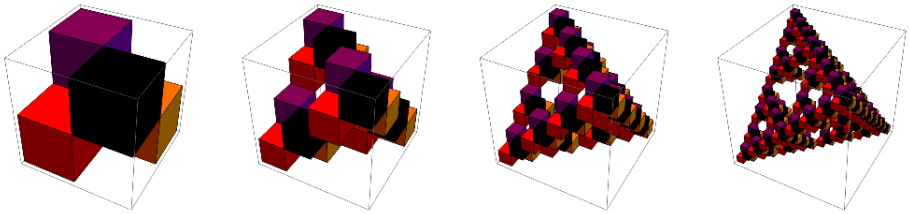


Figure 2.10: Cubical Complexes representing epsilon neighborhoods of the Sierpinski Tetrahedron. The complexes have $\beta_1 = 0$, 3, 15, and 63 respectively.

It is rare that we are able to analytically calculate these growth rates with such ease. We therefore wish to use computational methods to find calculated growth rates. As is often the case for dynamical system attractors, we begin from a point cloud which approximates the set. This is obtained by taking the orbit of an iterated function system with the fractal as the attractor. We use 100,000 points. We then calculate the persistent homology of this point cloud using the alpha complex, through the GUDHI alpha_complex_3d_persistence program [22]. Any features that have persistence intervals of length less than 0.0001 are trimmed to remove excess noisy features. The Barcode diagram is shown below in figure 2.11.

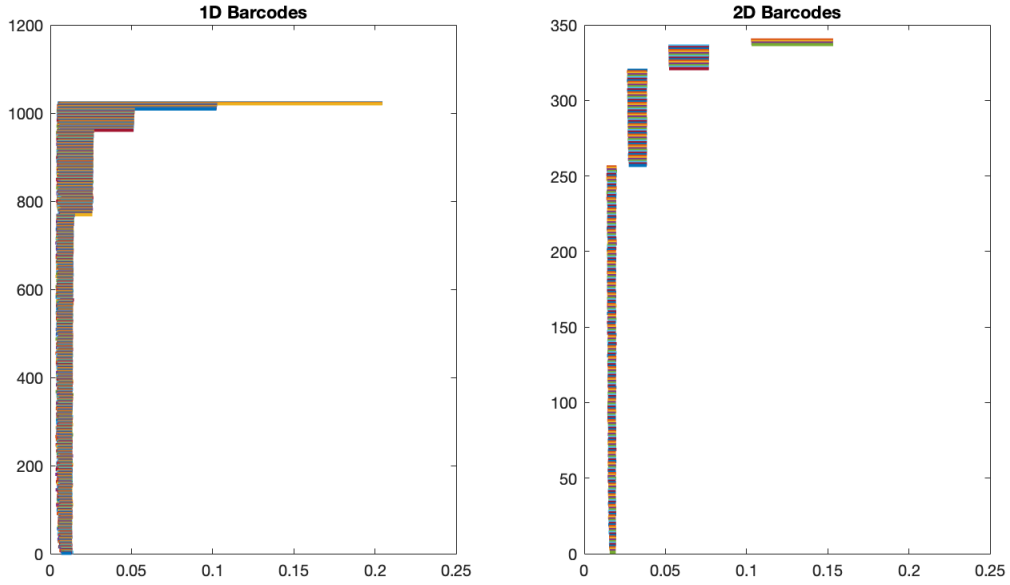


Figure 2.11: The Barcodes for the Alpha complex on 10^5 points approximating the Sierpinski Tetrahedron

Given the persistence intervals, to compute the growth rates of the Betti numbers we need the cuttoff value ρ above which we can be confident that the alpha complex has comparable homology to the ϵ neighborhood of the true fractal. In [4] Robins uses the heuristic:

$$\rho \approx \text{the } \epsilon \text{ value at which there are no isolated points in the filtration}$$

To see how ρ changes with the number of points we calculate ρ for a random orbit for each value between 10 and 200 points. We repeat this calculation 10 times. We plot these data points as well as their average and a curve matched to this data assuming a power-law relationship in figure

2.12. Using this power-law curve our expected cutoff value with 10^5 points should be ≈ 0.016. We conservatively use $\rho = 0.02$ for our calculations.

