Computational Topology at Multiple Resolutions: Foundations and Applications to Fractals and Dynamics

Vanessa Robins

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Applied Mathematics
June 2000

Abstract

Extracting qualitative information from data is a central goal of experimental science. In dynamical systems, for example, the data typically approximate an attractor or other invariant set and knowledge of the structure of these sets increases our understanding of the dynamics. The most qualitative description of an object is in terms of its topology — whether or not it is connected, and how many and what type of holes it has, for example. This thesis examines the degree to which such topological information can be extracted from a finite point-set approximation to a compact space. We consider both theoretical and computational aspects for the case of homology.

Any attempt to extract topological information from a finite set of points involves coarse-graining the data. We do this at multiple resolutions by forming a sequence of ϵ-neighborhoods with ϵ tending to zero. Our goal is to extrapolate the underlying topology from this sequence of ϵ-neighborhoods. There is some subtlety to the extrapolation, however, since coarse-graining can create spurious holes — a fact that has been overlooked in previous work on computational topology. We resolve this problem using an inverse system approach from shape theory.

The numerical implementations involve constructions from computational geometry. We present a new algorithm based on the minimal spanning tree that successfully determines the apparent connectedness or disconnectedness of point-set data in any dimension. For higher-order homology, we use existing algorithms that employ Delaunay triangulations and alpha shapes. We evaluate these techniques by comparing numerical results with the known topological structure of some examples from discrete dynamical systems. Most of the objects we study have fractal structure. Fractals often exhibit growth in the number of connected components or holes as ϵ goes to zero. We show that the growth rates can distinguish between sets with the same Hausdorff dimension and different homology. Relationships between these growth rates and various definitions of fractal dimension are derived.

Overall, the thesis clarifies the complementary role of geometry and topology and shows that it is possible to compute accurate information about the topology of a space from a finite approximation to it.

Acknowledgments

I am deeply indebted to many people for their support over the course of this project. I especially thank Liz Bradley and Jim Meiss for being such fantastic advisors and mentors. They were remarkably generous with their time and expert knowledge, and have taught me a tremendous amount about all facets of academic life. This thesis also benefited from many insights, questions, and suggestions from James Curry, Bob Easton, and Mike Eisenberg. I appreciate their interest and encouragement more than they probably realise. My library has expanded steadily over the years thanks to many contributions from Professor Curry.

I am eternally grateful to fellow graduate students and other friends for their moral support, constructive critiques of practice talks, and conversations about mathematics. The friendly atmosphere in the department has made the last five years a very enjoyable time. I hesitate to make a list in case someone is left out, but must particularly thank Travis Austin, Allison Baker, Lora Billings, Danielle Bundy, John Carter, Andrea Codd, Nancy Collins, Bernard Deconinck, Matt Easley, Rod Halburd, Laurie Heyer, Apollo Hogan, Rudy Horne, Joe Iwanski, Ken Jarman, Martin Mohlenkamp, Cristina Perez, Kristian Sandberg, David Sterling, Reinhard Stolle, Josh Stuart, Matt Tearle, David Trubatch, and Eric Wright. Tony Edgin, Hugh MacMillan, and Peter Staab were great housemates as well as helpful colleagues.

My studies were funded by Liz Bradley's NSF National Young Investigator award #CCR-9357740 and Packard Fellowship in Science and Engineering. Much-appreciated additional funding came from the University of Colorado Graduate School, the Sheryl Young Memorial Scholarship, and the Francis Stribic Graduate Fellowship. I am also grateful to the School of Mathematical Sciences at the Australian National University for granting me visitor privileges for the period July–December, 1997.

My partner, Karl Claxton, showed endless patience and good humour throughout this endeavour, and helped me in countless different ways. Finally, I thank my parents for imparting a love of learning, and for providing their unceasing support. This thesis is dedicated to them.

Contents

Chapter 1

Introduction

In this thesis we consider the problem of extrapolating information about the topological structure of a space from a finite approximation to it. The motivation for this work comes from the study of chaotic dynamical systems. A fundamental goal in this field is the extraction of qualitative information from data, e.g., geometric and topological properties of invariant sets. Much work on analyzing geometric structure focusses on the fractal dimensions of attractors. Techniques for extracting topological information, however, have been restricted to smooth sets. In this dissertation, we develop an approach to computational topology that is general enough for application to both smooth and fractal data. Our analysis of quantities such as the number of components and holes at multiple resolutions yields a new way to characterize fractal structure that is related to, but distinct from, the concept of fractal dimension. We contribute to the emerging field of computational topology by developing sound foundations for both the extrapolation and approximation problems.

This thesis sits at the intersection of three areas of research: dynamical systems, fractal geometry, and computational topology. We begin with a brief survey of the relevant literature from each of these fields in this chapter.

1.1 Extracting qualitative information from data

The qualitative theory of dynamical systems takes a global geometric perspective in understanding the time-evolution of a system. Typically, this involves describing the phase space — a geometric representation of all possible trajectories from the flow of a differential equation, for example, or the iteration of a map. Much information about the dynamics can be deduced from the structure of invariant subsets of phase space.[1] We rarely have an analytic description of these objects; experimental data are always finite, and even if the equations of motion are known, they are usually too complicated to solve exactly. Instead, numerical integration of the governing equations is used to generate computer visualizations of the phase space. These visualizations help guide formal results, especially when the phase space is two or three dimensional. It is very difficult to visualize higher-dimensional spaces on a computer screen, however; even in a three-dimensional space it is almost impossible to see detailed structure in a cloud of points. Therefore, numerical tools that extract qualitative information from data are useful for providing intuition into the behavior of a dynamical system.

[1] In fact, for Hamiltonian systems of two degrees of freedom, geometric and topological information has been used in an artificial intelligence approach to the automatic classification of orbits [89].

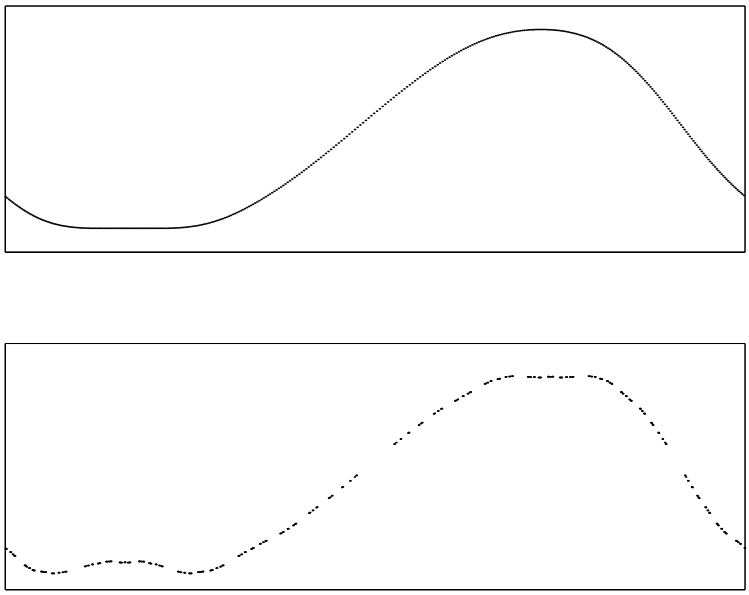


Figure 1.1: Approximations to a quasiperiodic orbit in the standard map for different values of its parameter, k. The top orbit ($k = 0.7$) approximates an invariant circle; the bottom orbit ($k = 1.0$) covers a Cantor set. The transition from circle to Cantor set is a dramatic change in the topology and has important consequences for the dynamics. The invariant circles trap chaotic orbits, which implies the dynamics is relatively confined. A Cantor set is totally disconnected so the chaotic orbits can diffuse through the gaps and the momentum variable is no longer bounded. We study this transition in Chapter 4 by counting the number of connected components of the data as a function of resolution.

Geometric properties of an invariant set that are of interest in the study of chaotic dynamics include its Lebesgue measure, the density distribution of points from an orbit, and its fractal dimension. For a certain class of chaotic attractors, the Lyapunov exponents (which are basically averaged eigenvalues) are related to the box-counting dimension of the attractor [64]. The study of dynamical systems has driven a substantial amount of research on fractal geometry, including multifractal analysis and fat fractal exponents. We discuss these concepts in Section 1.2.2.

Topological properties, such as the number of connected components or holes, are more fundamental but more difficult to extract from data. In Figures 1.1, 1.2 and 1.3 we sketch some examples that illustrate the type of topological properties we are interested in. We examine these examples in detail in Chapter 4.

Previous work on extracting topological information from data ranges from determining the topological dimension of an attractor [54, 65], to applications of knot theory to model flows in $\mathbb{R}^3$ [28, 57, 82], to the computation of homology groups [37, 58, 60]. The topological dimension of an attractor is a measure of the number of degrees of freedom of the dynamics. Fluid flow, for example, is modelled by partial differential equations, but in some situations, the essential behavior can be described by a low-dimensional differential equation. For attractors that are embedded in $\mathbb{R}^3$, the knot and link invariants of unstable periodic orbits can be used to build a template for the dynamics. The template generates equations that model the flow and allow the prediction of other periodic orbits. This gives a way to test the validity of a given template against the experimental data by verifying the presence of the predicted periodic or-

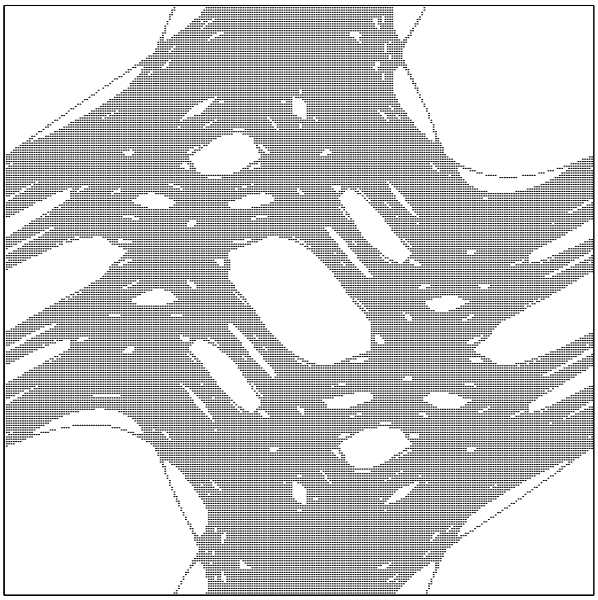


Figure 1.2: A chaotic region of the standard map with $k = 1.1$. The picture is generated by recording which boxes in a 256×256 grid are visited by a single chaotic orbit of 10^8 points. The chaotic region is connected and appears to have positive area which tells us that a significant proportion of initial conditions will lead to chaotic motion. There are holes on a range of scales, however, so not all trajectories are chaotic. These holes are caused by resonance zones around periodic orbits. This set is an example of a fat fractal because it has positive area and holes on arbitrarily fine scales.

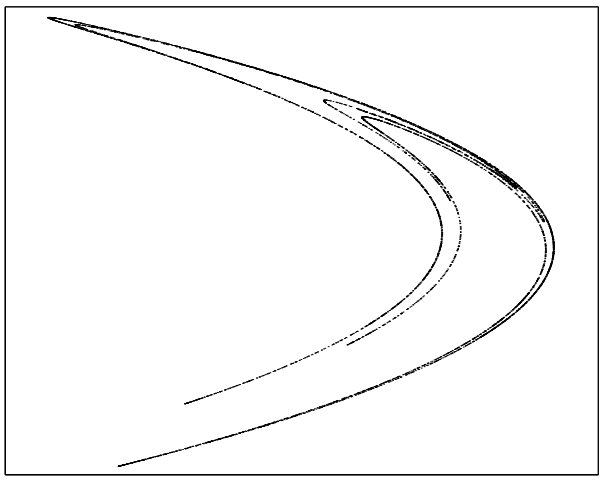


Figure 1.3: The Hénon attractor, one of the standard examples in chaotic dynamics. The set is connected and it is often described as having a Cantor set cross-section. We confirm this structure using our computational techniques in Chapter 4. The attractor has a box-counting dimension of approximately 1.27. Orbits cover the attractor in a slightly non-uniform way which gives the attractor multifractal properties.

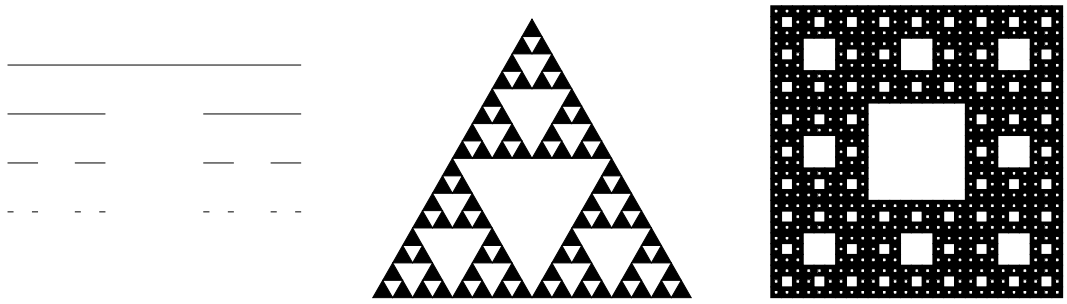


Figure 1.4: The Cantor set, Sierpinski triangle, and Sierpinski curve — standard fractal examples.

bits. Homology groups are topological invariants that use algebraic techniques to describe the topology of a space in terms of equivalence classes of k-dimensional cycles. If two spaces have different homology groups then they cannot be homeomorphic. An implication of this is that if two attractors have different homology, then the dynamics cannot be equivalent. Homology groups are also important in Conley index theory; see [15] for an introduction to this topic. The first step in computing homology groups is to construct a finite triangulation of the space. Muldoon *et al.* [60] do this for embedded time series data that cover a manifold using a standard procedure called taking the nerve of a cover. The cover is made of overlapping patches of approximately linear subsets of the embedded data. The simplices in the nerve are generated by the intersections of the patches. Mischaikow *et al.* [37, 58] use cubical, rather than simplicial, complexes in their application of Conley index theory to chaotic time series data. We discuss other approaches to computational homology in Section 1.3.

1.2 Fractal geometry

As we have already observed, invariant sets from chaotic dynamical system are often fractals, and this is a fundamental consideration in our work. In this section, we give an overview of some basic concepts in fractal geometry. We start with some examples, then discuss fractal dimensions and other ways to characterize fractal structure. We finish with a description of iterated function systems and the similarity dimension.

The term "fractal" is nebulous — it is difficult to give a precise definition without excluding some interesting cases. Instead, Falconer [23] lists some properties that are common to most fractals, but these are neither essential nor exhaustive.

1. A fractal, X, typically has fine structure, i.e., detail on arbitrarily small scales.
2. X is too irregular to be described in traditional geometrical language.
3. The "fractal dimension" of X exceeds its topological dimension.
4. Often X is self-similar, at least in an approximate or statistical sense.
5. Many examples have a simple, recursive definition.

Classical examples include the middle-third Cantor set, the Sierpinski triangle, and the Sierpinski curve; see Figure 1.4. The *middle-third Cantor set* is the subset of $[0, 1]$ that remains after removing the middle-third interval, then repeatedly removing the middle third of the remaining

intervals, *ad infinitum.* This set is *compact* (closed and bounded), *totally disconnected* (each connected component is a single point), and *perfect* (every point is a limit point). We study these properties in more detail in Chapter 2. Since the Cantor set is totally disconnected, it has a topological dimension of zero, its fractal dimension is $\log 2/\log 3$.

The *Sierpinski triangle* (or *Sierpinski gasket*) is constructed in a similar manner. One starts with a filled triangle and remove the central filled triangle with vertices at the midpoints of the edges, then does this repeatedly to the remaining triangles. The resulting set is connected and topologically one-dimensional, but its fractal dimension is $\log 3/\log 2$. We use the Sierpinski triangle and some related fractals as examples throughout this thesis; see Figure 1.5.

If we start with a square, divide it into nine squares, remove the central one and repeat with the remaining eight, and so on, then the resulting set is the *Sierpinski curve (or carpet).* Although this set appears to have very similar topology to the Sierpinski triangle, they are fundamentally different. The Sierpinski curve contains a homeomorphic image of every plane continuum [87], but the Sierpinski triangle does not. The difference has to do with the possible values of the branching order of a point. The Sierpinski triangle has points with branching order of $2, 3$, and 4, but the Sierpinski carpet has points of every order; see [66] for details. The techniques we develop in this thesis fail to distinguish this difference in topology.

1.2.1 Fractal dimensions

The most popular tools for describing fractal structure are the various formulations of fractal dimension. Roughly speaking, a fractal dimension is a number that represents the amount of space occupied by a set. It generalizes the intuitive notion that a point is zero-dimensional, a line is one-dimensional, and so on. If a curve has infinite length, say, but zero area, then it should have a fractal dimension between one and two. The two most commonly used definitions are the Hausdorff dimension and the box-counting dimension. The former is based on a construction from measure theory and therefore has the most useful mathematical properties, but it is difficult to compute for specific examples. The box-counting dimension is based on the notion of "measurement at scale ϵ." This definition is easy to work with and straightforward to implement computationally. We give formal definitions of these dimensions below.

Hausdorff dimension

Given a subset, X, of a separable metric space, the definition starts with a countable ϵ-cover, $\mathcal{U}_\epsilon$ — a collection of open sets U such that $\operatorname{diam} U \leq \epsilon$, and whose union contains the set X. (Recall that the diameter of a set is the largest distance between any two points in the set: $\operatorname{diam} U = \sup\{d(x, y) \mid x, y \in U\}$.) The *s-dimensional Hausdorff outer measure* of X is

$$\mathcal{H}^s_\epsilon(X) = \inf \sum_{U \in \mathcal{U}_\epsilon} (\operatorname{diam} U)^s \tag{1.1}$$

where the infimum is taken over all countable ϵ-covers of X. The *s-dimensional Hausdorff measure* of X is then:

$$\mathcal{H}^s(X) = \lim_{\epsilon \to 0} \mathcal{H}^s_\epsilon(X). \tag{1.2}$$

If s is an integer, then $\mathcal{H}^s(X)$ is equivalent to Lebesgue measure.

When $\epsilon < 1$ we know $\epsilon^s > \epsilon^t$ if $s < t$. It follows that $\mathcal{H}^s_\epsilon(X) \geq \mathcal{H}^t_\epsilon(X)$, and therefore that $\mathcal{H}^s(X) \geq \mathcal{H}^t(X)$. In fact, the following theorem holds; see Falconer [23] or Edgar [19].

Theorem 1. *If $\mathcal{H}^s(X) < \infty$ and $s < t$, then $\mathcal{H}^t(X) = 0$. Conversely if $\mathcal{H}^s(X) < \infty$ and $t < s$, then $\mathcal{H}^t(X) = \infty$.*

This theorem implies that there is a unique value of s where the Hausdorff measure jumps from infinity to zero. This value of s is the *Hausdorff dimension*:

$$\dim_H(X) = \inf\{s : \mathcal{H}^s(X) = 0\} = \sup\{t : \mathcal{H}^t(X) = \infty\} \tag{1.3}$$

We do not explicitly use the Hausdorff dimension for any of our examples or proofs, but include the definition since it is typically what mathematicians mean when they use the term fractal dimension.