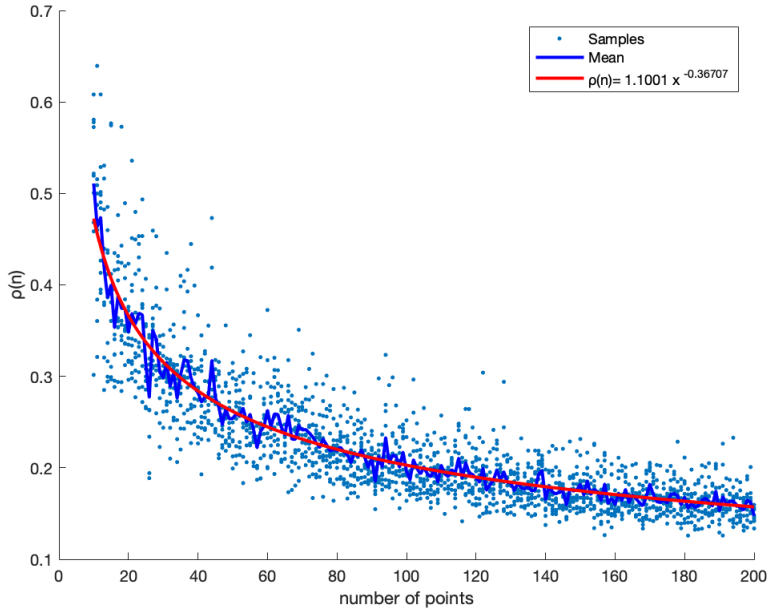


Figure 2.12: The cut-off value ρ for n points distributed randomly on the Sierpinski Tetrahedron

Once we have our cutoff value we can calculate the persistent Betti numbers from the intervals. Figure 2.13 shows both the 1st and 2nd Betti numbers as a function of the filtration parameter. Below these graphs are the persistent Betti numbers $\beta_1^{\rho,\epsilon}$ and $\beta_2^{\rho,\epsilon}$. This shows the important difference between the concepts.

For β_1 we see a large spike near the beginning and then a stair-case like effect as ϵ grows. This is what we expect. Initially there are no holes as each point is an isolated point. As ϵ grows, the points begin connecting and forming holes, many are short lived and noisy. This causes the initial spike. Around ρ the ϵ neighborhood of the points begins to more accurately reflect that of the true underlying attractor. As we grow the epsilon neighborhood from here, there are critical epsilon values at which large groups of holes will close as ϵ exceeds the diameter of the triangular-shaped holes described previously. These values are where there are sudden drops in the number of holes, making the stair effect. The persistent Betti number graph is very similar to the standard Betti number graph, but is only valid for $\epsilon > \rho$. This indicates there are few extra holes that form due to the geometry of the fractal's ϵ neighborhood.

There is however a drastic difference between the graph of $\beta_2(\epsilon)$ and that of $\beta_2^{\rho,\epsilon}$. The top graph shows us that there are voids which form inside of the complex as it grows, and they seem to form and disappear at predictable intervals. This is what we expect. As the triangular holes close, a void forms within the space between them, but the void is then quickly filled in as ϵ reaches the diameter of the space between all 4 triangular loops. The graph of the persistent Betti number however shows these voids "don't count" when we are looking at the growth rates. While some voids have non-trivial persistent interval lengths, the features mostly do not form until after ρ, because they can not correspond to topological features of the true fractal. The persistent Betti number is a constant 0, the growth rate γ_2 is therefore 0 as well.

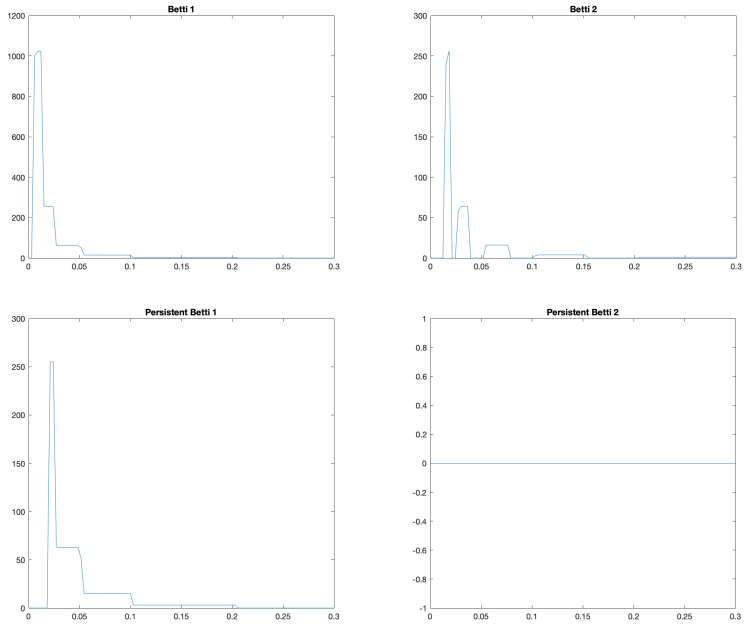


Figure 2.13: The Betti numbers and persistent Betti numbers as a function of ϵ

To find the calculated growth rate of β_1 we find the negative slope of the line formed by plotting the persistent Betti number vs ϵ on a log-log plot. We use data from the range $\rho \approx 0.02 \leq \epsilon \leq 0.2$. This is shown in Figure 2.14. The calculated negative slope is with 95% confidence interval 2.07 ± 0.06. This is above the true value likely because our cutoff value gives us only half of the smallest "step" which overestimates the true value. Using more points could give us a better picture of the true fractal but computation becomes unwieldy. Results are summarized in Figure 2.15.

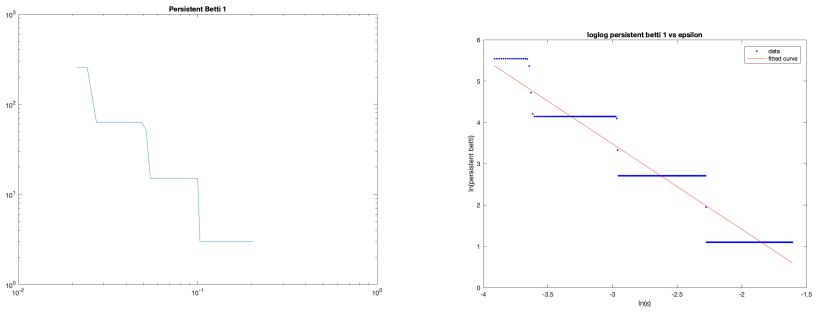


Figure 2.14: The persistent Betti numbers as a function of ϵ

	Exact	Calculated
Dim	2	-
γ_0	0	-
γ_1	2	2.07
γ_2	0	-

Figure 2.15: Results for various indicies of the Sierpinski Tetrahedron

2.1.4 A "Holey" Fractal

The fractals that remove only one subcube have high fractal dimension $\frac{\log(7)}{\log(2)} \approx 2.80735$ and are very dense. If the symmetries are chosen correctly, it would make sense that these fractals have infinitely many cube-shaped voids within them, a perfect example of a non-zero "holiness" index. The fractal shown in figure 2.16, fractal (0, -1, 0, 0, 0, 0, 0, 0) is one of these fractals.

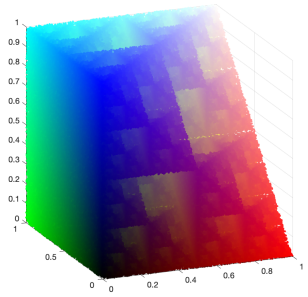


Figure 2.16: A fractals with voids, fractal (0, -1, 0, 0, 0, 0 , 0, 0).

The sequence of β_2 for the cubical complexes of the fractal as shown in figure 2.17, as calculated using Perseus [21] are as follows:

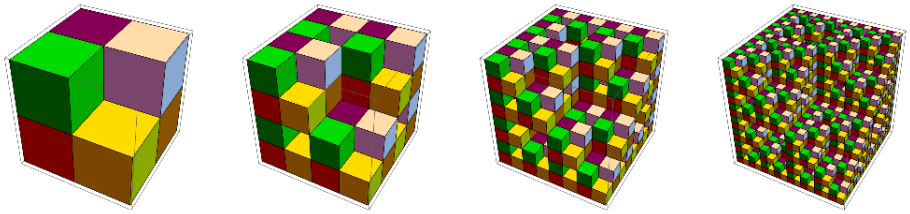


Figure 2.17: Cubical Complexes representing epsilon neighborhoods of fractal. The complex has $\beta_2 = 0$, 1, 17, and 156.