Box-counting dimension

Suppose that at least $N(\epsilon)$ sets of diameter ϵ are needed to cover X. Then $\mathcal{H}^s_\epsilon(X) \leq N(\epsilon)\epsilon^s$. If $s = \dim_H$ and $\mathcal{H}^s(X)$ is finite, then we expect $N(\epsilon) \sim \epsilon^{-s}$ as $\epsilon \to 0$. The box-counting dimension is therefore defined as

$$\dim_B(X) = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{-\log \epsilon}. \tag{1.4}$$

Of course, this limit may not exist, in which case the $\limsup$ and $\liminf$ are used. The above heuristic does not guarantee that box-counting and Hausdorff dimensions are equivalent; in general, $\dim_H \leq \dim_B$ and there are compact sets for which the two differ. See Falconer [23] for further details.

The number $N(\epsilon)$ can be defined in many ways, all of which yield an equivalent value of $\dim_B$ (see [23] for details). Some definitions of $N(\epsilon)$ include:

1. the smallest number of closed balls of radius ϵ that cover X;
2. the smallest number of cubes of side ϵ that cover X;
3. the number of ϵ-mesh cubes that intersect X;
4. the smallest number of sets of diameter at most ϵ that cover X;
5. the largest number of disjoint balls of radius ϵ with centers in X.

One uses whichever definition is most convenient. Numerical algorithms for estimating the box-counting dimension mostly use definition 3 and meshes of boxes with side 2^k [4]. The box-counting dimension is typically what physicists mean by the term fractal dimension.

1.2.2 Other characterizations of fractal structure

The dimension of a fractal is a measure of geometric irregularity on small scales. The Hausdorff and box-counting dimensions are each invariant under bi-Lipschitz transformations: those for which there exist c_1, c_2 such that

$$c_1|x - y| \leq |f(x) - f(y)| \leq c_2|x - y|. \tag{1.5}$$

This makes them geometric invariants of sorts, but not topological invariants. All one can say about the topology of a set given its Hausdorff dimension is that $\dim_H < 1$ implies the set is totally disconnected. The converse is certainly not true — Cantor sets can be constructed with

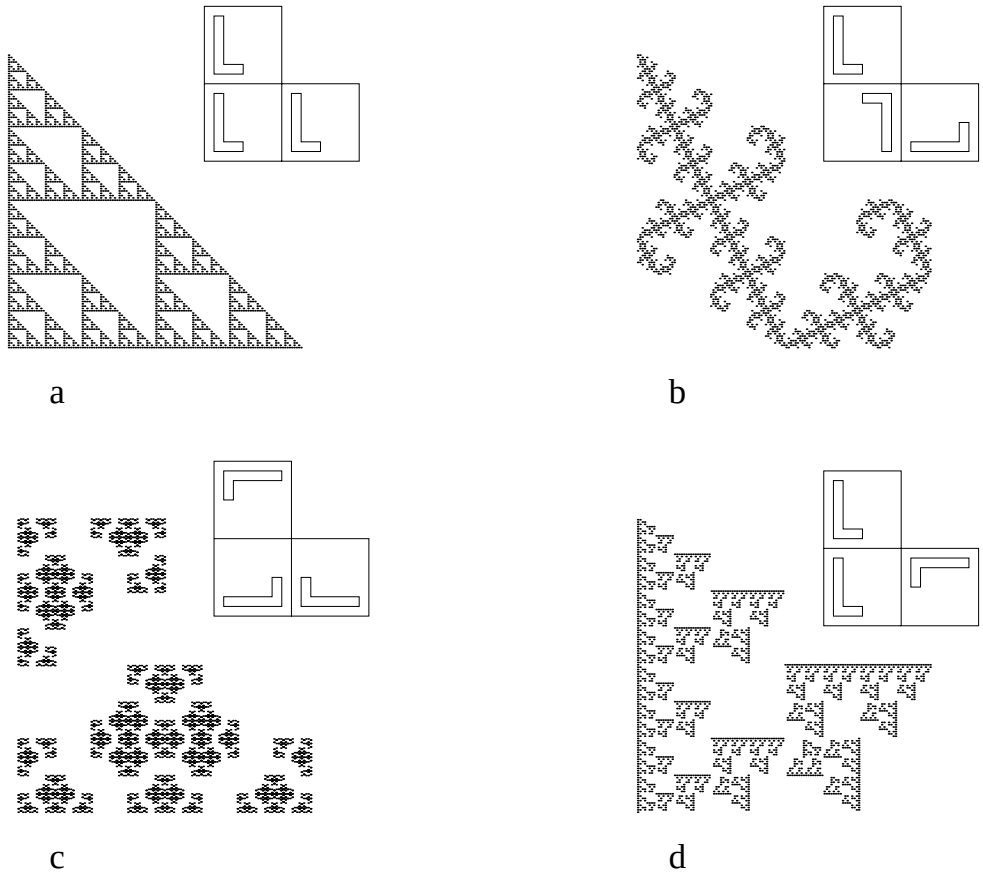


Figure 1.5: Four relatives of the Sierpinski triangle. Each of these fractals is generated as the attractor for an iterated function system (Section 1.2.3) that maps the unit square into three squares of one half the size. The "L" notation designates the rotation or reflection used in each case. These fractals have exactly the same Hausdorff dimension ($\log 3 / \log 2$) but their topology is different. We use these fractals in Chapters 2 and 3 as test examples for our numerical techniques. By looking at the number and size of connected components, and the number of holes as a function of resolution we can distinguish their different topological structure of these four fractals.

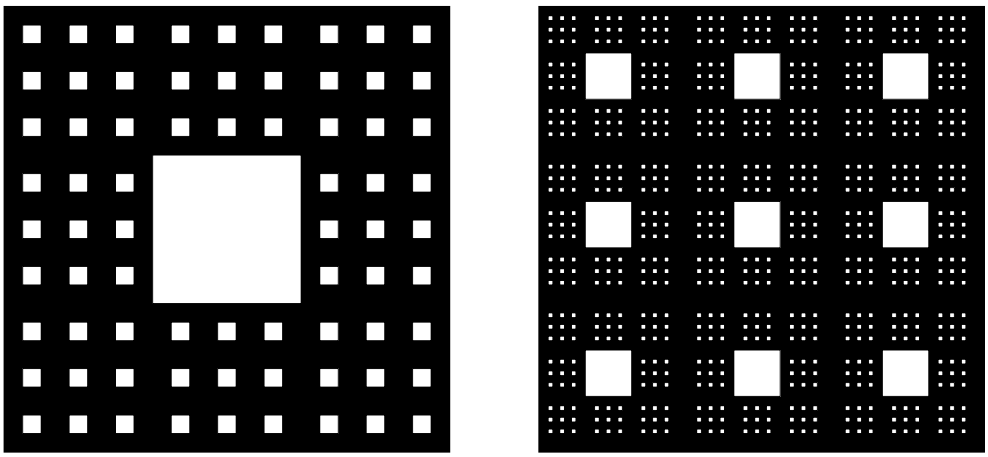


Figure 1.6: Two fractals with the same dimension but different lacunarity (after Mandelbrot [51]). Each fractal is built from $9^2 - 9$ copies of the unit square with an edge of $\frac{1}{9}$. They differ in the positioning of the 9 deleted squares, and this gives them different coarse-scale structure. The one on the left has a large central hole and therefore has high lacunarity. The fractal on the right has a more uniform structure and is therefore of lower lacunarity. Both the fractals have the topology of a Sierpinski carpet.

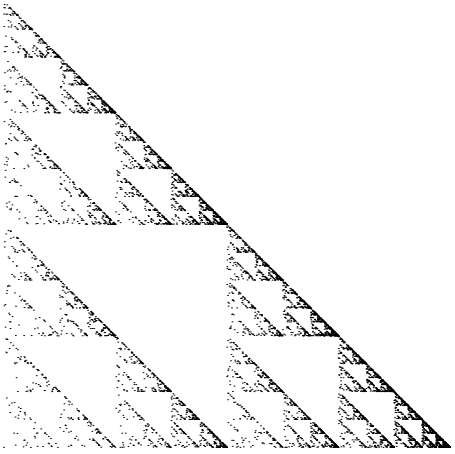


Figure 1.7: A multifractal measure on the Sierpinski triangle. This picture is generated by choosing each of the three functions from the IFS with uneven probabilities. Such a nonuniform distribution of points leads to inaccurate estimations of the box-counting dimension. In our work, the effect is an undesirably high "cutoff resolution;" see Section 2.4 for more discussion. Multifractal theory characterizes the fractal properties of the measure rather than the set that supports the measure.

any Hausdorff dimension. Thus, as many people have observed, the dimension is just one part of a complete characterization of the structure of a fractal. In this section, we review some approaches to distinguishing between two sets that have the same fractal dimension. Each theory addresses a different context, which we illustrate with the examples in Figures 1.5, 1.6, and 1.7.

If two fractals have the same Hausdorff dimension, $\dim_H = s$, then the first step towards telling them apart is to compare their s-dimensional Hausdorff measures, $\mathcal{H}^s(X)$. We have already remarked, however, that Hausdorff dimensions and measures are difficult to work with computationally. In any case, this fails to distinguish between the Sierpinski triangle relatives in Figure 1.5, since the Hausdorff measure is related to the self-similar scaling (see the following section), and this is identical for each of these examples. One can also compare objects like this by determining their topological dimension. This would distinguish the Cantor set relative from the other three topological types in Figure 1.5, for example, but not between the other three fractals which are topologically one-dimensional. In this thesis, we obtain a finer characterization of topological structure by examining the number of connected components and holes as a function of resolution. We elaborate on this when we return to these examples in Chapters 2 and 3.

Mandelbrot's work on *lacunarity* [5, 51] aims to distinguish between fractals with the same dimension and different coarse-scale structure. The problem is illustrated by the examples in Figure 1.6. Lacunarity measures the degree of translational invariance within the fractal and is interpreted as a texture parameter. This has implications for experimental measurements of dimension, since fractals with low lacunarity (i.e., very uniform coarse-scale structure) can appear to fill out a set of positive Lebesgue measure. An approximation to a Cantor set with low lacunarity may therefore appear to be a connected interval at coarse resolutions. A precise mathematical definition of lacunarity is not yet agreed upon; see [2, 5, 51, 76] for more discussion. Our work does not address this issue, since we are interested in the limiting scaling of components or holes, just as the dimension characterizes the limiting scaling of the measure.

It is possible for sets with positive Lebesgue measure to have structure on arbitrarily fine scales; such sets are called *fat fractals*. An example is the chaotic region in Figure 1.2. The Hausdorff or box-counting dimension of a fat fractal is an integer and therefore fails to characterize the fractal nature of the set. Scaling properties of these sets are studied in [20, 22, 30, 80, 84], mostly by examining the rate of convergence of the measure of the ϵ-neighborhoods, X_ϵ, as $\epsilon \to 0$. The resulting fat fractal exponents are interpreted as an exchange index by Tricot [80]. In Chapter 5 we derive some inequalities that relate our topological growth rates to the fat fractal exponents.

We finish this section by mentioning *multifractals*. Multifractal theory stems from the observation that the distribution of points on a set (i.e., a measure) can have fractal properties. This is often the case in dynamical systems, for example, when orbits cover an attractor in a nonuniform manner. The example in Figure 1.7 shows an approximation to a multifractal measure on the Sierpinski triangle. The first step in multifractal analysis is a pointwise localization of the concept of dimension. This is done by analyzing the scaling of the fractal measure, μ, of balls $B_r(x)$ centered at x with radii $r \to 0$:

$$\dim_{\mathrm{loc}} \mu(x) = \lim_{r \to 0} \frac{\log \mu(B_r(x))}{\log r}.$$

One then considers subsets of the fractal that consists of points with identical local dimension. The distribution of the dimensions of these subsets is the multifractal spectrum. See Falconer [23, 24] for further discussion. We consider the effect of nonuniform point distributions on our computational techniques in Section 2.4, but we have not yet attempted to adapt

them to this context.

1.2.3 Iterated function systems

We finish our review of fractal geometry by describing a tool for generating and analyzing fractals with some degree of self-similarity. The concept of an *iterated function system* (IFS) was formalized by Hutchinson [36]. Given a finite collection of functions, $f_i : \mathbb{R}^n \to \mathbb{R}^n$, for $i = 1, \ldots, m \geq 2$, we study sets that are invariant under the joint action of these functions:

$$X = f(X) = \bigcup_{i=1}^{m} f_i(X). \tag{1.6}$$

Examples are the Sierpinski triangle relatives, shown in Figure 1.5. When the f_i are contractions on a closed domain $D \subset \mathbb{R}^n$ (i.e., for $x, y \in D$ there is a number $0 < c_i < 1$ such that $|f_i(x) - f_i(y)| \leq c_i|x - y|$) the following results hold:

1. There is a unique closed, bounded set satisfying (1.6).

2. This set is the closure of the set of fixed points of arbitrary finite compositions, $f_{i_1} \circ \cdots \circ f_{i_k}$, with $i_j \in \{1, \ldots, m\}$.

3. Given any set $A \subset D$, then $f^k(A) \to X$ in the Hausdorff metric as $k \to \infty$.

Proofs of the above are based on the contraction mapping theorem; see [23] or [36]. Further properties of iterated function systems are explored in detail in Barnsley [4]. Many of the examples we use throughout this thesis are generated by iterated function systems, e.g., the Sierpinski triangle relatives in Figure 1.5.

One property we make use of in our numerical work is that IFS attractors are perfect, i.e., they have no isolated points. This follows from result 2 above. We must show that every point, $x \in X$, is the limit of a sequence of other points in the IFS attractor. From 2, x is either a fixed point of a finite number of compositions or in the closure of these points. In the latter case, x is (by definition) the limit of a sequence of points from X. For the other case, let $g = f_{i_1} \circ \cdots \circ f_{i_k}$ and suppose $x = g(x)$. Now consider the fixed point of one of the IFS functions, $y = f_{i_j}(y)$. Again, from result 2 we know $y \in X$; since there is more than one function in the IFS, we can assume that $y \neq x$. Now let $y_k = g^k(y)$ (g^k is the k-fold composition of g). Since $X = f(X)$, we know $y_k \in X$ for all k; the contraction mapping principle implies $y_k \to x$, so we are done.

The question of whether an IFS attractor is connected or disconnected is studied by Barnsley [4] in the context of generating Mandelbrot sets for parameterized families of iterated function systems. An image based algorithm for studying the connectedness of IFS attractors is presented in [7]. Although we use iterated function system attractors as examples in this thesis, our algorithms for determining connectedness are designed to apply in a much broader context.

Similarity dimension

Many simple fractals, such as the Sierpinski triangle and most of the examples in Chapter 2, are attractors for iterated function systems of similarities. This means each function, f_i, satisfies

$$|f_i(x) - f_i(y)| = c_i|x - y| \quad \text{for all} \quad x, y \tag{1.7}$$

where $0 < c_i < 1$ is the contraction or similarity ratio. Suppose X is the invariant set for a family of m similarities:

$$X = \bigcup_{i=1}^{m} f_i(X). \tag{1.8}$$

There is a very simple definition of dimension which is easy to compute — the similarity dimension, $\dim_S$, which is the number s that makes the following hold:

$$\sum_{i=1}^{m} c_i^s = 1. \tag{1.9}$$

In general, $\dim_H(X) \leq \dim_S(X)$. The Hausdorff and similarity dimensions are equivalent when the IFS satisfies the *open set condition.* An IFS of similarities satisfies the open set condition if there exists a nonempty bounded open set, V, such that

$$V \supset \bigcup_{i=1}^{m} f_i(V), \tag{1.10}$$

with the union disjoint. For example, the Sierpinski triangle relatives satisfy the open set condition with V as the open unit square. See Falconer [23] for a proof that Hausdorff, box-counting, and similarity dimensions agree under this condition.

1.3 Computational topology

By computational topology, we mean the study of topological properties of an object that can be computed to some finite accuracy. There is a growing literature on the formalization and representation of topological questions for computer applications, and on the study of appropriate algorithms; see [13] for a survey of the field. Application areas include digital image processing, topology-preserving morphing in computer graphics, solid modelling for computer aided design, mesh generation for finite elements, 3-d models of protein molecules, and the analysis of experimental time-series data. Other distinctly different fields that combine topology and computer science include topological techniques in the theory of computing and computer visualization of complicated topological spaces.

The earliest work on extracting topological information from data targeted digital images. These are typically represented by binary data on a fixed regular grid in two or three dimensions, e.g., pixels and voxels. This field has many applications including algorithmic pattern recognition, which plays an important role in computer vision (e.g., determining whether a robot-width corridor exists between two obstacles [6]), and remote sensing (e.g., computing the boundaries of a drainage basin from satellite data [88]). The fundamental concept in this field is that of adjacency, the definition of which depends upon the grid structure. Much work in this area focuses on algorithms for the labeling of components [40], boundaries [83], and other features of digital images. Basic results include consistent notions for connectedness [40], simple connectedness [33], a digital Jordan curve theorem [74], and algorithms for the Euler characteristic of digital sets [39, 45].

The data we are interested in analyzing are typically finite sets of points from a finite-dimensional metric space. Existing work on extracting topological information from this type of data includes a number of approaches to computational homology. The first step in computing homology from point-sets is to build a triangulation or other regular cell complex that

reflects the topology of the data. Once this is done, it is possible — though costly — to compute representations of the homology groups from the complex. It is much faster to find only the ranks of the groups — the *Betti numbers* — and often this is enough information for applications. Fast[2] algorithms for computing Betti numbers take many forms. Friedman [26] uses an isomorphism between homology groups and the null space of a combinatorial Laplacian to compute Betti numbers from an arbitrary simplicial complex. In Chapter 3, we use a multiresolution approach to building simplicial complexes called *alpha shapes* which is due to Edelsbrunner *et al.* [17, 18]. For subsets of $\mathbb{R}^2$ and $\mathbb{R}^3$ there is a fast incremental algorithm for computing Betti numbers from alpha shapes [11]. The implementations use fundamental constructions from computational geometry such as Delaunay triangulations and Voronoi diagrams; we describe these in more detail in Section 3.4.1. To the best of our knowledge, Edelsbrunner's approach is the only existing algorithm for computing Betti numbers at multiple resolutions. However, previous work with alpha shapes has not formally investigated the problem of extrapolating information about the underlying space from the finite data — the topic of this thesis. A drawback of the alpha shape implementation for our applications is that it is not suited to the large data sets typically encountered in dynamical systems applications.

1.4 Overview of the thesis

1.4.1 The basic assumptions

In both experimental and simulated dynamical systems, the data to be analyzed are typically points along a trajectory. These points approximate the omega-limit set of the orbit, which in turn may approximate an attractor or other invariant set. We abstract this setting by assuming the underlying set, X, is a compact subset of a metric space, and that the data, S, are a finite set of points that approximate X. We measure the accuracy of the approximation using the Hausdorff metric,

$$d_H(S, X) = \min\{\epsilon \mid X \subset S_\epsilon \text{ and } S \subset X_\epsilon\}.$$

The notation S_ϵ represents the closed ϵ-neighborhood:

$$S_\epsilon = \{x \mid d(x, S) \leq \epsilon\}.$$

Thus, if $\rho = d_H(S, X)$, then every point of S is within a distance ρ of some point in X, and vice versa.

Our goal is to extract information about the topology of X from the finite approximation, S. We keep our approach as general as possible within the above context; the only requirement on X is that it must be compact. The definition of compactness — that given any covering of a space by open sets, it is possible to cover the space using only a finite number of those sets — implies that approximating a compact set by a finite set of points is not unreasonable. In the applications we describe above, compactness is a valid assumption since the omega-limit set of an orbit is compact if it is bounded.