$$1, 17, 156, 1210, 8831, 62907 \tag{2.2}$$

We use an inductive method to find an expression for the sequence. First imagine stacking 7 of the step 0^{th} iterations to create the 1^{st} iteration in figure 2.17 and imagine the single void within this iteration. Symmetry 0 is the identity so no rotations or inversions or reflections are performed. Thus if $\beta_2^{(k)}$ is the number of voids at step k we have $\beta_2^{(1)} = 1$. Now at iteration k consider the 6 faces of the unit cube. 3 faces, the faces along the coordinate planes (not visible in figure 2.17), are completely filled in at each iteration. The other 3 faces however each contain the k^{th} iteration of the classical Sierpinski gasket. Then at the $(\text{k}+1)^{\text{th}}$ iteration we shrink this down and arrange 7 of these k^{th} iteration approximations. This rearrangement creates one large void in the center of the approximation just like in the 1st iteration. Then in addition, each of the 7 k^{th} iteration approximations will give us 7×(the number of voids in in the kth step) or $7\beta_2^{(k)}$. However the arrangement is in such a way that we place 9 of the Sierpinski gasket sides up against 9 of the filled in sides. This creates a void for each one of the loops in the 9 k^{th} Sierpinski gaskets. That is $9\left(\frac{3^k-1}{2}\right)$ more voids. Thus in total we have that

$$\beta_2^{(k+1)} = 1 + 7\beta_2^{(k)} + 9\left(\frac{3^k-1}{2}\right) \qquad k = 1, 2, 3, ... \tag{2.3}$$

Solving this recurrence relation we find the explicit formula

$$\beta_2^{(k)} = \frac{91}{24}7^k - \frac{27}{8}3^k + \frac{7}{12} \tag{2.4}$$

which is fascinatingly always an integer, and matches the sequence (2.2) we calculated.

We wish to show that each of the voids in the cubical complex correspond to true voids in the fractal set and not just the finite approximation. This is needed to show we can use (2.4) as the

persistent Betti number in our calculation of γ_2, not just the Betti number. The 3 sides of the cube which lie on the axes planes

$$P_1 = \{(x, y, z) \mid x \in [0, 1], y \in [0, 1], z = 0\}$$
$$P_2 = \{(x, y, z) \mid y \in [0, 1], z \in [0, 1], x = 0\}$$
$$P_3 = \{(x, y, z) \mid z \in [0, 1], x \in [0, 1], y = 0\}$$

are subsets of the true set. This can be shown using the IFS representation of the the fractal. By self-similarity we thus know that the 3 sides of each subcube parallel to the axes planes are all also in the attractor, for example

$$P = \{(x, y, z) \mid y \in [1/2, 1], z \in [1/2, 1], x = 1/2\}$$

is a subset of the fractal as well. We can then show that each of the voids counted in (2.4) is contained within 6 of these planes and thus corresponds to a true void in the fractal set. This justifies the computation of the "holiness" index or the 2^{nd} growth rate to be

$$\gamma_2 = \lim_{\epsilon \to 0} \frac{\log\left(\beta_2^{0,\epsilon}\right)}{\log\left(1/\epsilon\right)} = \lim_{n \to \infty} \frac{\log\left(\frac{91}{24}7^k - \frac{27}{8}3^k + \frac{7}{12}\right)}{\log\left(2^n\right)} = \frac{\log 7}{\log 2} \approx 2.80735$$

which is exactly the same as the dimension of the fractal. Robins conjectures in [4] that if X is a self-similar fractal set with $\gamma_i \neq 0$ then necessarily $\gamma_i = \dim_S(X)$ where $\dim_S(X)$ is the similarity dimension. The similarity dimension is the same as the box-counting dimension in our case, see [23] for more on self-similarity. Both fractals we have analyzed support this conjecture.

We perform similar computational analysis for this fractal as we did for the Sierpinski tetrahedron. However even with 10^6 points we are unable to calculate γ_2 to a useable precision. The cutoff ρ decays very slowly with the number of points: using the same methods as before we estimate $\rho \approx 0.025$, even worse than the tetrahedron. This is due to the greater density of points needed to approximate this fractal with higher fractal dimension. With more points needed computation of the alpha complex also becomes difficult and slow. However, the persistent Betti numbers $\beta_2^{\rho,\epsilon}$ do qualitatively support our analysis as seen in figure 2.18.

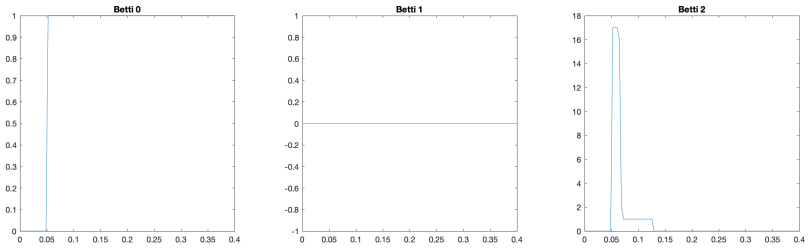


Figure 2.18: Persistent Betti numbers $\beta_i^{\rho,\epsilon}$ as a function of epsilon with $\rho = 0.025$. Betti 2 shows the stair-case pattern we expect, at values 1 and 17, the theoretical values $\beta_2^{(1)}$ and $\beta_2^{(2)}$ from (2.4).

	Exact
Dim	2.80735
γ_0	0
γ_1	0
γ_2	2.80735

Figure 2.19: Results for the indicies of the fractal

Chapter 3

Conclusion

Despite thier interesting topological and geometric features, fractals like the Sierpinski relatives can seem contrived. It is difficult to see an application for these abstract shapes. However, there is value in studying fractals like these both from an educational as well as an artistic standpoint.

Fractal sets are common in nature, dynamical systems, and other real-world applications. It is therefore useful to use somewhat contrived examples like the Sierpinski gasket to demonstrate the kinds of features fractals can have such as non-integer fractal dimension and topological growth rates. Because these examples are so artificial, they have the advantage that these properties can be derived analytically. Then techniques such as topological data analysis and persistent homology can be tested on these sets in order to verify their accuracy and effectiveness.

Our extension of the Sierpinski relatives to 3D extends the uses of the 2D relatives. Just as in the 2D case, we can analytically derive many properties of these fractals and then perform other computational techniques to test their accuracy. Adding a dimension adds complexity to the types of examples we can analyze, for example the classical relatives cannot have a non-zero γ_2 as our "holey" fractal does.

In addition to these academic uses, fractals like these are simply beautiful and interesting to look at. Self similarity and fractal geometry has been seen in art since ancient times, see [24] for an example of a Sierpinski triangle found on the floor of a medieval roman church. The 3D fractals are additionally intriguing and interesting to look at. We hope to be able to visualize and 3D print some of these fractals in the future.

In addition to future visualization it would be nice to completely categorize the topologies that

exist within our set of fractals as has been done with the 2D relatives. Given the large number of fractals that exist this may prove difficult, but if work is done to identify the number of unique fractals up to symmetry it may be a more manageable computation.

Persistent homology and topological data analysis have been shown to be useful data analysis techniques, from identifying a sub-group of breast cancers [25] to analyzing coverage in sensor networks [26]. As technology evolves, we are faced with the analysis of larger and larger data sets and having examples in 3D is important for education. In higher dimensions visualizing and understanding the "shape" of the data becomes increasingly difficult, but persistent homology allows us to quantify and explore the data in new ways.

Analyzing real-world data sets similarly to how we analyze the point-cloud approximation of the fractals would also be an interesting next-step. Mandelbrot discusses how common fractals are in nature [23] and how terrain, clouds, and plants can all be modeled with an iterative fractal process. Work has also been done to identify trees based on the fractal dimension of their leaves, see [27]. Perhaps 3D data from these types of natural objects could be used similarly, for example identifying plants based on their "loopiness", or rocks and clouds based on their "holiness". This could lead to new tools for object identification and recognition.

Bibliography

[1] A. Hatcher, C. U. Press, and C. U. D. of Mathematics, *Algebraic Topology*. Algebraic Topology, Cambridge University Press, 2002.

[2] J. Munkres, *Elements Of Algebraic Topology*. CRC Press, 2018.

[3] J. Hocking and G. Young, "Topology addison wesley," *Reading, Massachusetts*, 1961.

[4] V. Robins, *Computational topology at multiple resolutions: foundations and applications to fractals and dynamics*. PhD thesis, University of Colorado, 2000.

[5] H. J. S. Smith, "Xv. on systems of linear indeterminate equations and congruences," *Philosophical transactions of the royal society of london*, no. 151, pp. 293–326, 1861.