A finite set of points has no intrinsic topological structure, so it must be coarse-grained in some manner. To give the data non-trivial structure, we form the closed ϵ-neighborhood S_ϵ, as defined above. Our basic approach is to determine topological properties of the fattened set at different values of ϵ tending to zero. The idea is that the topological structure of X can

[2] I.e., time costs that are subquadratic in the number of points.

be extrapolated from that of the ϵ-neighborhoods of S. Naturally, this extrapolation is always constrained by the inherent accuracy of the data. We develop a criterion that identifies a cutoff resolution from the data; this provides a measure of confidence in the results.

The first step in justifying our approach is to examine its validity for general spaces. For example, a space that is "connected at resolution ϵ" for all $\epsilon > 0$ is only guaranteed to be connected in the usual sense if it is compact. The second step is to formally relate the topology of the ϵ-neighborhoods of S to that of the ϵ-neighborhoods of X. We do this in Chapter 3 using the assumption that the data and the underlying space are close in the Hausdorff metric. These sound mathematical foundations form the most significant contribution of the thesis to the emerging field of computational topology.

1.4.2 Organization of the thesis

The thesis has five chapters. The first two cover theoretical and computational results for our multiresolution approach to topology. We then describe some example applications in dynamical systems and finish with a chapter that derives inequalities involving our topological growth rates and various definitions of fractal dimension.

We begin, in Chapter 2, by investigating the most elementary properties of a space: the number and size of its connected components. Computing these quantities at multiple resolutions allows us to determine whether the data approximate a connected, totally disconnected, and/or perfect space. The idea of formulating connectedness using a resolution parameter ϵ goes back to Cantor's definition for connectedness in compact metric spaces. (In this chapter, ϵ is a distance between points, not the fattening by ϵ that we described earlier.) We introduce three functions, $C(\epsilon)$, $D(\epsilon)$, and $I(\epsilon)$, that are the number of components, largest component diameter, and number of isolated points, respectively. From the limiting behavior of these three functions, we are able to determine the connectedness properties of a compact space. For finite sets of points, we compute these quantities from the *minimal spanning tree* and the *nearest neighbor graph*. We show that the minimal spanning tree is the ideal data structure for describing connected components at multiple resolutions. For Cantor sets and other disconnected fractals, the number of ϵ-components goes to infinity as ϵ tends to zero, and for totally disconnected sets, the diameters go to zero. We characterize the rates of growth using a power law and compute the corresponding *disconnectedness* and *discreteness indices* for a number of examples.

In Chapter 3 we address the more challenging problem of computational homology. The definitions for homology theory are quite involved, so we give an overview of the basic concepts in Section 3.2. Homology quantifies structure via the Betti numbers, β_k, which essentially count the number of k-dimensional holes in a space. Our initial plan was to compute the Betti numbers of the ϵ-neighborhoods; we hoped that the limit as $\epsilon \to 0$ would give the Betti number of the underlying space. The process is more subtle than this, however, because fattening a set to its ϵ-neighborhood can actually introduce new holes. We resolve this problem using an inverse system approach from shape theory. This allows us to define the *persistent Betti numbers*, which count holes in an ϵ-neighborhood that correspond to a hole in the underlying space. The computer implementations of these ideas are not finalized. Instead, we use existing alpha shape software to analyze some simple examples, and illustrate why the regular Betti numbers are inadequate.

Chapter 4 explores applications of these techniques in dynamical systems. The examples we study have well-understood structure, which enables us to evaluate the usefulness of our techniques. The most extensive study of this chapter is the breakup of invariant circles in area-

preserving twist maps. We end by suggesting some other applications of our computational topology techniques to some open question in dynamics.

Finally, in Chapter 5 we explore connections between our topological growth rates and various definitions of fractal dimension. The results we give are far from complete and we outline some potential avenues for further work.

Chapter 2

Computing Connectedness

2.1 Introduction

This chapter examines some elementary concepts from point-set topology — specifically connectedness, total disconnectedness and perfectness. The goal is to obtain computational tools that allow us to determine the number and size of connected components of a data set at any given resolution.

The first part of the chapter reformulates the classical definitions in terms of a resolution parameter, ϵ. We then show that for a compact space, X, the topological properties of connectedness, total disconnectedness, and perfectness can be deduced by examining the limiting behavior of the number, $C(\epsilon)$, and size, $D(\epsilon)$, of ϵ-connected components as $\epsilon \to 0$. We characterize the limiting behavior of these two quantities by a power law, and compute the corresponding disconnectedness and discreteness indices.

In Section 2.3, we give a new algorithm based on the minimal spanning tree (MST) that implements these ideas numerically for arbitrary finite point-set data. We show that the MST is an ideal data structure for representing ϵ-components of a finite point-set. Essentially, this is because the MST always joins two subsets by the smallest possible edge. An important step is to determine a cutoff resolution, ρ, so that the computed results are a good representation of the true space for $\epsilon > \rho$. When the underlying space is perfect, ρ is well approximated by the resolution at which the data first appears to have an isolated point.

Finally, in Section 2.4, we demonstrate the effectiveness of our techniques by applying them to a variety of examples. We present data that exhibit different types of scaling in the number and size of their ϵ-components. We also investigate the dependence of the cutoff resolution ρ on the number of data points and on the uniformity of their distribution over the attractor. The first set of examples are fractals generated by closely related iterated function systems. Each has a distinct topology but all have the same Hausdorff dimension. We show that the disconnectedness and discreteness indices classify the sets according to their topology. We next analyze five Cantor sets to demonstrate different types of scaling that can occur in the functions $C(\epsilon)$ and $D(\epsilon)$ as $\epsilon \to 0$. These examples lead us to conjecture that Cantor sets with zero Lebesgue measure have disconnectedness index equal to their box-counting dimension. Results that prove some special cases of this conjecture are given in Chapter 5.

The material in this Chapter has been published in [71, 72].

2.2 Foundations for computing connectedness

2.2.1 Concepts from point-set topology

Connectedness is a very intuitive concept that captures the notion of continuity of a set of points in a topological way. The mathematical formulation goes back a little over one hundred years and the following definition can be traced to Jordan [42]. A topological space X is *connected* if it cannot be decomposed into the union of two non-empty, disjoint, closed sets. If such a decomposition exists then X is said to be *disconnected* — that is, if there are two closed sets U and V such that $U \cap V = \emptyset$ and $U \cup V = X$.

This definition does not lend itself to computation and in the following section we will give a definition that is equivalent to an early formulation of connectedness in metric spaces due to Cantor [35]. Cantor's definition uses ϵ*-chains*: a finite sequence of points $x_0, \ldots, x_n$ that satisfy $d(x_i, x_{i+1}) < \epsilon$ for $i = 1, \ldots, n$. A set, X, is *Cantor-connected* when every two points in X can be linked by an ϵ-chain for arbitrarily small ϵ. This definition agrees with the one above only in the special case of compact metric spaces. For example, the rational numbers are Cantor-connected but disconnected in the regular sense.

One object we are particularly interested in is the *connected component* of a point. Given $x \in X$, this is the largest connected subset of X containing x. For example, if $X = [0,1] \cup [2,3]$ then the connected component of $\frac{1}{2}$ is $[0,1]$. If the connected component of every point is only the point itself then the set is said to be *totally disconnected.* The rationals are totally disconnected, as is the middle-thirds Cantor set.

Another concept from point-set topology is the property of perfectness. This means every point has arbitrarily small neighborhoods containing infinitely many other points, so that no point is isolated. Formally a set is *perfect* if it is equal to the set of its accumulation points. Notice that this definition implies that only closed sets can be perfect.

Any compact metric space that is totally disconnected and perfect is homeomorphic to the middle-thirds Cantor set (see [35] for a proof). The combination of these properties may seem somewhat paradoxical, since they tell us that each point is isolated in one sense — its connected component is that single point, and yet no point is isolated, since each must be the limit of some sequence of points in the Cantor set.

In the following section we reformulate these definitions in a way that relies on extrapolation, making it possible to implement the ideas numerically. The basic approach is to look at the set with a finite resolution ϵ, see how certain properties change as $\epsilon \to 0$, and infer information about the topology.

2.2.2 ϵ-Resolution definitions

Given a compact subset X of a metric space, we say it is ϵ*-disconnected* if it can be written as the union of two sets that are separated by a distance of at least ϵ — i.e., there are two closed subsets, U and V with $U \cup V = X$ and $d(U,V) \equiv \inf_{x \in U, y \in V} d(x,y) \geq \epsilon$. Otherwise, X is ϵ*-connected.* As with Cantor's definition, a set that is ϵ-connected for any $\epsilon > 0$ is connected if and only if it is compact. This follows from the simple lemma below.

Lemma 2. *If a compact metric space X is disconnected, then it is ϵ-disconnected for some $\epsilon > 0$.*

Proof. Since X is disconnected, there are closed disjoint sets U and V such that $U \cup V = X$. Now suppose $d(U,V) = 0$. This implies that there are sequences $x_n \in U$, $y_n \in V$ with

$d(x_n, y_n) \to 0$. Since U and V are compact, there must be convergent subsequences $x_{n_i} \to x^*$ and $y_{n_j} \to y^*$ with $d(x^*, y^*) = 0$. But this implies $x^* = y^*$, which is impossible since U and V are disjoint, so there must be an $\epsilon > 0$ such that $d(U, V) = \epsilon$. □

Restricting our attention to compact sets is not unreasonable, since we are primarily interested in sets that are well approximated on a computer as finite point-sets. In the context of dynamical systems, we are interested in attractors and other invariant sets. These are closed subsets of metric spaces, and are therefore compact when they are bounded.

We now make an ϵ-resolution definition of connected component. A subset $A \subset X$ is an *ϵ-component* if A is ϵ-connected and $d(A, X \setminus A) \geq \epsilon$. Given a resolution, ϵ, X has a natural decomposition as the disjoint union of its ϵ-components. We can exploit this decomposition to deduce topological properties of the set. For example, if the only ϵ-component is X itself for all ϵ, then we can conclude that X is connected. We also know that at any fixed resolution $\epsilon > 0$, a compact set has a finite number of ϵ-components. To see this, take a covering of X by ϵ-balls. Since X is compact, there is a finite sub-cover. By their definition, every ϵ-component must contain at least one ϵ-ball$\cap X$, so there is a finite number of ϵ-components. This motivates the introduction of a function $C(\epsilon)$ that counts the number of ϵ-components at resolution ϵ. This function is monotonic: if $\epsilon_1 < \epsilon_2$ then $C(\epsilon_1) \geq C(\epsilon_2)$.

By looking at the size of the ϵ-components we can deduce the properties of total disconnectedness and perfectness. There are a number of ways to measure the size of a set; the quantity used depends on the context. We use the diameter, since this is defined in any metric space. Recall that the diameter of a set A is the largest distance between any two points in the set: $\mathrm{diam}(A) = \sup_{x,y \in A} d(x, y)$. At a resolution ϵ, let $\mathcal{D}(\epsilon)$ represent the set of diameter measurements of the ϵ-components. For notational convenience, we write $D(\epsilon) = \max \mathcal{D}(\epsilon)$ for the function that describes how the largest diameter changes with resolution. $D(\epsilon)$ is a monotonic non-increasing non-negative function, so the limit as $\epsilon \to 0$ must always exist.

From the definition of total disconnectedness, we have the following result.

Lemma 3. *A compact set X is totally disconnected if and only if* $\lim_{\epsilon \to 0} D(\epsilon) = 0$.

Proof. By way of obtaining a contradiction to the forward direction, suppose that

$$\lim_{\epsilon \to 0} D(\epsilon) = \delta > 0.$$

Take any sequence, $\epsilon_n \to 0$ and construct a tree as follows. At level n list all the ϵ_n-components with diameter $\geq \delta$. There are a finite number of these. Order the tree by set inclusion, i.e. an edge connects A_{ϵ_j} and $A_{\epsilon_{j+1}}$ if and only if $A_{\epsilon_{j+1}} \subset A_{\epsilon_j}$. We know that there must be ϵ-components with diameters $\geq \delta$ for all ϵ, so this tree must have an infinite branch. This gives a sequence of nested components A_{ϵ_n}, with $\mathrm{diam}(A_{\epsilon_n}) \searrow \delta$. Since the sets are nested, they have a limit, $A_* = \lim_{n \to \infty} A_{\epsilon_n} = \bigcap_n A_{\epsilon_n}$. It then follows that A_* is ϵ-connected for all ϵ and $\mathrm{diam}(A_*) = \delta$. This implies A_* has at least two points in it and so X could not be totally disconnected.

The converse follows directly from the definition. If the diameters of the ϵ-components go to zero, then the connected component of any $x \in X$ is just $\{x\}$, so X is totally disconnected. □

In ϵ-resolution terms, a compact set X is perfect if and only if $\min \mathcal{D}(\epsilon) > 0$ for all $\epsilon > 0$. Alternatively, we can look for isolated points, since these are easy to detect numerically. We say a point $x \in X$ is isolated at resolution ϵ if $d(x, X - x) \geq \epsilon$. For a set to be perfect, the number of ϵ-isolated points, $I(\epsilon)$, must be zero for any $\epsilon > 0$.

The three quantities $C(\epsilon)$, $D(\epsilon)$, and $I(\epsilon)$ form the basis of our computational approach to determining the connectedness properties of data.

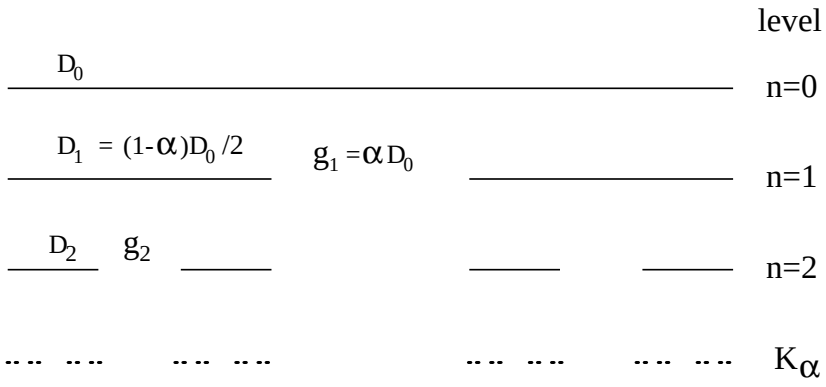


Figure 2.1: The construction of a K_α Cantor set.

2.2.3 Disconnectedness and discreteness growth rates

In some situations of interest — Cantor sets, for example — it is expected that $C(\epsilon) \to \infty$ and $D(\epsilon) \to 0$ as $\epsilon \to 0$. An effective way to evaluate this behavior is to assume an asymptotic form and compute appropriate indices. We use a simple polynomial power law, although finer scales of functions exist [81]. Thus, near $\epsilon = 0$, we look for real numbers γ and δ such that $C(\epsilon) \sim \epsilon^{-\gamma}$ and $D(\epsilon) \sim \epsilon^{\delta}$. The exponents may be found as the following limits:

$$\gamma = \lim_{\epsilon \to 0} \frac{\log C(\epsilon)}{\log(1/\epsilon)} \tag{2.1}$$

$$\delta = \lim_{\epsilon \to 0} \frac{\log D(\epsilon)}{\log \epsilon}. \tag{2.2}$$

If the limits do not exist we can compute the $\liminf$ and the $\limsup$ of each quantity.

The component growth rate, γ, is called the *disconnectedness* index. A positive value of γ implies that the set has infinitely many components. We call δ the *discreteness* index. If δ is positive, the set must be totally disconnected. The function $D(\epsilon)$ relates the size of the ϵ-components to the distance between them, so δ measures the relative rate of decrease of component and gap sizes. If the set is connected or has a finite number of components, then γ and δ are both zero. Three simple examples of Cantor subsets of $\mathbb{R}$ are given below.

Middle-α Cantor sets

These Cantor sets arise in piecewise-linear one-dimensional maps. Let $0 < \alpha < 1$ and consider the Cantor set $K_\alpha \subset [0,1]$ constructed by successively removing the middle α-proportion of each remaining interval. This construction has a natural correspondence with the ϵ-components. At a given level n, there are $C_n = 2^n$ intervals of equal length $D_n = \frac{1}{2}(1-\alpha)D_{n-1}$, separated by gaps of at least $g_n = \alpha D_{n-1}$; see Figure 2.1. With $D_0 = 1$, the recursion relations may be solved to find $D_n = [\frac{1}{2}(1-\alpha)]^n$ and $g_n = \alpha[\frac{1}{2}(1-\alpha)]^{n-1}$, so that

$$\delta = \lim_{n\to\infty} \frac{\log D_n}{\log g_n} = \lim_{n\to\infty} \frac{n \log[\frac{1}{2}(1-\alpha)]}{(n-1)\log[\frac{1}{2}(1-\alpha)] + \log\alpha} = 1 \ ,$$

and

$$\gamma = \lim_{n\to\infty} -\frac{\log C_n}{\log g_n} = \lim_{n\to\infty} \frac{-n \log 2}{(n-1)\log[\frac{1}{2}(1-\alpha)] + \log\alpha} = \frac{\log 2}{\log 2 - \log(1-\alpha)} \ .$$

The discreteness index, δ, is independent of α because the Cantor set is constructed in such a way that the ϵ-components and gaps decrease at the same rate. The disconnectedness index, γ, has the same value as the Hausdorff dimension $\dim_H$. Since $\dim_H < 1$, it follows that these Cantor sets have zero Lebesgue measure. This also follows from the fact that the sum of the gap-lengths is equal to one.

A Cantor set with positive measure

Now consider a Cantor set with gaps that decrease more rapidly. Let K be the subset of $[0,1]$ obtained by successively removing gaps from the center of remaining intervals, with widths $g_n = (\frac{1}{2})^{2n-1}(\frac{1}{10})$ for $n = 1, 2, \ldots$. The total Lebesgue measure of the gaps is just $\frac{1}{10}$, so the measure of K is $\frac{9}{10}$. After a bit of algebra, we find that for $(\frac{1}{2})^{2n+1}(\frac{1}{10}) \le \epsilon < (\frac{1}{2})^{2n-1}(\frac{1}{10})$, there are $C_n = 2^n$ components with diameters $D_n = (\frac{1}{2})^n[\frac{9}{10} + (\frac{1}{2})^n(\frac{1}{10})]$. So,

$$\delta = \lim_{n\to\infty} \frac{n\log(\frac{1}{2}) + \log[\frac{9}{10} + (\frac{1}{2})^n(\frac{1}{10})]}{(2n-1)\log(\frac{1}{2}) + \log(\frac{1}{10})} = \frac{1}{2} \ ,$$

and

$$\gamma = \lim_{n\to\infty} \frac{-n\log 2}{(2n-1)\log(\frac{1}{2}) + \log(\frac{1}{10})} = \frac{1}{2} \ .$$

Since K has positive Lebesgue measure, its Hausdorff dimension is 1. Therefore, the disconnectedness index and the dimension are distinct. In fact, in Chapter 5 we show that for subsets of $\mathbb{R}$, the disconnectedness index is equivalent to the fat fractal exponent.

Cantor sets with zero Hausdorff dimension

It is possible for a Cantor set to have $\dim_H = 0$. Such Cantor sets are observed as invariant sets for the symplectic twist maps we study in Chapter 4. As an example, let $0 < \lambda < 1$ and $c = (1-\lambda)/\lambda$, and suppose the Cantor set has a single gap $g_n = c\lambda^n$ for each $n = 1, 2, \ldots$. We construct the Cantor set as a subset of $[0,1]$, so the constant c is chosen the make the sum

$$\sum_{i=1}^{\infty} g_i = \sum_{i=1}^{\infty} c\lambda^i = 1.$$

This means the Lebesgue measure of the Cantor set is zero. By setting $\epsilon = g_n$, we have that $C(g_n) = n+1$ and therefore

$$\gamma = \lim_{n\to\infty} -\frac{\log(n+1)}{\log g_n} = \lim_{n\to\infty} -\frac{\log(n+1)}{n\log\lambda + \log c} = 0.$$

The diameters of the ϵ-components depend on the exact positioning of the gaps in $[0,1]$. However, if $\epsilon = g_n$, then the largest remaining ϵ-component must be longer than the next gap to be resolved, and its length cannot exceed the sum of all the remaining gaps. That is,

$$g_{n+1} \le D(g_n) \le 1 - \sum_{i=1}^{n} g_i.$$

It follows that

$$\frac{\log g_{n+1}}{\log g_n} \ge \frac{\log D(g_n)}{\log g_n} \ge \frac{\log(1 - \sum_{i=1}^{n} g_i)}{\log g_n}.$$

Now,

$$1 - \sum_{i=1}^{n} g_i = 1 - \sum_{i=1}^{n} c\lambda^i = 1 - \frac{c\lambda}{1-\lambda}(1-\lambda^{n+1}) = \lambda^{n+1}.$$

Thus,

$$\frac{\log c\lambda^{n+1}}{\log c\lambda^n} \geq \frac{\log D(g_n)}{\log g_n} \geq \frac{\log \lambda^{n+1}}{\log c\lambda^n}.$$

Taking the limit as $n \to \infty$ in each expression, we find that $\delta = 1$.

In Section 2.4, we examine some Cantor subsets of $\mathbb{R}^2$. For many of these examples we find that γ is approximately equal to the box-counting dimension, $\dim_B$, and that δ is close to one. We give formal results relating fractal dimensions and the disconnectedness and discreteness indices in Chapter 5.

2.3 Implementation

The theory outlined in the previous section applies to arbitrary compact sets. Recall that compactness is the property that any covering of the set by open subsets can also be accomplished by a finite number of those subsets. This is why it is possible to represent a compact set by a finite number of points. One way to do this is to take a covering of the set by open balls, choose a finite sub-cover, and let the centers of the balls be the finite point-set approximation. Suppose each point in the compact set, X, is within a maximum distance $\rho/2$ of some point in the approximating set, S. The original set and its finite point-set approximation therefore exhibit essentially the same $C(\epsilon)$, $D(\epsilon)$, and $I(\epsilon)$ behavior when $\epsilon > \rho$. For example, if the original set, X, is connected and perfect, then examining S with resolution $\epsilon > \rho$ we see $C_S(\epsilon) = 1$, $\mathrm{diam}(X) - \rho \leq D_S(\epsilon) \leq \mathrm{diam}(X) + \rho$, and $I_S(\epsilon) = 0$. For $\epsilon < \rho$, S has many components and the majority of these are isolated points. For general compact sets we have the bound $C_S(\epsilon - \rho) \geq C_X(\epsilon) \geq C_S(\epsilon + \rho)$ for $\epsilon > \rho$. These ideas are developed further in Chapter 3.

In practice, we are given only the finite point-set approximation and wish to determine the connectedness properties of the underlying set. To do this, we need to estimate the minimum resolution, ρ, from the point-set itself. We can then attempt to extrapolate the limiting behavior of $C(\epsilon)$ and $D(\epsilon)$ from the data with $\epsilon > \rho$. The question that naturally arises is: how confident can we be that the limiting topology is that implied by the data for $\epsilon > \rho$? It is possible to construct examples that appear to be connected down to a given resolution, but are in fact Cantor sets. Conversely, there are connected sets whose finite point-set approximations may appear to be totally disconnected. Both these problems can be addressed to some extent by checking the effect on $C(\epsilon), D(\epsilon)$ and $I(\epsilon)$ of increasing the number of points approximating the underlying set. Ultimately, though, we are restricted by machine precision.

In order to compute $C(\epsilon)$, $D(\epsilon)$ and $I(\epsilon)$ numerically, we need an appropriate way to organize the finite point-set data. The structure we use is a graph — the Minimal Spanning Tree (MST) [67]. This choice was inspired by K. Yip's work on computer recognition of orbit structures in two dimensional area-preserving maps [89]. The following section describes why the edge lengths of the MST naturally define the resolutions at which one should see a change in the number of components. In fact, the MST of a data set contains all the information we need about the ϵ-components.

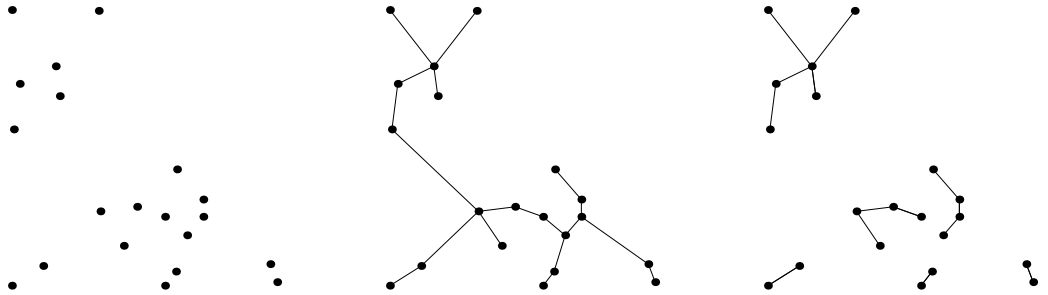


Figure 2.2: A finite set of points (a), its minimal spanning tree (b), and the nearest neighbor graph (c). The weight of an edge is the Euclidean distance between the points it joins.

2.3.1 Minimal spanning trees

A *graph* is a structure consisting of a finite set of points called *vertices* and a list of pairs of these points called *edges*. A *weighted graph* has a weight or cost assigned to each edge. A graph is *connected* if there is a sequence of edges (a *path*) joining any point to any other point. When it contains no closed paths, a connected graph is called a *tree*. Given a graph, a *spanning tree* is a subgraph that is a tree and contains all the vertices from the original graph. For our purposes, the vertices are data points in a metric space (usually $\mathbb{R}^n$), the edges are lines joining two points, and the weight of an edge is just the metric distance between the two points joined by that edge.

The *minimal spanning tree* (MST) of a graph is a spanning tree of that graph that has minimal total weight. It is unique when all the edge weights are distinct. For our application, we build the MST from the complete graph — i.e., one that has an edge between every pair of points. The intuitive way to do this is to delete edges from the complete graph as follows. Start with the longest edge (largest weight), and then remove successively shorter ones, provided that doing so leaves the sub-graph connected. The algorithm ceases when removing any edge would leave a disconnected graph. An alternative constructive algorithm, due to R.S. Prim [68], is more readily implemented on a computer. The initial subtree consists of any point, its nearest neighbor, and the edge between them. The subtree grows by adding the point that has the shortest distance to some point (called its parent) in the subtree. This step is repeated until all points are in the tree. The cost of this algorithm is $O(n^2)$ where n is the number of points in the set. It is also possible to construct the MST from a special graph called the Delaunay triangulation (c.f., Section 3.4). For sets in the plane, this results in an algorithm with cost $O(n \log n)$. See [67] for details.

The property of connecting closest points makes the MST a natural structure for our analysis. This is illustrated by the data set and MST presented in Figure 2.2. Notice that the MST bridges the "gap" between two subsets with the shortest edge possible, automatically organizing the point-set in a way that corresponds closely to ϵ-connected components of the set. The following results formalize this intuition. More results about minimal spanning trees and their relationship to clustering can be found in [90].

First, we show how the edge lengths in the MST relate to ϵ-connectedness and the ϵ-components.

Lemma 4. *If S is a finite set of points that is ϵ-connected for all $\epsilon \geq \epsilon_0 > 0$, then the longest edge in the MST has length $l < \epsilon_0$.*

Conversely, if the MST of a finite set of points, S, has largest edge length $l < \epsilon_0$, then S is ϵ-connected for all $\epsilon \geq \epsilon_0$.

Proof. Recall from Section 2.2 that S is ϵ-connected if and only if there is an ϵ-chain joining every pair of points in S. This implies that in the complete graph over the points of S, it is possible to move from one vertex to any other along edges of length less than ϵ. Therefore, in the intuitive construction of the MST, all edges of length greater than ϵ_0 are removed, since they could not disconnect the graph. It follows that the MST of S can have no edge longer than ϵ_0.

Conversely, the MST is a connected graph, so given any points $x, y \in S$, we can find an ϵ_0-chain: $x_0 = x, \ldots, x_n = y$, with $d(x_i, x_{i+1}) < \epsilon_0$. This implies that S is ϵ-connected for all $\epsilon \geq \epsilon_0$. □

The next result shows that the MST is a suitable structure for representing ϵ-components. First, we note that every edge in a MST defines a partition of S (the point set) into two subsets P and Q containing the points from the two subtrees generated by removing the edge. It follows that:

Lemma 5. *Removing an edge from a MST generates two subgraphs, each of which is a MST of its points.*

Suppose that the longest edge has length l and assume that this edge length is unique. If we remove this edge, we are left with two subsets P and Q that are ϵ-connected for some $\epsilon < l$. These subsets genuinely are ϵ-components since $d(P, Q) = l > \epsilon$. If there are $n > 1$ edges of length l then removing them leaves $n + 1$ connected components.

The minimal spanning tree, once constructed, holds all the information we need to deduce connectedness properties from ϵ-components. The ϵ-components of the set S correspond directly to the connected components of the sub-graph that is generated by removing edges from the MST of lengths, $l \geq \epsilon$. Also, the edge lengths of the MST automatically give the resolutions at which one sees a change in the number of components. We now discuss ways to extract this information efficiently from the MST.

2.3.2 Practical issues

This section addresses some of the details of the numerical implementation. In particular, it describes how we compute $C(\epsilon)$, $D(\epsilon)$, and $I(\epsilon)$. We give Matlab code for the following algorithms in an appendix to this thesis.

Finding the number of ϵ-components is straightforward: $C(\epsilon)$ is just one more than the number of edges with length greater than ϵ. We build the MST using Prim's algorithm with the Euclidean distance between points as the edge weight, and store the MST as three arrays. The first lists the coordinates of the data points in arbitrary order, the second contains the index of the parent of each point, and the third, the length of the edge between the point and its parent. This allows us to compute $C(\epsilon)$ from the cumulative distribution of edge-lengths, obtained by simply sorting the array.

We know that a point will be isolated at resolution ϵ if all the MST edges incident on that point are longer than ϵ. One way to compute $I(\epsilon)$ is to delete all edges longer than ϵ from the MST and then count the number of isolated points. Another method is to use the *nearest neighbor graph* (NNG); see Figure 2.2 for an example. The nodes of this directed graph are again the data points; an edge points from one node, x_i, to another, x_j, if

$$d(x_i, x_j) \leq d(x_i, x) \quad \text{for all} \quad x \in S \setminus \{x_i, x_j\}. \tag{2.3}$$

This relationship is not reflexive, i.e., $x_i \to x_j$ does not imply $x_j \to x_i$. Ignoring their directions, the edges of the NNG are a subset of those in the MST. It is therefore very simple to build the NNG from the MST, again by sorting the MST edges by length and then recording the shortest edge incident on each point. The number of ϵ-isolated points, $I(\epsilon)$, is just the number of edges in the NNG that are longer than ϵ (counting an edge that points both ways twice). $I(\epsilon)$ is then a cumulative distribution of edge-lengths in the NNG.

To find the diameter of an ϵ-component we first need to list the points in the ϵ-component. This involves a tree walk algorithm, which we will discuss in the following three paragraphs. Finding the Euclidean metric diameter of a set of n points in the plane is an order $O(n \log n)$ algorithm, since the computations can be restricted to points lying on the boundary of the convex hull. For subsets of higher-dimensional spaces, this restriction does not necessarily help [67]. Instead, the algorithm we use is the brute-force comparison of distances between all pairs of points, which is $O(n^2)$. Finding the diameter using the supremum metric ($d_{\sup}(x, y) = \sup_i |x_i - y_i|$) is faster, since all that is required is to find the maximum and minimum value in each coordinate over the set of points, and this is linear in the number of points.

A tree walk on the MST is not particularly efficient since its root is arbitrary and little can be said in general about its branching structure. There is, however, a natural binary structure to the MST, since each edge defines a partition of the MST into two components. This fact can be used to construct a binary tree from the MST, which is then faster to search.

The binary tree represents information about the MST in the following way. Nodes (vertices) of the binary tree represent edges from the MST, ordered by length; the root is the longest edge. The two children of an edge node are the longest edges of the two sub-MSTs generated by removing that edge. The leaves (nodes with no children) of the binary tree represent the data points; the parent node of a leaf is the shortest edge incident to that point in the MST. In essence, each edge node in the binary tree represents a connected component of the MST. Given a value of ϵ, each ϵ-connected component is represented by an edge-node with length less than ϵ, but whose parent has length greater than ϵ. The points in an ϵ-component can be found by listing the leaves "under" its representative node. This binary tree can be built so that each edge-node has information about the component it represents — for example, the number of points in the component.

The time to list the points in a component is proportional to the depth of the tree, which in turn depends on the number of points in the data set and how well the tree is balanced. For fairly uniformly distributed data, the tree is well balanced, but for non-uniformly distributed data this is not the case and the time to list the points in an ϵ-component can become very long.

Our algorithms to build the MST and the binary component tree are not the most efficient implementations. There is a large literature on special data structures for graph algorithms [9] that could be exploited to give faster implementations.

Finally, we must address the problem of how to determine the finest appropriate resolution, ρ, as discussed at the beginning of Section 2.3. To do this, we examine how the number of isolated points in the ϵ-decomposition of the set varies with resolution —i.e., the function $I(\epsilon)$. In all of the examples below, the underlying sets are perfect, so the finite point-set approximation is "bad" at any resolution for which there are isolated points. It follows that the resolution at which we start to see isolated points is an estimate of ρ, i.e. $\rho = \inf\{\epsilon : I(\epsilon) = 0\}$. The validity of this approach is supported by the numerical evidence in the next section; the data for $C(\epsilon)$ and $D(\epsilon)$ diverge from their true values at the resolution when isolated points are first detected.

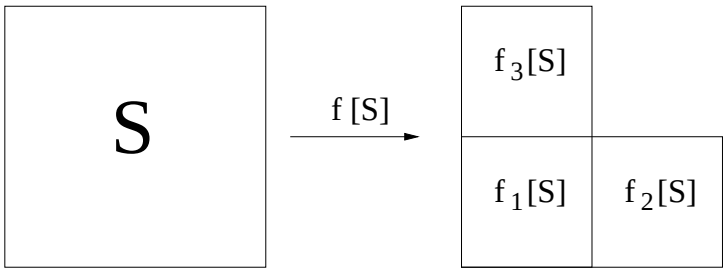


Figure 2.3: Template for the iterated function system that generates the Sierpinski triangle relatives

2.4 Examples

In this section we present some examples that illustrate the behavior of the number of ϵ-components, $C(\epsilon)$, the largest diameter, $D(\epsilon)$, and the number of isolated points, $I(\epsilon)$, for fractals with different topology. The goal is to show that these quantities give useful information about the topology.

The first examples are relatives of the Sierpinski triangle. These sets are generated from a family of iterated function systems (IFS). The Hausdorff dimension of each set is identical, even though they have different topological structure, as the disconnectedness and discreteness indices highlight. For these fractals, we show that the cutoff resolution decreases when the number of data points is increased, which is not surprising, since more points sampled from an attractor constitute a better approximation of the underlying set. We also vary the way in which the data cover the set and find that for a fixed number of points the cutoff resolution is a minimum when the data are uniformly distributed. Again, this is exactly what we expect, since isolated points appear at larger values of ϵ when points are not evenly spaced.

We next present a number of Cantor-set examples to illustrate different types of scaling behavior in $C(\epsilon)$ and $D(\epsilon)$. We observe that, for Cantor sets with zero Lebesgue measure, the computed value of γ is approximately equal to the dimension of the set. For a Cantor set with positive measure, though, γ and the dimension are significantly different.

2.4.1 Relatives of the Sierpinski triangle

Sierpinski triangle relatives are attracting fixed sets of a family of iterated function systems:

$$S = f[S] = f_1[S] \cup f_2[S] \cup f_3[S].$$

In each case, the functions f_i are similarity transformations[1] of the unit square with contraction ratio $\frac{1}{2}$, as illustrated by the template in Figure 2.3. The functions that generate the Sierpinski triangle, Figure 2.4, are simple contractions composed with a translation; the generators of the examples in Figure 2.11 and Figure 2.13 involve additional rotation or reflection symmetries of the square. There are 232 different fractals in this family [66] and their Hausdorff dimensions are identical: $\dim_H = \log 3 / \log 2$. Their topology, however, ranges from simply connected to connected to totally disconnected to a class of examples with infinitely many connected components of non-zero diameter [71]. This range of structure makes them ideal test cases for our techniques.

[1]Recall from Section 1.2.3 that similarity transformation, S, is an affine transformation that contracts or dilates distance uniformly; i.e., for all x and y, there is a positive number r such that $|S(x) - S(y)| = r|x - y|$.

It is easy to generate a finite number of points on the attractor of an iterated function system. One way (Barnsley's chaos game [4]) is to chose an initial point x_0 in the domain of the IFS and then record its trajectory under the iteration $x_{n+1} = f_{i_n}(x_n)$, setting $i_n = 1, 2$ or 3 with probability p_1, p_2 and p_3 respectively. If x_0 is in the attractor then its entire orbit is in the attractor; if not, the iterates converge to it. In the examples below, we choose $x_0 = (0, 0)$. Thus, the orbit can be viewed as a random sampling of the attractor by a finite number of points. When $p_1 = p_2 = p_3 = \frac{1}{3}$ the data cover the fractal uniformly; if the probabilities are not equal the distribution of points is nonuniform and their density approximates a multifractal measure. In the appendix, we give Matlab code for generating the Sierpinski triangle relatives in the above manner.

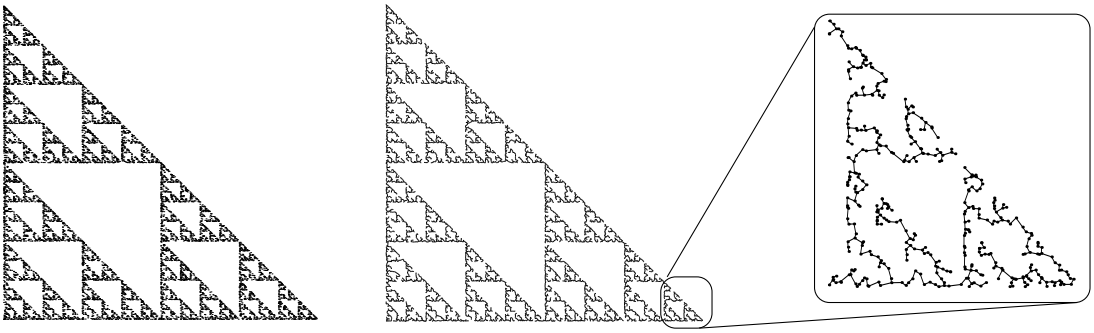
The Sierpinski triangle

The generating functions for the Sierpinski triangle are:

$$\begin{aligned} f_1(x, y) &= \tfrac{1}{2}(x, y) \\ f_2(x, y) &= \tfrac{1}{2}(x + 1, y) \\ f_3(x, y) &= \tfrac{1}{2}(x, y + 1). \end{aligned} \tag{2.4}$$

A finite point-set approximation to the triangle and the corresponding minimal spanning tree are shown in Figure 2.4. The underlying set is connected and perfect, so we expect to see $C(\epsilon) = 1$, $D(\epsilon) = \sqrt{2}$, and $I(\epsilon) = 0$ for $\epsilon > \rho$. This is corroborated by the calculations of $C(\epsilon)$ and $D(\epsilon)$ for 10^4 and 10^5 point approximations to the triangle, as shown in Figure 2.5. We see that for ϵ above a threshold value, the computed values of $C(\epsilon)$ and $D(\epsilon)$ are in exact agreement with our expectations. The point at which $C(\epsilon)$ and $D(\epsilon)$ deviate from the ideal values is the value of ϵ at which the number of isolated points, $I(\epsilon)$, becomes positive. This ϵ value is, of course, the cutoff resolution ρ discussed in Section 2.3.2. At finer resolutions — i.e., $\epsilon < \rho$ — we see a sharp transition in the number of connected components from one to the number of points in the set; the diameters show a correspondingly sharp decrease. Both these effects are due to the narrow distribution of edge-lengths of the MST. Clearly, the value of ρ depends on the number of points, N, covering the set. For the 10^4-point approximation, $\rho \approx 0.008$ and for $N = 10^5$, $\rho \approx 0.0022$; We expect the relationship to be $\rho \approx 1/\sqrt{N}$, since the data are homogeneously distributed on a subset of $\mathbb{R}^2$. This is supported by the data in Figure 2.6(a). Here, we plot cutoff resolution versus the number of points for $10^3 \leq N \leq 10^5$; the slope of the least-squares fit line is -0.58.