[6] R. W. Ghrist, *Elementary applied topology*, vol. 1. Createspace Seattle, 2014.

[7] R. Ghrist, "Barcodes: the persistent topology of data," *Bulletin of the American Mathematical Society*, vol. 45, no. 1, pp. 61–75, 2008.

[8] E. Carlsson, G. Carlsson, and V. De Silva, "An algebraic topological method for feature identification," *International Journal of Computational Geometry & Applications*, vol. 16, no. 04, pp. 291–314, 2006.

[9] H. Edelsbrunner, "Alpha shapes—a survey,"

[10] H. Edelsbrunner and E. P. Mücke, "Three-dimensional alpha shapes," *ACM Transactions on Graphics (TOG)*, vol. 13, no. 1, pp. 43–72, 1994.

[11] N. F. Sanderson, "Topological data analyses of time series using witness complexes," 2018.

[12] K. Mischaikow, T. Kaczynski, and M. Mrozek, "Computational homology," *Applied Mathematical Sciences 157*, 2004.

[13] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," in *Proceedings 41st annual symposium on foundations of computer science*, pp. 454–463, IEEE, 2000.

[14] A. Tausz, M. Vejdemo-Johansson, and H. Adams, "JavaPlex: A research software package for persistent (co)homology," in *Proceedings of ICMS 2014* (H. Hong and C. Yap, eds.), Lecture Notes in Computer Science 8592, pp. 129–136, 2014. Software available at `http://appliedtopology.github.io/javaplex/`.

[15] H. Edelsbrunner and J. Harer, "Persistent homology-a survey," *Contemporary mathematics*, vol. 453, pp. 257–282, 2008.

[16] N. Sanderson, E. Shugerman, S. Molnar, J. D. Meiss, and E. Bradley, "Computational topology techniques for characterizing time-series data," in *International Symposium on Intelligent Data Analysis*, pp. 284–296, Springer, 2017.

[17] S. Strogatz, "Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering (studies in nonlinearity)," 2001.

[18] V. Robins, J. D. Meiss, and E. Bradley, "Computing connectedness: An exercise in computational topology," *Nonlinearity*, vol. 11, no. 4, p. 913, 1998.

[19] H.-O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and fractals: new frontiers of science*. Springer Science & Business Media, 2006.

[20] M. F. Barnsley, *Fractals everywhere*. Academic press, 2014.

[21] V. Nanda, "Perseus, the persistent homology software," 2020.

[22] The GUDHI Project, *GUDHI User and Reference Manual*. GUDHI Editorial Board, 3.1.1 ed., 2020.

[23] B. B. Mandelbrot, *The fractal geometry of nature*, vol. 173.

[24] E. Conversano and L. T. Lalli, "Sierpinski triangles in stone, on medieval floors in rome," *J. Appl. Math*, vol. 4, pp. 114–122, 2011.

[25] 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*, vol. 108, no. 17, pp. 7265–7270, 2011.

[26] V. De Silva and R. Ghrist, "Coverage in sensor networks via persistent homology," *Algebraic & Geometric Topology*, vol. 7, no. 1, pp. 339–358, 2007.

[27] O. M. Bruno, R. de Oliveira Plotze, M. Falvo, and M. de Castro, "Fractal dimension applied to plant identification," *Information Sciences*, vol. 178, no. 12, pp. 2722–2733, 2008.

[28] D. S. Dummit and R. M. Foote, *Abstract algebra*, vol. 3. Wiley Hoboken, 2004.

[29] E. W. Weisstein, "Octahedral group."

Appendix A

Group Theory

Some definitions from group theory follow. These are necessary for understanding homology theory as discussed in chapter 1. Often groups and free groups are described using multiplicative notation ($\cdot$ instead of $+$), however the groups we deal with in this thesis, particularly the chain groups and their subsets, use additive notation so we adopt that here.

Definition of a Group

A **Group**, denoted $G = (S, +)$, is a set S along with an operation $+$ that satisfies the following properties:

- **Closure:** for all $a, b \in S$, $a + b$ is in S as well.