The results discussed so far are for uniformly distributed data. In general dynamical systems, though, orbits cover attractors nonuniformly. Therefore, we want to understand the effects that nonuniformly distributed points have on the connectedness data. As described earlier, we can change the way an orbit covers the IFS attractor by choosing the functions f_1, f_2, and f_3 with different probabilities. To generate Figure 2.7(a), for example, we set $p_1 = 0.05$ and $p_2 = p_3 = 0.475$. This highly nonuniform distribution of points induces perceptible changes in the $C(\epsilon)$, $D(\epsilon)$ and $I(\epsilon)$ data, as shown in Figure 2.8, but the graphs remain qualitatively similar to those in Figure 2.5. The cutoff resolution is significantly larger: $\rho \approx 0.04$ compared with 0.008 for the uniform distribution with the same number of points. The growth in the number of ϵ-components for $\epsilon < \rho$ is also less rapid than that for the uniform data. Both of these changes are due to a greater spread in the edge-lengths of the MST. The geometry of the distribution is reflected in the graph of $D(\epsilon)$; the densely covered diagonal means that $D(\epsilon) = \sqrt{2}$ for ϵ values significantly less than ρ.



(a) (b) (c)

Figure 2.4: (a) 10^4 points uniformly distributed on the Sierpinski triangle. (b) The corresponding MST. (c) A close up of the bottom right corner of the MST.

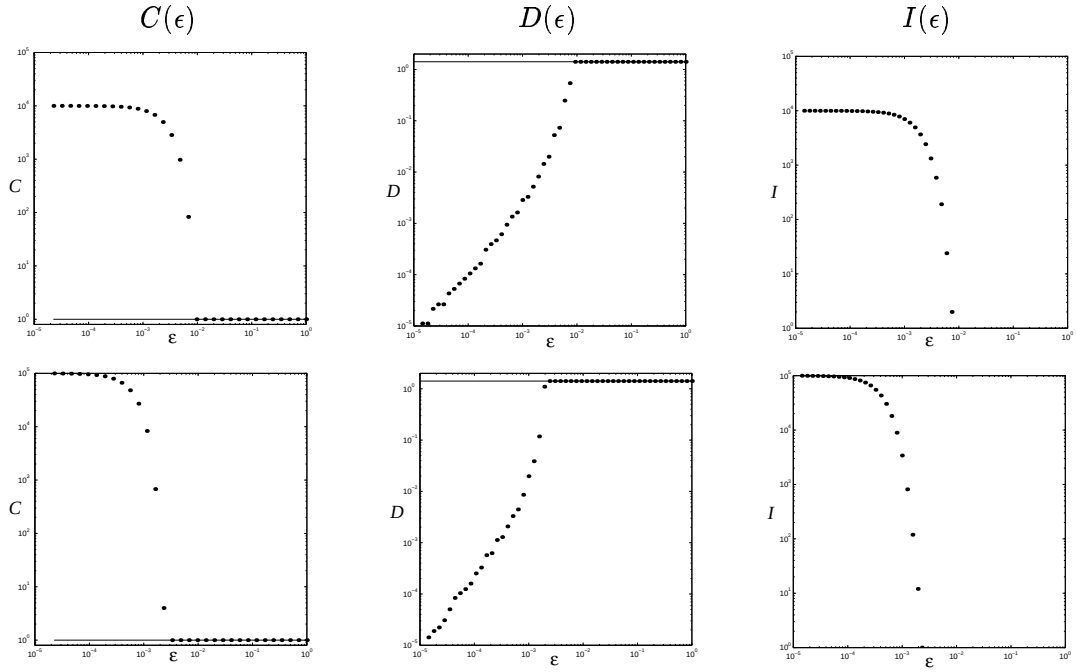


Figure 2.5: $C(\epsilon), D(\epsilon)$ and $I(\epsilon)$ for the Sierpinski triangle. The top row gives results for 10^4 uniformly distributed points on the fractal and the bottom row for 10^5 points. All axes are logarithmic. The horizontal axis range is $10^{-5} < \epsilon < 1$. The solid lines represent $C(\epsilon)$ and $D(\epsilon)$ for ideal data; the dots are the computed values.

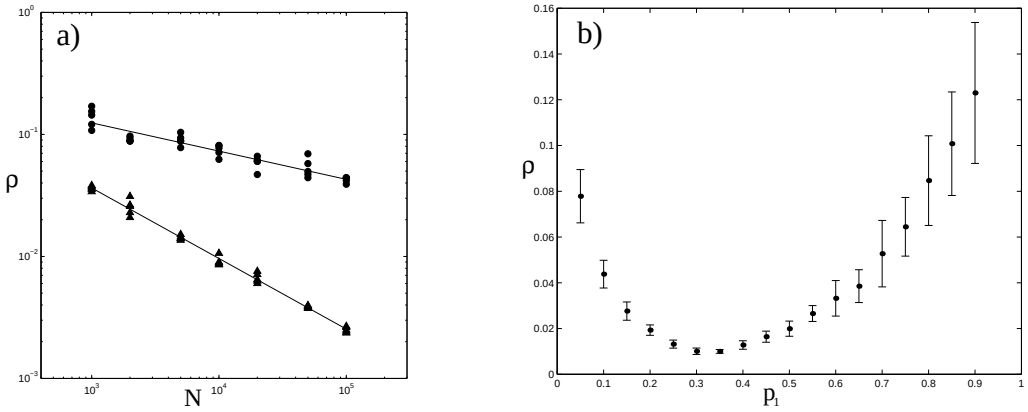
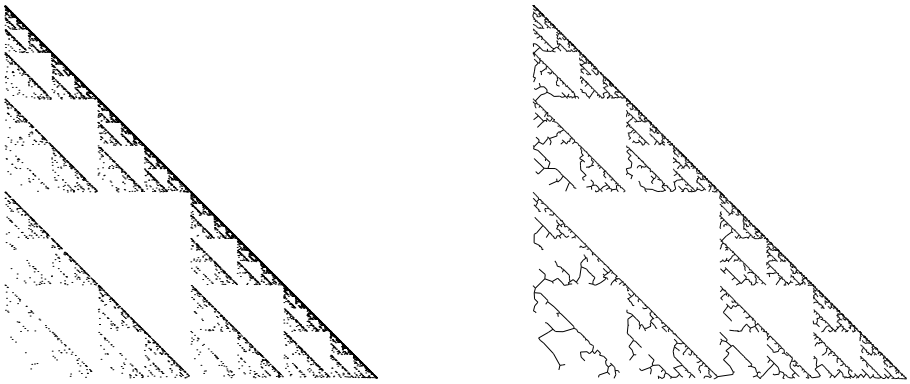


Figure 2.6: (a) Cutoff resolution, ρ, as a function of the number of points, $10^3 \leq N \leq 10^5$, covering the Sierpinski triangle for two values of p_1; ● marks data for the nonuniform distribution with $p_1 = 0.05$ and ▲ marks data for $p_1 = \frac{1}{3}$, i.e., a uniform distribution. (b) Cutoff resolution as a function of p_1 for 10^4 data points on the Sierpinski triangle. The error bars are the standard deviation about the mean of twenty calculations of ρ for each value of p_1.

We can lower the cutoff resolution by increasing the number of data points but, as shown in Figure 2.6(a), the rate at which ρ decreases is $\rho \approx N^{-0.23}$, significantly slower than that for the uniform data. This is an interesting result that deserves further investigation. The distribution of edge-lengths in the MST must depend on the distribution of points, and perhaps there is a way to formally relate the two distributions. For a connected set, ρ is the largest edge-length in the MST. Thus, given an analytic form for the distribution of edge-lengths, it may be possible to predict the scaling of ρ with the number of points. We arrive at similar questions in Chapter 4 when we consider point distributions on invariant circles of the standard map.

Finally, the graph in Figure 2.6(b) summarizes the variation of the cutoff resolution with the nonuniformity in the distribution of points. The measure of nonuniformity in the data is p_1, the probability of choosing f_1; we set $p_2 = p_3 = (1 - p_1)/2$. Twenty orbits of 10^4 points were generated for values of p_1 in the range 0.05 to 0.9. The cutoff resolution reaches a minimum at $p_1 = \frac{1}{3}$, i.e., for uniformly distributed data, as we expect. The other feature to note is that the standard deviation also depends on the distribution, and is greatest for highly nonuniform data. This is because when one function in the IFS is chosen with very low probability, there is greater variability in the way an orbit fills out the attractor; this in turn leads to greater variation in the edge-lengths of the MST.

We conclude that for moderate amounts of nonuniformity (for this example, $0.2 \leq p_1 \leq 0.5$) the cutoff resolution is at a level comparable to that for perfectly uniform data and our techniques are not adversely affected. For highly nonuniform coverings of an attractor, significantly more data points are needed to reach the same cutoff resolutions as for uniform data. The only effect this has is to generate inconclusive, rather than incorrect, diagnoses of the topology of the underlying set.



(a) (b)

Figure 2.7: (a) 10^4 points on the Sierpinski triangle generated by setting $p_1 = 0.05$ and $p_2 = p_3 = 0.475$. (b) The corresponding MST.

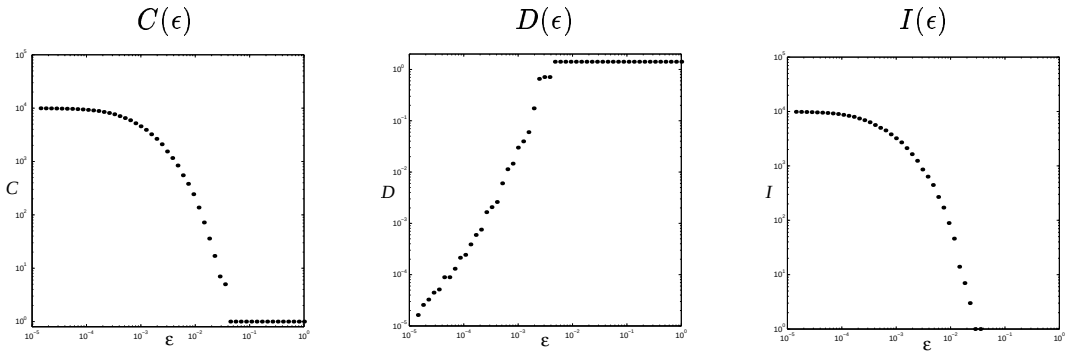
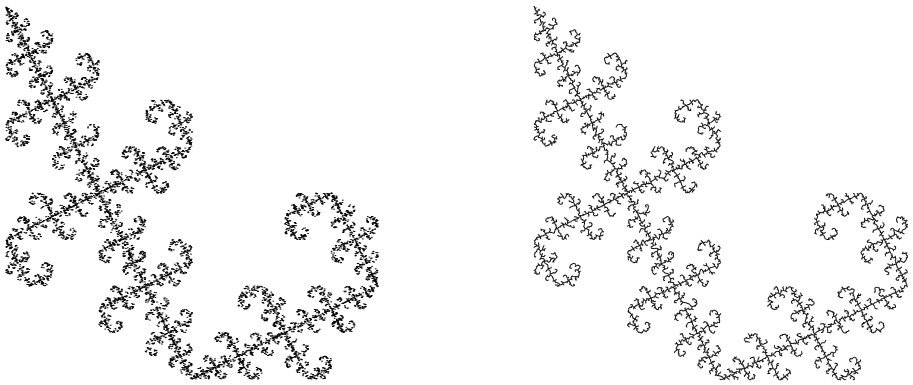


Figure 2.8: $C(\epsilon), D(\epsilon)$ and $I(\epsilon)$ for the nonuniformly distributed data set in Figure 2.7. All axes are logarithmic. The horizontal axis range is $10^{-5} < \epsilon < 1$.



(a) (b)

Figure 2.9: (a) 10^4 points on the set generated by (2.5) and (b) the corresponding MST.

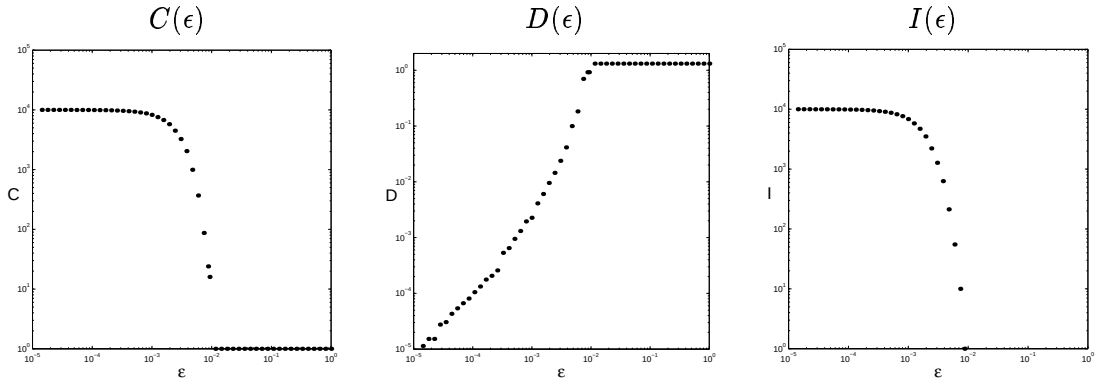
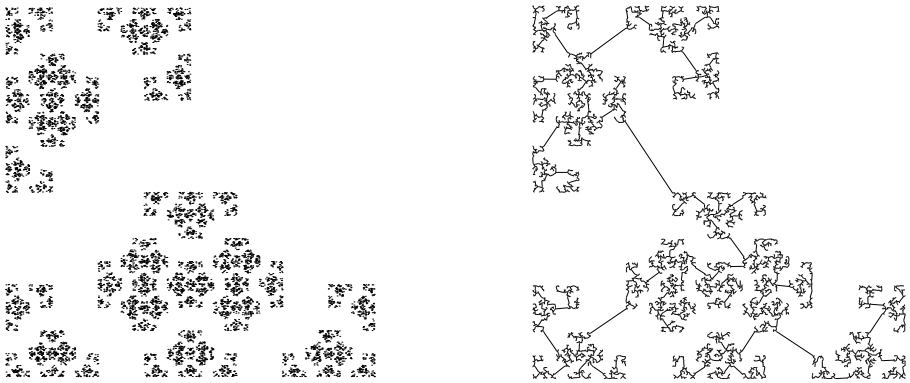


Figure 2.10: $C(\epsilon), D(\epsilon)$ and $I(\epsilon)$ for the simply connected gasket relative. The data is for 10^4 points on the set. All axes are logarithmic. The horizontal axis range is $10^{-5} < \epsilon < 1$.



(a) (b)

Figure 2.11: (a) 10^4 points on the Cantor set generated by (2.6) and (b) the corresponding MST.

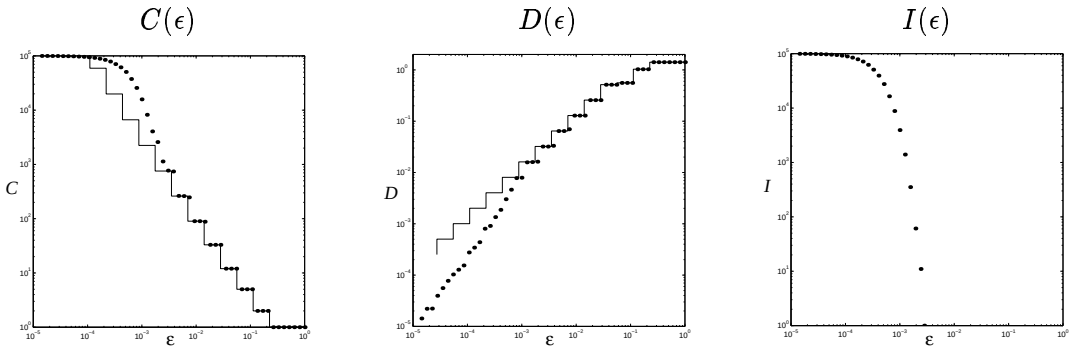


Figure 2.12: $C(\epsilon)$, $D(\epsilon)$ and $I(\epsilon)$ for 10^5 points uniformly distributed over the Cantor set triangle relative. All axes are logarithmic. The horizontal axis range is $10^{-5} < \epsilon < 1$. The solid lines represent $C(\epsilon)$ and $D(\epsilon)$ for ideal data; the dots are the computed values.

A simply connected relative

The fractal shown in figure 2.9 is generated by:

$$\begin{aligned} f_1(x,y) &= \tfrac{1}{2}(-x+1,-y+1) \\ f_2(x,y) &= \tfrac{1}{2}(-y+2,x) \\ f_3(x,y) &= \tfrac{1}{2}(x,y+1). \end{aligned} \tag{2.5}$$

This set is simply connected, in contrast to the Sierpinski triangle which has infinitely many holes. Despite this difference, we see that data for $C(\epsilon)$ and $D(\epsilon)$ are almost identical to those for the triangle. The problem of detecting holes is addressed in Chapter 3.

A Cantor set relative

Figure 2.11 shows the attractor for the iterated function system generated by

$$\begin{aligned} f_1(x,y) &= \tfrac{1}{2}(-y+1, x) \\ f_2(x,y) &= \tfrac{1}{2}(y+1, x) \\ f_3(x,y) &= \tfrac{1}{2}(y, -x+2). \end{aligned} \tag{2.6}$$

This fractal is a Cantor set, so we should see $C(\epsilon) \to \infty$ and $D(\epsilon) \to 0$ as $\epsilon \to 0$. We can derive analytic expressions for both functions, since we know that the set is three copies of itself at half the size. We call the region that disconnects two ϵ-components a "gap". Its "width" is the metric distance between the two components. Clearly, the set has a single component whenever ϵ is greater than the width of the largest gap, g_0. As ϵ decreases, subsequent components are resolved at gap sizes $g_n = g_0/2^n$. The diameters of the components follow the same pattern after some initial transient behavior. Thus for $n \geq 3$,

$$D(\epsilon) = D(g_2)/2^{n-2} \quad \text{for} \quad g_{n+1} < \epsilon \leq g_n. \tag{2.7}$$

This gives $\delta = 1$, confirming that the set is totally disconnected.

The number of components, $C(\epsilon)$, is a little harder to determine. Since $C(\epsilon)$ is just one more than the total number of gaps, we calculate it by first deriving an expression for the latter. Let N_n be the number of gaps of size g_n. Since the fractal contains three copies of itself, one might think that $N_n = 3^n$. By careful inspection of the IFS template we find that this is not the whole story — some gaps merge into one. In fact, with $N_0 = 1$ we have the recursion:

$$N_n = \begin{cases} 3N_{n-1} & \text{if } n \text{ is odd,} \\ 3N_{n-1} - 2 \cdot 3^{n/2-1} & \text{if } n \text{ is even.} \end{cases}$$

These can be solved to find:

$$N_n = \begin{cases} 2 \cdot 3^{n-1} + 3^{(n-1)/2} & \text{if } n \text{ is odd,} \\ 2 \cdot 3^{n-1} + 3^{n/2-1} & \text{if } n \text{ is even.} \end{cases}$$

We then have that for $g_{n+1} < \epsilon \leq g_n$,

$$C(\epsilon) = 1 + \sum_{j=0}^{n} N_j = \begin{cases} 3^n + 2 \cdot 3^{(n-1)/2} & \text{if } n \text{ is odd,} \\ 3^n + 3^{n/2} & \text{if } n \text{ is even.} \end{cases}$$

The leading power is the same for all n, so we may use either case to evaluate the limit:

$$\gamma = \limsup_{n\to\infty} \frac{\log(C(g_n))}{\log(1/g_n)} = \lim_{n\to\infty} \frac{\log[3^n + 3^{n/2}]}{\log[2^n/g_0]} = \frac{\log 3}{\log 2}\,.$$

Since there are no isolated points, the attractor is perfect and therefore a Cantor set, as claimed.

We can see in Figure 2.12 that the numerical calculations agree very well with the theory down to the cutoff resolution $\rho \approx 0.003$. When $\epsilon < \rho$, the computed values of $C(\epsilon)$ are larger than the predicted values because isolated points are counted as extra components. For still-smaller values of ϵ, every point is resolved as an isolated point and the $C(\epsilon)$ curve levels off. The meaningful portion of the data — between these extremes — shows a staircase periodicity about a linear trend. The slope of the linear trend is an estimate of γ. We determine γ numerically,

using a least-squares fit, to be 1.41 ± 0.05. This is lower than the true limiting value given above (1.58) because of second-order effects at the relatively large values of ϵ for which the $C(\epsilon)$ data are valid. We estimate the slope of the true curve over the same range to be 1.48, which is closer to the value computed above.

The numerically calculated values of $D(\epsilon)$ also show a staircase periodicity about a linear trend. The data have a systematic bias for the jumps at slightly larger values of ϵ than predicted by theory. This is due to finite-data effects: the points are not filling out the corners, so edges in the MST are a little longer than the "true" gaps and the diameters of the ϵ-components are a little less. The data points fall below the true curve when $\epsilon < \rho$ because inter-point distances are comparable to the inter-component distances at these resolutions. Since $D(\epsilon)$ measures the largest diameter, the flat tails of the $D(\epsilon)$ data in Figure 2.12 are due to the presence of at least one triple/pair combination within a distance ϵ of each other, whilst almost all the other points have become isolated. The slope of the linear trend of $D(\epsilon)$ is an estimate of δ. We estimate the slope over the range $\rho < \epsilon < 0.06$ because the first few steps are shallower than the limiting trend. Using a least squares fit, again, we calculate δ to be 1.00 ± 0.03, as predicted above.

The cutoff resolution for data from this IFS varies with the number of points and their distribution in exactly the same manner as the data from the Sierpinski triangle. For moderately nonuniform data, estimates of γ and δ are the same as those given above. When the data is very unevenly distributed, ρ increases significantly and the ϵ range may be too small to allow an estimate of the slope. Again, this leads to inconclusive results rather than incorrect ones.

A relative with infinitely many connected components

A third triangle relative, shown in Figure 2.13, is generated by the following similarities:

$$\begin{aligned} f_1(x,y) &= \tfrac{1}{2}(x,y) \\ f_2(x,y) &= \tfrac{1}{2}(y+1,-x+1) \\ f_3(x,y) &= \tfrac{1}{2}(x,y+1). \end{aligned} \tag{2.8}$$

The attractor for this system has infinitely many connected components, yet is not totally disconnected because the components have positive diameters. We therefore expect to see $C(\epsilon) \to \infty$ and $D(\epsilon) \to 1$ as $\epsilon \to 0$. Again, the gaps decrease simply as $g_n = g_0/2^n$. This time, the number of components is just

$$C(\epsilon) = \tfrac{1}{2}(3^{n+1}+1) \quad \text{for} \quad g_{n+1} < \epsilon \le g_n, \tag{2.9}$$

giving $\gamma = \log 3/\log 2$. The largest ϵ-component always contains the line segment $x = 0$, $0 \le y \le 1$, so $D(\epsilon) \to 1$.

The graphs of $C(\epsilon)$ and $D(\epsilon)$ in Figure 2.14 reflect this; the former has characteristics similar to those of the Cantor set above, but $D(\epsilon)$ looks like that for the Sierpinski triangle. The slope of the $C(\epsilon)$ staircase is estimated from the data for a 10^5 point approximation, giving a value of $\gamma = 1.55 \pm 0.03$. This is in very close agreement with the theoretical value of $\gamma = \log 3/\log 2 \approx 1.585$.

2.4.2 Cantor sets in the plane

One of our objectives is to use our techniques to identify and characterize phase-space structures in dynamical systems. Cantor sets are often present in chaotic dynamical systems, so it is useful to examine some simple Cantor set examples to gain a better understanding of the different

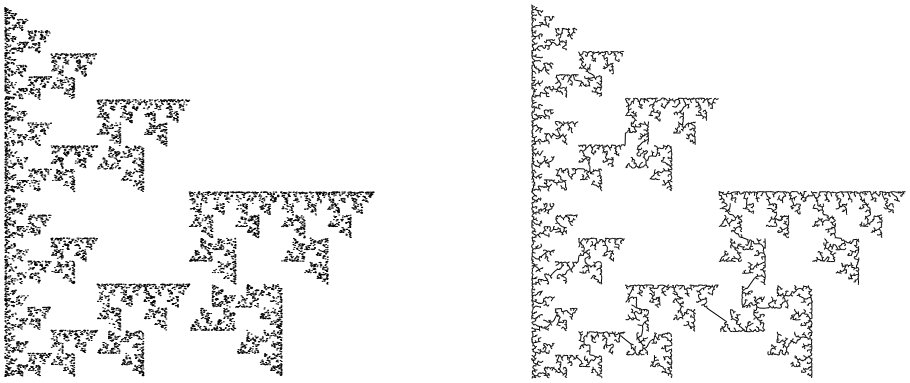


Figure 2.13: (a) 10^4 points on the fractal generated by (2.8) and (b) the corresponding MST.

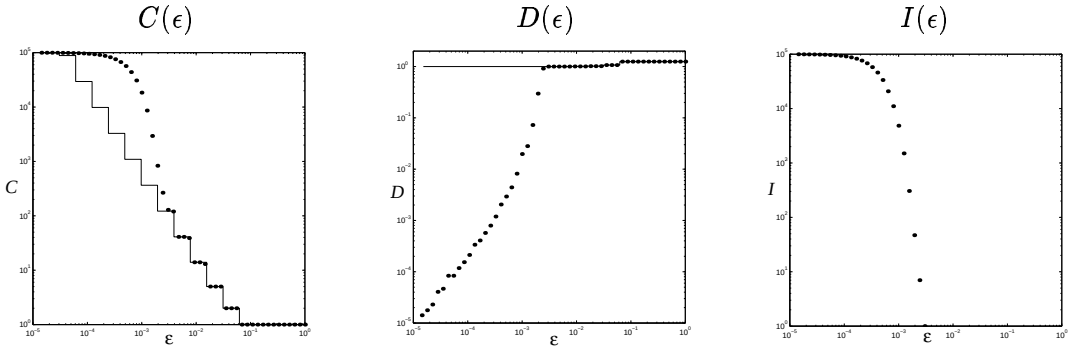


Figure 2.14: $C(\epsilon)$, $D(\epsilon)$ and $I(\epsilon)$ for a triangle relative with infinitely many components. Again, the data is for 10^5 points uniformly distributed on the set. All axes are logarithmic. The horizontal axis is $10^{-5} < \epsilon < 1$. The solid line represents $C(\epsilon)$ and $D(\epsilon)$ for ideal data; the dots are the computed values.

types of scaling that can occur in the $C(\epsilon)$ and $D(\epsilon)$ graphs. In Figure 2.15, Figure 2.17, and Figure 2.19, we show five Cantor sets in the plane. In each case, the orbit has 50000 points. Four of these have zero Lebesgue measure and one (Figure 2.17(b)) has positive measure, so it is termed a fat Cantor set (this is analogous to the term "fat fractal" for fractals with positive measure [22]). All are attractors of iterated function systems of the form

$$S = f[S] = f_1[S] \cup f_2[S] \cup f_3[S] \cup f_4[S].$$

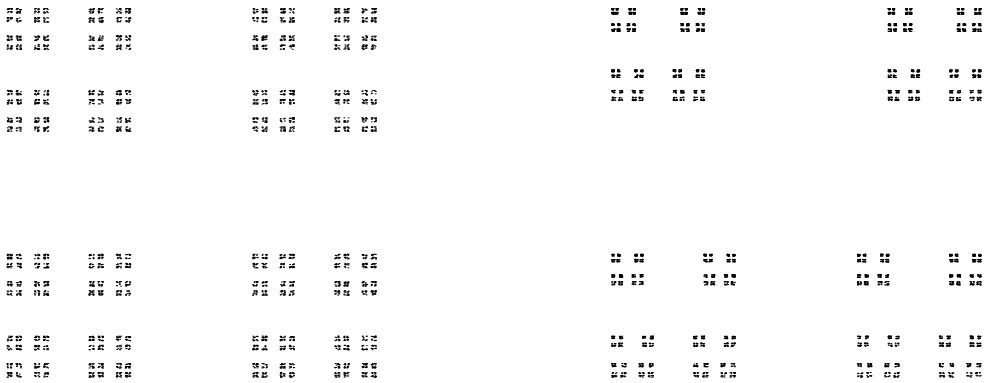
The generating functions, f_i, become increasingly complex in this series of examples. The three simplest involve only affine transformations and a fourth uses conformal functions; the functions that generate the fat Cantor set cannot be written in closed form. The geometric structure of each set is reflected in the type of staircase seen in the graphs of $C(\epsilon)$ and $D(\epsilon)$. For the four examples with zero Lebesgue measure, we expect to see $\gamma = \dim_B$, the box-counting dimension; this is supported by our results, summarized in Table 2.1. For the Cantor set with positive measure, the value of γ is significantly different from the dimension. We again observe, for all of the examples, that the cutoff resolution, ρ, is well approximated by the ϵ-value where the number of isolated points ceases to be zero. Again, this is because all of the underlying sets are perfect.

We start with a simple example where each f_i is a similarity transformation with contraction ratio $\frac{1}{3}$, as shown in Figure 2.15(a). The numerical calculations of $C(\epsilon)$, $D(\epsilon)$, and $I(\epsilon)$ are presented in the top row of Figure 2.16. These graphs show staircase scaling behavior similar to the triangle-relative Cantor set presented in Section 2.4.1; this makes sense because both are generated by iterated function systems of similarity transformations. For this example, the jumps in $C(\epsilon)$ and $D(\epsilon)$ are at $\epsilon_n = 1/3^n$; because there are four self-similar copies at one third the size, we see $C(\epsilon_n) = 4^n$ and $D(\epsilon_n) = 1/3^n$, which gives the theoretically determined limits of $\gamma = \log 4/\log 3 \approx 1.262$ and $\delta = 1$. This is in close agreement with our numerical estimate of $\gamma = 1.23 \pm 0.05$ and $\delta = 0.97 \pm 0.02$.

The second example, Figure 2.15(b), is also generated by similarities. This time, the lower two have a contraction ratio of $\frac{1}{3}$ and the upper two have a ratio of $\frac{1}{4}$. As can be seen from the second row of Figure 2.16, this leads to a more complicated staircase pattern in the $C(\epsilon)$ and $D(\epsilon)$ graphs. Jumps in these graphs occur at values of ϵ corresponding to edge lengths in the MST. The structure from the IFS means these edge lengths are of the form $l(\frac{1}{3})^m(\frac{1}{4})^n$, for all integers m and n, where l is one of the two longest edges. Values of γ and δ, presented in Table 2.1, are again very close to the expected values.

To generate the set in Figure 2.17(a), more-general affine transformations are used, each contracting by $\frac{1}{3}$ horizontally and $\frac{1}{4}$ vertically. The corresponding graphs of $C(\epsilon)$ and $D(\epsilon)$ in Figure 2.18 show the now-familiar staircase scaling pattern. Compared to the second Cantor set example, the larger steps in these graphs reflect the more regular geometric structure of the set.

The fourth example is a Cantor set with positive Lebesgue measure and therefore a dimension of 2. It is possible to represent this set as the attractor of an iterated function system of the general form above. The functions involved, however, are limits of piecewise linear approximations and it is not possible to write them in closed form. Instead, we generate the set as the cross product of two positive measure Cantor subsets of the unit interval. These sets are constructed as follows: at each level, $n \geq 1$, 2^{n-1} gaps of length $a/2^{pn-1}$ are removed from the center of an interval remaining from level $n-1$. The sum of the gap lengths is $a/(2^{p-1}-1)$; choosing p and a to make this length less than one ensures the Cantor set has positive measure. It is easy to recursively generate the end points of the gaps (down to some level) and these points are used as the finite point-set approximation. For the set in figure 2.17(b), we set $a = 2/3$ and $p = 2$ for



(a) (b)

Figure 2.15: Cantor sets generated by iterated function systems of four similarity transformations. Both sets have 50000 points. (a) Similarities with contraction ratio $\frac{1}{3}$. (b) The upper two similarities have ratio $\frac{1}{4}$ and the lower two have ratio $\frac{1}{3}$.

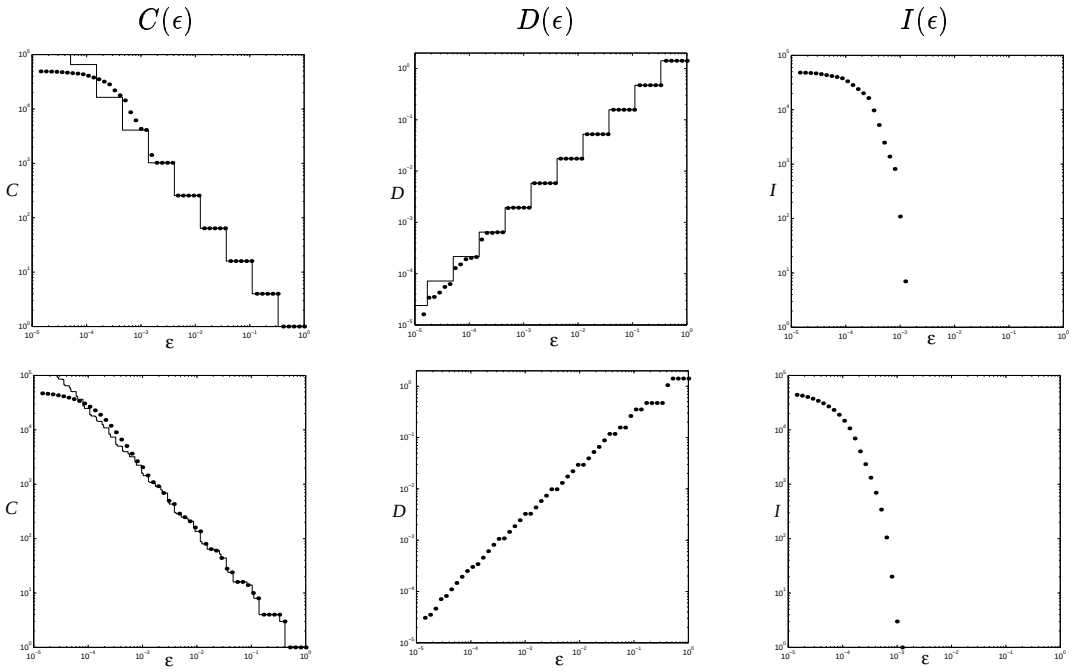


Figure 2.16: $C(\epsilon), D(\epsilon)$ and $I(\epsilon)$ for the Cantor sets in figure 2.15. The top row is data for Figure 2.15(a); the second row is for Figure 2.15(b). All axes are logarithmic. The horizontal axis range is $10^{-5} < \epsilon < 1$. The solid lines represent $C(\epsilon)$ and $D(\epsilon)$ for ideal data; the dots are the computed values.

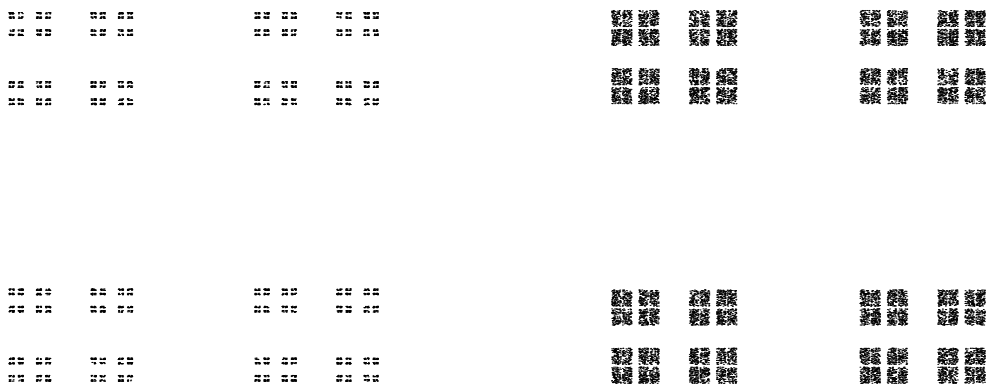


Figure 2.17: Two Cantor sets with largest gaps of $1/2$ and $1/3$. (a) A set generated by an IFS of four affine transformations with horizontal contraction of $\frac{1}{3}$ and vertical contraction of $\frac{1}{4}$. (b) A fat Cantor set, generated as the cross product of two Cantor sets of positive measure in the real line.