- **Associativity:** for all $a, b, c \in S$, we have $(a + b) + c = a + (b + c)$

- **Identity:** there is a unique identity element e such that $a + e = e + a = a$ for all $a \in S$

- **Inverse:** for all $a \in S$ there is an inverse element $-a$ such that $a + (-a) = (-a) + a = e$

Definition of a Free Group

A **Free Group** is the infinite group on the set of all "words" or sums that can be constructed from a generating set S and its inverse S^{-1}. For example consider the set $S = \{a, b\}$ and its inverse $S^{-1} = \{-a, -b\}$. We denote the Free group with S as a generating set as F_s.

$$F_S = (\{a, b, -a, -b, a + a = 2a, a + b, b - a, a + (-a) + b + b = 2b, ...\}, +)$$

Appendix B

Symmetry Groups

We discuss the symmetry groups for both the square and cube. These are important to understand the Sierpinski relatives. For more see [19].

A square has a number of symmetries: ways in which we can flip/rotate the shape in order to give us the same square. For example we can flip the square along its diagonal and the resulting shape is identically a square. We could rotate the square 90 degrees and again get a square. These examples are illustrated in Figure B.1. We inscribe an L in the square to visualize the operation performed.

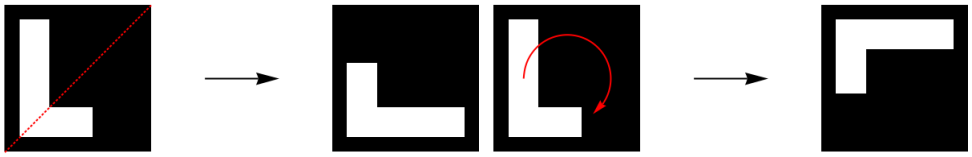


Figure B.1: Examples of two symmetries of the square, reflection about the diagonal (d_6) and rotation by 90 degrees (d_3)

The set of possible symmetries of a square can be equipped with the operation of concatenation, thus creating a group. For example, if we flip the square down its middle and then rotate it 90 degrees clockwise, we get the same square we would if we simply reflected the square along its diagonal. This is illustrated in Figure B.2. We can denote each element of the symmetry group of the square by d_i. The index of each symmetry is arbitrary, but we use the ordering as defined in [19]. Flipping along the vertical center line is d_5 and rotating 90 degrees is d_3, so we have $d_5 d_3 = d_6$

where d_6 is the symmetry representing reflection about the diagonal. There are 8 symmetries total. This is the 4th dihedral group, D_4. The elements of D_4 representing the symmetries of the square are shown in Figure B.3. See [28] for more on dihedral groups and group theory.

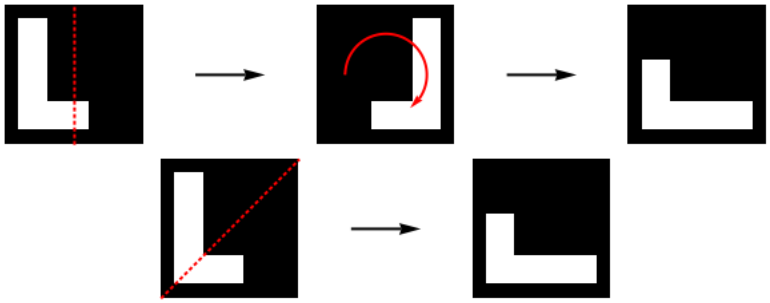


Figure B.2: Preforming two symmetry operations results in a third different operation

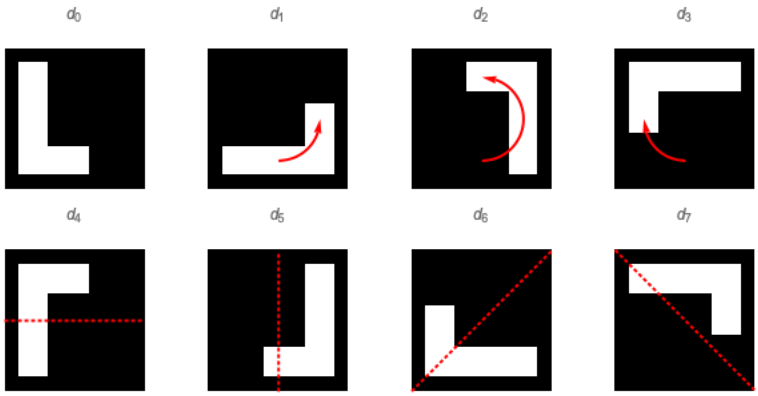


Figure B.3: All 8 symmetries of the square

Since we know two elements concatenated will give us a third element, we can create a "multiplication table" of sorts which shows what the "product" of two group elements is. This is commonly called a Cayley table. We show the Cayley table of D_4 in figure B.4.