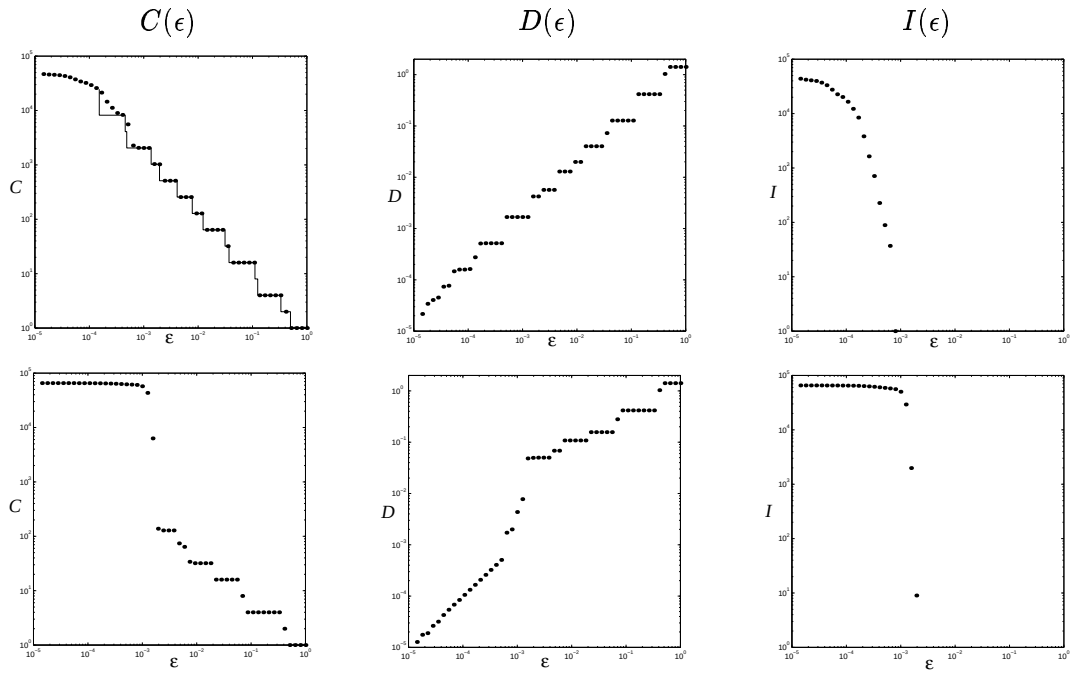
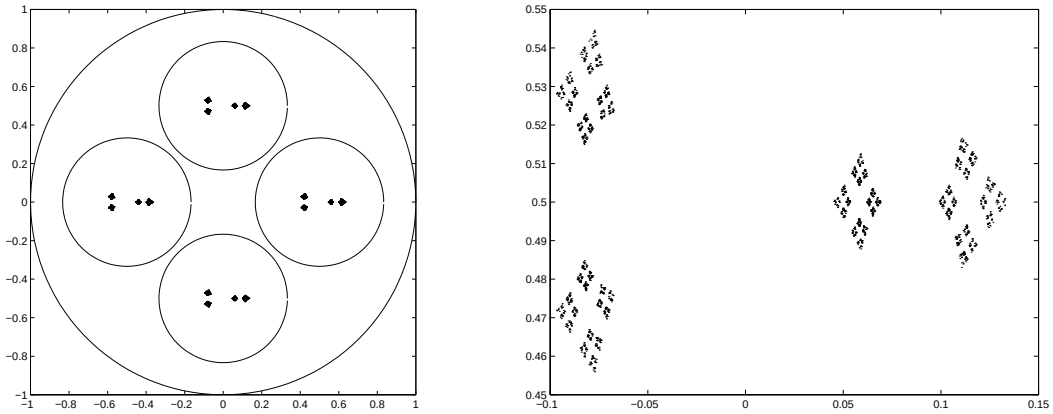


Figure 2.18: $C(\epsilon), D(\epsilon)$ and $I(\epsilon)$ for the 2-D Cantor sets in Figure 2.17. The top row is data for Figure 2.17(a); the second row for Figure 2.17(b), the fat Cantor set; All axes are logarithmic. The horizontal axis range is $10^{-5} < \epsilon < 1$.



(a) (b)

Figure 2.19: A Cantor set generated by an IFS consisting of four nonlinear affine transformations, each mapping the unit circle into a circle of radius $\frac{1}{3}$. (a) The data set with circle boundaries. (b) A close up of one of the four clusters.

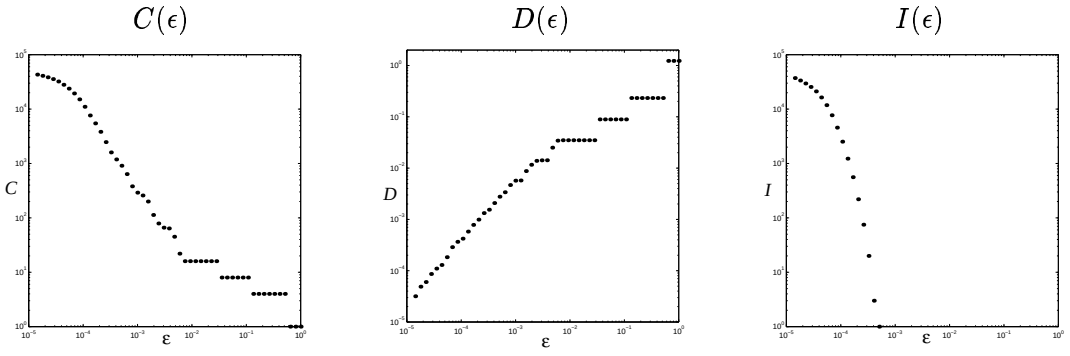


Figure 2.20: $C(\epsilon), D(\epsilon)$ and $I(\epsilon)$ for the nonlinear Cantor set of Figure 2.19. All axes are logarithmic. The horizontal axis range is $10^{-5} < \epsilon < 1$.

the horizontal cross-section, and $a = 2$, $p = 3$ for the vertical one. The behavior of $C(\epsilon)$ and $D(\epsilon)$, shown in the bottom row of Figure 2.18, is not unlike that of the previous example. The slopes are significantly shallower, however, because the gaps are decreasing at a faster rate than the component diameters, giving $\delta \approx 0.46$. The component growth rate, γ, is approximately 0.80, which is clearly distinct from the box-counting (and Hausdorff) dimension of 2.

The final IFS example uses nonlinear, conformal transformations. A function, F, is conformal if its derivative matrix at each point, $DF(x)$ is a similarity transformation. For the set illustrated in Figure 2.19, the functions are of the form $f_k(z) = z^2/3 + c_k$, where $z = x + iy$, and the translations, c_k for $k = 1, \ldots, 4$, take the values $\{\pm\frac{1}{2}, \pm\frac{i}{2}\}$. Notice that although we choose the f_i with equal probability, the nonlinearity introduces a nonuniformity to the distribution of points over the Cantor set. The cutoff resolution $\rho \approx 5 \times 10^{-4}$ is nevertheless comparable to the previous examples of uniformly distributed data. Scaling in the graphs of $C(\epsilon)$ and $D(\epsilon)$ occurs in two distinct ϵ intervals; see Figure 2.20. For $0.005 < \epsilon < 1$, there are three shallow steps reflecting the large-scale structure that is visible in figure 2.19(a). The

Table 2.1: A summary of the values of γ and δ for Cantor subsets of the plane. The numbers are estimated using a least-squares linear fit to logarithmic plots of $C(\epsilon)$ and $D(\epsilon)$, respectively; the error margins are estimated by varying the scaling range. The second column gives the box-counting dimension, $\mathrm{d}_{\mathrm{box}\,B}$, for each set; these numbers are computed using formulas from Falconer [23].

Data Set	$\dim_B$	γ	δ
Fig. 2.15(a)	1.262	1.23 ± 0.02	0.96 ± 0.04
Fig. 2.15(b)	1.126	1.11 ± 0.02	1.00 ± 0.03
Fig. 2.17(a)	1.131	1.13 ± 0.01	0.98 ± 0.03
Fig. 2.17(b)	2	0.80 ± 0.05	0.46 ± 0.05
Fig. 2.19	$1.21 < \dim_B < 1.34$	1.36 ± 0.03	0.95 ± 0.05

second portion of the data, for $\rho < \epsilon < 0.005$, has a steeper slope, corresponding to the limiting small-scale structure of the set. The values of γ and δ given in Table 2.1 are slopes of the $C(\epsilon)$ and $D(\epsilon)$ over the interval $\rho < \epsilon < 0.005$. We find, as for the previous zero-measure Cantor sets, that γ is close to the box-counting dimension and $\delta \approx 1$.

2.5 Concluding remarks

Our results demonstrate that the minimal spanning tree of a finite set of points can provide accurate information about the topology of the underlying set, down to a numerically computable resolution $\rho > 0$. In particular, we are able to identify sets that are connected, totally disconnected, or have infinitely many connected components with non-zero diameter. Confidence in the extrapolation can be increased by sampling more points in order to get a better approximation to the underlying set, but of course we are still ultimately restricted by the machine precision. Connected sets have disconnectedness index $\gamma = 0$, and discreteness index $\delta = 0$. Based on the examples in the chapter, we conjecture that Cantor sets with zero Lebesgue measure have γ equal to the box-counting dimension and $\delta = 1$. Results in this direction are given in Chapter 5. The fat Cantor set example shows that γ and dimension are not the same when the set has positive Lebesgue measure.

The points in most of the examples of this chapter are fairly evenly distributed over the underlying set. However, it is often the case that orbits cover an attractor in a highly nonuniform way. Previous work on characterizing fractals has dealt with this by introducing a concept of dimension for measures, and developing a theory of multifractals [24]. The results of Section 2.4.1 show that our techniques are most effective for uniformly distributed data; for a fixed number of data points, the cutoff resolution, ρ, is minimal when the points are evenly spread over the attractor. Our techniques can still give valid results for highly nonuniform data; the difference is that ρ may be too high in these cases to make strong statements about the topology of the underlying set. It may be possible to weight the edges of the MST by the density of the data distribution and thereby lower ρ in these situations.

A natural question that arises from studying the Sierpinski triangle relatives is how to distinguish between simply connected sets and ones with holes. This involves reformulating concepts from homology theory by introducing a resolution parameter, similar to the way we treat the definition of connectedness. This is the subject of the next chapter.

Chapter 3

Computing Homology

3.1 Introduction

We now turn our attention to the more difficult problem of deducing the homology of a compact metric space from a finite amount of data. The homology groups of a space characterize the number and type of holes in that space and therefore give a fundamental description of its structure. This type of information is used, for example, in understanding the structure of attractors from embedded time series data [58, 60], or for determining similarities between proteins in molecular biology [13].

The computability of homology groups from a given triangulation is well-known and the algorithm uses simple linear algebra [61]. This algorithm has extremely poor numerical behavior, however, so the study of computational homology remains an active area of research. In many applications knowledge of the entire group structure is unnecessary — all that is needed is the rank of the homology group, i.e., the Betti number. This information can be computed indirectly from a triangulation; many algorithms exist [11, 14, 26, 37]. The extraction of homology from data involves the additional problem of generating a triangulation or other regular cell-complex that reflects the topology of the underlying space. There are many different approaches to this [18, 37, 60].

Our goal is to develop computational techniques that allow us to extrapolate the homology of an underlying compact space, X, given only a finite approximation, S. We assume that S approximates X in a metric sense, i.e., that every point of S is within distance ρ of X and vice versa. As in Chapter 2, the basic trick is to coarse-grain the data at a sequence of resolutions that tend to zero. Of course, the extrapolation will always be constrained by the accuracy of the data — this is measured by ρ, the cutoff resolution. We model the coarse-graining of a set, X, by closed ϵ-neighborhoods:[1]

$$X_\epsilon = \{x \mid d(x, X) \leq \epsilon\}.$$

Our main contribution to the literature on computational homology is sound mathematical foundations for relating the homology of the ϵ-neighborhoods of S to the homology of X. Our multiresolution approach has the additional advantage of being applicable to fractal sets.

We begin this chapter with an overview of the relevant concepts from homology theory in Section 3.2. The section starts by describing simplicial homology theory. This theory is based on finite triangulations so is readily adapted to computer implementation. Fractal sets do not

[1] This is a different coarse-graining procedure to that in Chapter 2 — the resolution parameter there is related to the distance between points. A set is ϵ-connected in the sense of Chapter 2 if its $\epsilon/2$-neighborhood is connected.

have finite triangulations, so a more general homology theory is needed. The appropriate formulation is Čech homology. The basis of Čech homology is an inverse system of finite triangulations that approximate a space. This idea is extended by shape theory which considers inverse systems of approximating spaces in a more general class. Our sequence of ϵ-neighborhoods fits this framework.

The central results of this chapter are given in Section 3.3. This section describes the inverse system of ϵ-neighborhoods of X and the corresponding inverse systems of homology groups. We would like to quantify the topological structure by computing Betti numbers as functions of ϵ. There is a problem with this however, since the ϵ-neighborhood X_ϵ can have holes that do not exist in X. We resolve this problem by introducing the concept of persistent Betti number, which counts the number of holes in X_ϵ that correspond to a hole in X. When X has fractal structure, it is possible to see unbounded growth in the persistent Betti numbers as $\epsilon \to 0$. We characterize this growth by assuming an asymptotic power law. The next part of this section derives formal relationships between the ϵ-neighborhoods of X and a finite point-set approximation. For the finite approximations, we show that it is possible to reduce the computation of ρ-persistent Betti numbers to linear algebra.

In Section 3.4, we turn to the practical problem of how to implement these ideas computationally. As mentioned above, there are a number of existing approaches for generating triangulations and computing Betti numbers. The one that is closest to our needs is due to Edelsbrunner *et al.* [11, 18]. Their algorithms are based on subcomplexes of the Delaunay triangulation called alpha shapes. We describe their approach in some detail since we use the NCSA implementation of these algorithms to generate data in Section 3.5. We also give a brief overview of some other computational homology algorithms. In the final part of this section, we outline a multiresolution approach to computational homology that may be a more efficient implementation of our ideas.

We use the Sierpinski triangle relatives as test examples again in Section 3.5. Since we have not yet implemented algorithms for computing the persistent Betti numbers, we give data for the regular Betti numbers. This distinguishes between the connected Sierpinski triangle and the simply connected relative. The examples demonstrate that the regular Betti numbers can be misleading and that the persistent Betti numbers are necessary for a proper characterization of the underlying topology.

The material in Sections 3.3–3.5 is published in [69].

3.2 An overview of homology theory

Homology theory is a branch of algebraic topology that attempts to distinguish between spaces by constructing algebraic invariants that reflect the connectivity properties of the space. The field has it origins in the work of Poincaré. In Section 3.2.1 we review the basic definitions for simplicial homology — a theory based on triangulations of spaces. Simplicial homology lends itself to computational implementation because triangulations of data are common numerical constructions and there is a well defined algorithm for computing homology groups from a given triangulation. Fractals typically require infinitely many simplices in their triangulations. This means the groups associated with a fractal should also be infinite, but this is not possible with simplicial homology. A different approach is needed, therefore, to describe spaces with infinitely detailed structure. The appropriate formulation of limit for this situation is given by the machinery of inverse systems, which we describe in Section 3.2.3. An inverse limit system is then used to define Čech homology in Section 3.2.4. We also give a brief outline of

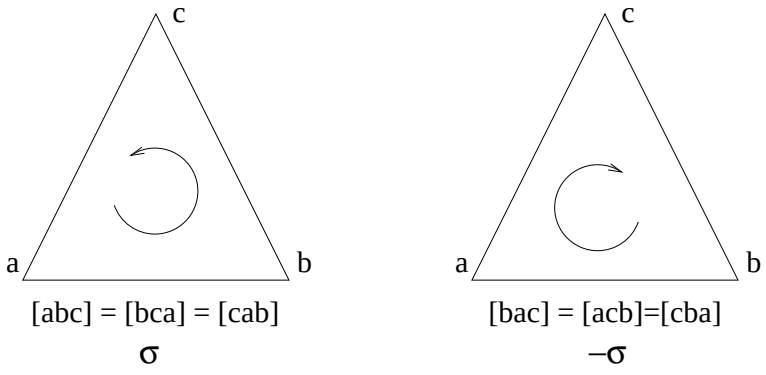


Figure 3.1: Two triangles (2-simplices) with opposite orientations.

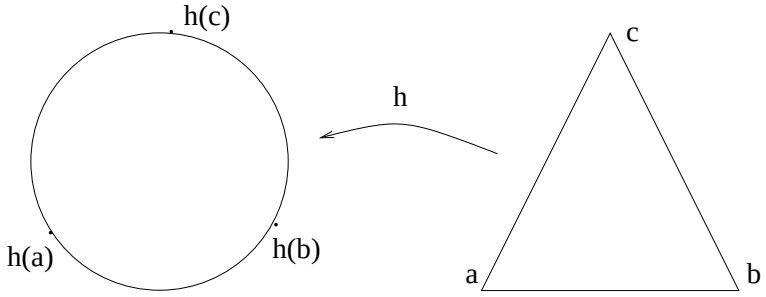


Figure 3.2: A triangulation of the circle. The simplicial complex contains the 0-simplices a, b, and c, and the 1-simplices $[ab], [bc]$, and $[ca]$. The homeomorphism from the complex to the circle is denoted by h.

shape theory, another branch of algebraic topology. The basic result of shape theory is that any compact metric space can be approximated by an inverse system of polyhedra. We use this idea to formulate the mathematical foundations for our approach to computational homology.

3.2.1 Simplicial homology

There are a number of different, but equivalent, formulations of homology theory. The simplest to understand is simplicial homology. This theory is based on triangulations of topological spaces (simplicial complexes). Singular homology is a more general technique that uses maps of simplices into a general topological space. Another common approach uses CW-complexes that are built from general n-dimensional cells rather than simplices. This is probably the most popular tool in current research, and there is a good introduction to algebraic topology from this perspective by Hatcher [32]. We focus on simplicial homology, since it is the easiest to adapt for implementation on a computer. The notation we use in this section is based on that of Munkres [61].

Simplicial complexes

The basic building block is an *oriented k-simplex*, σ^k — the convex hull of $k+1$ geometrically independent points, $\{x_0, x_1, \ldots, x_k\} \subset \mathbb{R}^n$, with $k \leq n$. For example, a 0-simplex is just a point, a 1-simplex is a line segment, a 2-simplex a triangle, and a 3-simplex is a tetrahedron. We write $\sigma^k = [x_0, x_1, \ldots, x_k]$ to denote a k-simplex and its vertices. The ordering of the vertices

defines an orientation of the simplex. This orientation is chosen arbitrarily but is fixed. Any even permutation of the vertices in a simplex gives another simplex with the same orientation, while an odd permutation gives a simplex with negative orientation; see Figure 3.1.

An abstract *simplicial complex*, $\mathcal{C}$, is a collection of oriented simplices with the property that a non-empty intersection of two simplices in $\mathcal{C}$ must itself be a simplex in $\mathcal{C}$. If the simplicial complex is finite then it can always be embedded in $\mathbb{R}^n$ for some n; certain complexes with infinitely many simplices can also be embedded in finite-dimensional space (see Figure 3.10 for an example). Such an embedded complex is a geometric realization of $\mathcal{C}$. The subset of $\mathbb{R}^n$ occupied by the geometric complex is denoted by $|\mathcal{C}|$ and is called a *polytope* or *polyhedron*.

When a topological space X is homeomorphic to a polytope, $|\mathcal{C}|$, it is called a *triangulated space*, and the simplicial complex $\mathcal{C}$ is a *triangulation* of X. For example, a circle is homeomorphic to the boundary of a triangle, so the three vertices a, b, c and three 1-simplices, $[ab], [bc], [ca]$ are a triangulation of the circle; see Figure 3.2. A complete characterization of the class of topological spaces that have a triangulation is not known.

Homology groups

We now define the group structures associated with a space, X, that is triangulated by a finite simplicial complex, $\mathcal{C}$. Although the triangulation of a space is not unique, the homology groups for any triangulation of the same space are identical; this makes simplicial homology well-defined. The essential ingredients for constructing the homology groups are the free groups of sums of k-simplices, and the boundary operator that maps a k-simplex to the $(k-1)$-simplices in its boundary.

The set of all k-simplices from $\mathcal{C}$ form the basis of a free group called the *kth chain group*, $C_k(X)$. The group operation is an additive one; recall that $-\sigma^k$ is just σ^k with the opposite orientation, so this defines the inverse elements. In general, a *k-chain* is the formal sum of a finite number of oriented k-simplices: $c_k = \sum_i a_i \sigma_i^k$. The coefficients, a_i, are elements of another group that is typically the integers: $a_i \in \mathbb{Z}$, but can be any abelian group. For Čech homology and our computational work we use rational or real coefficients.