This symmetry group is important for our later examples of the Sierpinski relatives. The size of this group is directly related to how many fractals we can create. See the following section for more on this.

In chapter 2 we will extend the 2D Sierpinski relatives to a new class of 3D fractals. In 2D we need the symmetry group of the square, in 3D we are interested in the symmetries of the cube. This is also a commonly studied group, called the full octahedral symmetry group [29]. Whereas the square's symmetry group has 8 elements/symmetries, this group has 48. We show a few of

		i							
		0	1	2	3	4	5	6	7
	0	0	1	2	3	4	5	6	7
	1	1	2	3	0	7	6	4	5
	2	2	3	0	1	5	4	7	6
	3	3	0	1	2	6	7	5	4
j	4	4	6	5	7	0	2	1	3
	5	5	7	4	6	2	0	3	1
	6	6	5	7	4	3	1	0	2
	7	7	4	6	5	1	3	2	0

Figure B.4: The Cayley Table for the D_4 group. Choosing a column i and a row j gives the subscript for operation $d_i d_j$

the elements of this group in Figure B.6. A 3D analog of the inscribed L is placed in the cube to visualize the operation on the cube in 3D space. The ordering of the elements is again arbitrary but we use the conventional ordering where the first 24 elements follow the ordering of the symmetric group S_4 (the subgroup excluding inversions). The final 24 are the inversions of those first 24. We also show the Cayley table of this group, but since it is so large we visualize the table with colors instead of numbers, see Figure B.5.

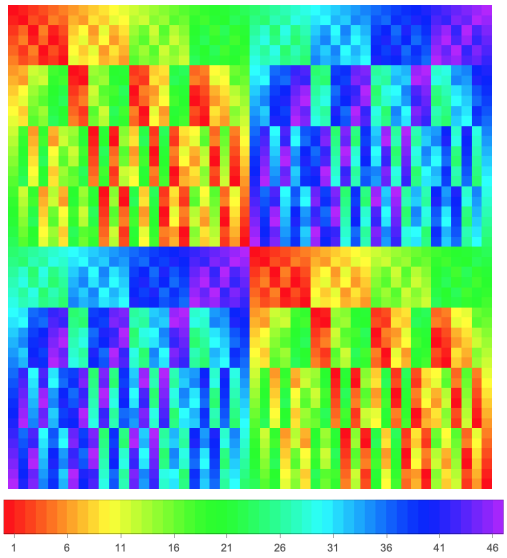


Figure B.5: Visualization of the Cayley table for the 3d Cube's symmetry group. Each element from 0-47 is represented by a color ranging from red to pink.

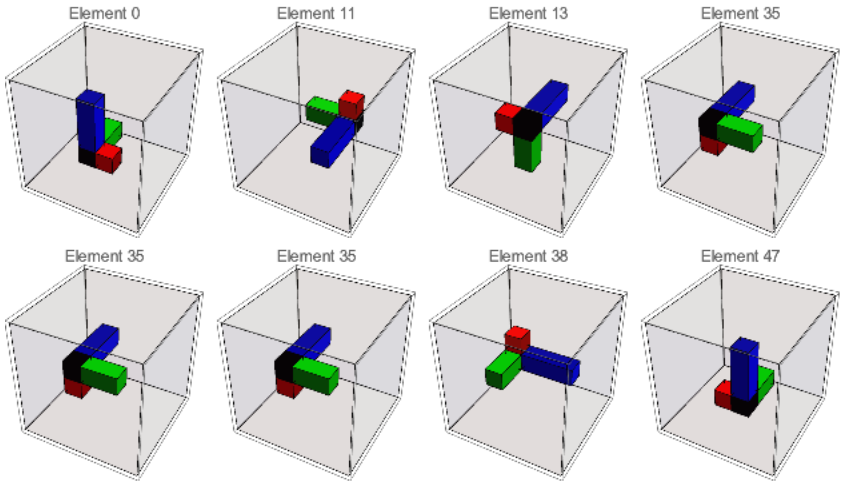


Figure B.6: 8 symmetries of the cube. Element 0 is the identity, others show rotations/reflections/inversions.