When $k \geq 1$, 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. If $\sigma^k = [x_0, x_1, \ldots, x_k]$ is a k-simplex, we have

$$\partial_k(\sigma^k) = \sum_{i=0}^{k} (-1)^i [x_0, \ldots, \hat{x}_i, \ldots, x_k] \tag{3.1}$$

where $[x_0, \ldots, \hat{x}_i, \ldots, x_k]$ represents the $(k-1)$-simplex obtained by deleting the vertex x_i. The action of the boundary operator on general k-chains is obtained by linear extension from its action on the k-simplices: $\partial_k(\sum_i a_i \sigma_i^k) = \sum_i a_i \partial_k(\sigma_i^k)$. We drop the subscript from the boundary operator when the dimension is understood.

There are two ways to define the boundary operator ∂_0 on 0-chains. The first approach is to make $\partial_0(c_0) = 0$ for all 0-chains. With this definition, the dimension of the resulting homology group counts the number of path-connected components of a space. An alternative definition is to say that the boundary of a 0-chain $c_0 = \sum_i a_i \sigma_i^0$ is zero only when the integer coefficients add to zero: $\sum_i a_i = 0$ — this is called augmented homology. The dimension of the augmented homology group is one less than the number of path-connected components. We use the first definition, because it links back to the work in Chapter 2, where we counted the number of ϵ-connected components.

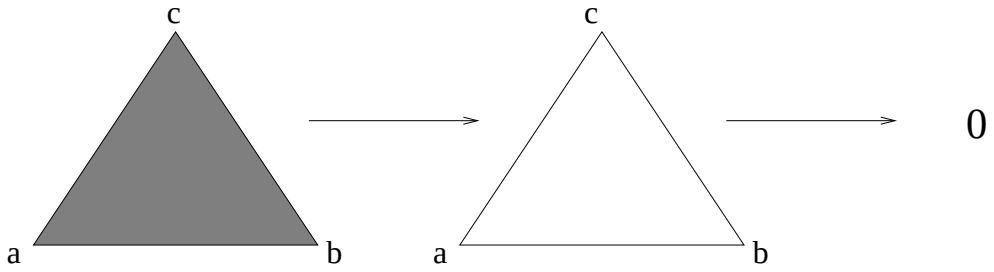


Figure 3.3: The boundary of a 2-simplex is the sum of its three edges and the boundary of this 1-chain is zero.

As an example, consider the simplicial complex consisting of a triangle and all its edges and vertices, as shown in Figure 3.3. The boundary of the 2-simplex $[a, b, c]$ is

$$\partial([abc]) = [bc] - [ac] + [ab],$$

and the boundary of this 1-chain is:

$$\partial([bc] - [ac] + [ab]) = (c - b) - (c - a) + (b - a) = 0.$$

This illustrates the fundamental property of the boundary operator, namely that

$$\partial_k \circ \partial_{k+1} = 0. \tag{3.2}$$

The proof is straightforward, and involves some combinatorics; see [35].

We now consider two subgroups of C_k that have important geometric interpretations. The first subgroup consists of k-chains that map to zero under the boundary operator. This group is the *cycle group*, denoted Z_k — it is the kernel of ∂_k and its elements are called k-cycles. The second is the group of k-chains that bound a $k+1$-chain. This is the *boundary group*, B_k — it is the image of ∂_{k+1}. It follows from (3.2) that every boundary is a cycle, i.e. B_k is a subgroup of Z_k.

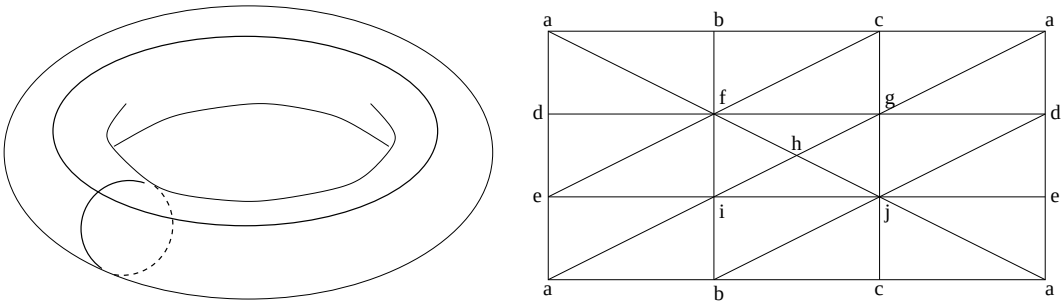
Since $B_k \subset Z_k$, we can form the quotient group, $H_k = Z_k/B_k$. This is the precisely the *homology group*. The elements of H_k are equivalence classes of k-cycles that do not bound any $k+1$ chain — this is how homology characterizes k-dimensional holes. Formally, two k-cycles $z_k^1, z_k^2 \in Z_k$ are in the same equivalence class if $z_k^1 - z_k^2 \in B_k$. Such cycles are said to be homologous. We write $[z_k] \in H_k$ for the equivalence class of cycles homologous to z_k.

The homology groups of a finite simplicial complex are finitely generated abelian groups, so the following theorem tells us about their general structure.

Theorem 6 (Munkres, Thm 4.3). *If G is a finitely generated abelian group then it is isomorphic to the following direct sum:*

$$G \simeq (\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}) \oplus \mathbb{Z}/t_1 \oplus \cdots \oplus \mathbb{Z}/t_m. \tag{3.3}$$

The number of copies of the integer group $\mathbb{Z}$ is called the *Betti number* β. The cyclic groups $\mathbb{Z}/t_i$ are called the *torsion subgroups* and the t_i are the *torsion coefficients.* The torsion coefficients have the property that $t_i > 1$ and t_1 divides t_2 which divides t_3 and so on. The torsion coefficients of $H_k(\mathcal{C})$ measure the twistedness of the space in some sense. The Betti number of the kth homology group H_k is denoted β_k. For $k \geq 1$, β_k is the number non-equivalent non-bounding k-cycles and this can be interpreted as the number of k-dimensional holes. As we mentioned earlier, β_0 counts the number of path-connected components of $|\mathcal{C}|$. The Betti numbers are therefore exactly the type of information we seek.



(a) (b)

Figure 3.4: (a) The torus has two non-equivalent, non-bounding 1-cycles as shown. The homology groups are $H_0 = \mathbb{Z}$, $H_1 = \mathbb{Z} \oplus \mathbb{Z}$, and $H_2 = \mathbb{Z}$. (b) A triangulation of the torus. The left and right edges of the rectangle are identified, and the top and bottom edges.

Examples

Two simple examples are the circle, Figure 3.2, and the torus, Figure 3.4. The circle is homeomorphic to the triangle, and the chain groups have the following bases:

$$\begin{aligned} C_2: &\quad \{0\} \\ C_1: &\quad \{[ab], [bc], [ac]\} \\ C_0: &\quad \{a, b, c\}. \end{aligned}$$

There is a single generator for the 1-cycles, $[ab] + [bc] - [ac]$ and it is not boundary of a 2-chain. The circle is path-connected and there are no 2-simplices, so the homology groups are:

$$\begin{aligned} H_0 =& \mathbb{Z} \\ H_1 =& \mathbb{Z} \\ H_2 =& \{0\}. \end{aligned}$$

The torus is triangulated by the simplicial complex in Figure 3.4(b). It has two non-homologous 1-cycles, $[ab] + [bc] + [ca]$ and $[ae] + [ed] + [da]$. These correspond to the loops in Figure 3.4(a). There is also a non-bounding 2-cycle, σ equal to the sum of all the 2-simplices. The homology groups are therefore:

$$\begin{aligned} H_0 =& \mathbb{Z} \\ H_1 =& \mathbb{Z} \oplus \mathbb{Z} \\ H_2 =& \mathbb{Z} \\ H_3 =& \{0\}. \end{aligned}$$

Smith normal form

There is a well defined algorithm for computing the homology groups of a given simplicial complex. This algorithm is based on finding the *Smith normal form* (SNF) for a matrix representation of the boundary operators. Recall that the oriented k-simplices form a basis for the kth chain group, C_k. This means it is possible to represent the boundary operator, $\partial_k : C_k \to C_{k-1}$,

by a matrix with entries in $\{0, 1, -1\}$. We denote the SNF of the boundary matrix by D_k. If m_k is the number of k-simplices then D_k has m_k columns and m_{k-1} rows.

The algorithm to reduce an integer matrix to SNF is similar to Gaussian elimination, but at all stages the entries remain integers. The resulting matrix has the form:

$$D_k = \begin{bmatrix} B_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \text{where} \quad B_k = \begin{bmatrix} b_1 & & 0 \\ & \ddots & \\ 0 & & b_{l_k} \end{bmatrix}. \tag{3.4}$$

The nonzero entries satisfy $b_i \geq 1$ and b_1 divides b_2, divides b_3, and so on. For a full description of the basic algorithm see Munkres [61].

The SNF matrices for ∂_{k+1} and ∂_k give a complete characterization of the kth homology group, H_k. The torsion coefficients of H_k are the diagonal entries, b_i, of D_{k+1} that are greater than one. The rank of the cycle group, Z_k, is the number of zero columns of D_k, i.e., $m_k - l_k$. The rank of the boundary group, B_k, is the number of non-zero rows of D_{k+1}, i.e., l_{k+1}. The kth Betti number is therefore

$$\beta_k = \operatorname{rank}(Z_k) - \operatorname{rank}(B_k) = m_k - l_k - l_{k+1}.$$

Bases for Z_k and B_k (and hence H_k) are determined by the row operations used in the SNF reduction.

There are two practical problems with the algorithm for reducing a matrix to SNF as it is described in Munkres [61]. First, the time-cost of the algorithm is of a high polynomial degree in the number of simplices; second, the entries of the intermediate matrices typically become extremely large and create numerical problems. Devising algorithms that overcome these problems is an area of active research. When only the Betti numbers are required, it is possible to do better. In fact, if we construct the homology groups over the rationals, rather than the integers, then we need only apply Gaussian elimination to diagonalize the boundary operator matrices — a process that requires on the order of n^3 arithmetic operations. Doing this means we lose all information about the torsion, however. We discuss some other approaches to computational homology in Section 3.4.

3.2.2 The role of homotopy in homology

The study of homotopy leads to a substantial branch of algebraic topology. In this section we give some elementary definitions that are necessary for considering equivalence classes of maps between spaces and the corresponding homomorphisms of homology groups. For more details, see Munkres [61] or Hocking and Young [35].

Homotopy equivalence

Two maps or two spaces are homotopy equivalent if there is a continuous deformation from one to the other. This type of equivalence is usually easy to visualize and gives us a powerful tool for computing homology groups. We start by defining a homotopy between two continuous maps.

Let X and Y be any topological spaces. Two maps $f, g : X \to Y$ are homotopic if their images, $f[X]$ and $g[X]$, can be continuously deformed into one another. Formally, $f, g : X \to Y$ are *homotopic* if there is a mapping $F : X \times \mathbb{I} \to Y$ such that for each $x \in X$, $F(x, 0) = f(x)$ and $F(x, 1) = g(x)$. The map F is called a *homotopy* between f and g, and $X \times \mathbb{I}$ is the

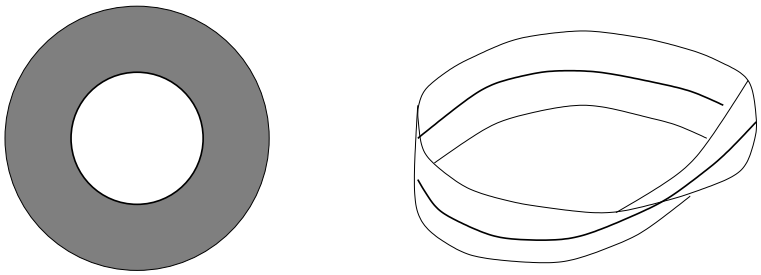


Figure 3.5: An annulus and a Möbius band are both homotopy equivalent to a circle.

homotopy cylinder. We use the notation $f \simeq g$ when f is homotopic to g. For example, let $f, g : S^1 = [0, 2\pi) \to [-1, 1]$, be defined as $f(\theta) = 0$ and $g(\theta) = \sin\theta$. A homotopy between f and g is given by $F(\theta, t) = t \sin\theta$. If Y is a convex subset of $\mathbb{R}^n$, and X is an arbitrary space, then any two functions $f, g : X \to Y$ are homotopic via $F(x, t) = tg(x) + (1 - t)f(x)$ [77]. In general, however, it can be a difficult problem to find a homotopy between two maps, if one exists.

Two spaces, X and Y have the same homotopy type, or are *homotopically equivalent*, if there are mappings $f : X \to Y$ and $g : Y \to X$ such that $fg : Y \to Y$ is homotopic to the identity on Y and $gf : X \to X$ is homotopic to the identity on X (note that throughout this chapter, the composition of two functions, f and g is written as fg). We can show that the unit circle, S^1, and the annulus, A, have the same homotopy type as follows. Let

$$S^1 = \{(r, \theta) \mid r = 1\} \quad \text{and} \quad A = \{(r, \theta) \mid 1 \leq r \leq 2\}$$

Let $f : S^1 \to A$ be the inclusion map $f(1, \theta) = (1, \theta)$, and $g : A \to S^1$ map all points with the same angle to the corresponding point on the unit circle: $g(r, \theta) = (1, \theta)$. Then $gf : S^1 \to S^1$ is given by $gf(1, \theta) = (1, \theta)$, which is exactly the identity map. The other composition is $fg : A \to A$ is $fg(r, \theta) = (1, \theta)$. This is homotopic to the identity, $i_A = (r, \theta)$, via the homotopy $F(r, \theta, t) = (1 - t(r - 1), \theta)$.

The following theorem is one of the central results in homology theory. The proof uses constructions from the following section; see [61] for details.

Theorem 7. *Homotopically equivalent spaces have isomorphic homology groups.*

This means that homology classifies spaces up to homotopy equivalence. Homeomorphic spaces are necessarily homotopically equivalent, since if $h : X \to Y$ is a homeomorphism, then it has an inverse, $h^{-1} : Y \to X$, and by definition $hh^{-1} = i_Y$, and $h^{-1}h = i_X$. This fact, together with Theorem 7, imply that homology groups are topological invariants. It is not the case that homotopy equivalent spaces are homeomorphic. For example, a circle is homotopy equivalent to an annulus and to the Möbius band, but no two of these spaces are homeomorphic; see Figure 3.5. Theorem 7 is also a useful tool for determining homology groups: if a given space is homotopy equivalent to a simpler one, then we need only find homology groups for the latter.

Induced Homomorphisms

When we study the system of ϵ-neighborhoods in Section 3.3, we need to consider how continuous mappings from one neighborhood into another induce homomorphisms on their homology

groups. A large amount of machinery has been developed for studying this type of problem. We give a brief summary here and refer to Munkres [61] for more details.

Suppose K and L are two simplicial complexes and consider a continuous map of their underlying spaces $f : |K'| \to |L|$. The first step is to make a continuous piecewise linear approximation of f, called a *simplicial approximation*, $h : K \to L$. The simplicial complex K is a subdivision of K', so $|K'| = |K|$. The subdivision is constructed to ensure that h is close to f. "Close" means that given $x \in |K'|$, then $f(x)$ and $h(x)$ lie in the same simplex of L.

The simplicial approximation, h, maps vertices of K to vertices of L and extends linearly to each simplex of K. Such a simplicial map, induces homomorphisms on the chain groups $h_\sharp : C_k(K) \to C_k(L)$ in a natural way. If $\sigma^k = [x_0, \ldots, x_k]$ is an oriented k-simplex of K, then we define

$$\begin{aligned} h_\sharp(\sigma^k) =& [h(x_0), \ldots, h(x_k)] \quad \text{if} \quad [h(x_0), \ldots, h(x_k)] \in L, \\ h_\sharp(\sigma^k) =& 0 \quad \text{otherwise.} \end{aligned}$$

This definition is extended linearly to all chains $c_k \in C_k(K)$.

The crucial property of the chain map $h_\sharp$ is that it commutes with the boundary operator, i.e., $\partial(h_\sharp(c_k)) = h_\sharp(\partial(c_k))$. This implies that $h_\sharp$ maps cycles in $Z_k(K)$ to cycles in $Z_k(L)$ and also boundaries to boundaries. This in turn means that $h_\sharp$ induces a homomorphism of the homology groups, $h_* : H_k(K) \to H_k(L)$. This homomorphism is defined on equivalence classes of k-cycles $[z_k] \in H_k(K)$ by: $h_*([z_k]) = [h_\sharp(z_k)]$. See [35] for a proof that this definition satisfies the properties of a homomorphism.

In the special case that $h : |K| \to |K|$ is the identity map, then h_* is the identity homomorphism. If $h : |K| \to |L|$ is a homeomorphism of the underlying spaces, then h_* is an isomorphism of the corresponding homology groups.

For a continuous map, f, any simplicial approximation to f induces the same group homomorphism, denoted f_*. The result we need for Sections 3.2.4 and 3.3 is that homotopic maps induce identical homomorphisms.

Theorem 8. *Suppose $f, g : |K| \to |L|$ are two continuous maps and that f is homotopic to g. Then they induce the same homomorphism of the homology groups, i.e., $f_* = g_*$.*

This is a standard result in homology; see [61] for a proof.

3.2.3 Inverse systems

In order to generalize homology to spaces with infinite structure we need to use a limit. There are two ways to generalize the notion of limit to general index sets and spaces: direct and inverse limit systems. Our problem involves the latter and we describe the associated definitions below; see Hocking and Young [35] or Spanier [77] for more details.

An *inverse system* of topological spaces consists of collection of spaces, X_λ, indexed by a directed set $(\Lambda, \succeq)$, and continuous mappings $p_{\lambda\mu} : X_\mu \to X_\lambda$, for each pair $\mu \succeq \lambda$. The maps are called *bonding morphisms* and must satisfy the following two conditions.

$$p_{\lambda\lambda} = 1_{X_\lambda}, \quad \text{the identity map on} \quad X_\lambda, \quad \text{and} \tag{3.5}$$

$$p_{\lambda\mu} p_{\mu\nu} = p_{\lambda\nu}, \quad \text{for any choice of} \quad \nu \succeq \mu \succeq \lambda. \tag{3.6}$$

Notice that these maps act against the order relation, which is why the system is an "inverse" one. We write $\mathbf{X} = (X_\lambda, p_{\lambda\mu}, \Lambda)$ for the inverse system, or $(X_\lambda, p_{\lambda\mu})$ when the index set is clear from context.

The *inverse limit space*, $\lim_{\leftarrow} \mathbf{X}$, is the subspace of $\Pi_\Lambda X_\lambda$ consisting of all "preorbits" in $\mathbf{X}$. That is,

$$\lim_{\leftarrow} \mathbf{X} = \{(x_\lambda) \mid x_\lambda \in X_\lambda \text{ and } p_{\lambda\mu}(x_\mu) = x_\lambda \text{ for } \mu \succeq \lambda\}. \tag{3.7}$$

The *projections*, $p_\mu : \lim_{\leftarrow} \mathbf{X} \to X_\mu$, are the continuous maps $p_\mu((x_\lambda)) = x_\mu$. In fact, the inverse limit space is defined up to an isomorphism: any space that is homeomorphic to $\lim_{\leftarrow} \mathbf{X}$ is also an inverse limit of $\mathbf{X}$.

We can often simplify the inverse limit calculation by using a subset of the index set, $\Lambda' \subset \Lambda$. The subset must be *cofinal* for this restriction to work, i.e., given any $\lambda \in \Lambda$, there is an element $\mu \in \Lambda'$ with $\mu \succeq \lambda$. We then have the following theorem [52].

Theorem 9. *If Λ' is a cofinal subset of Λ, then the inverse systems $(X_\lambda, p_{\lambda\mu}, \Lambda)$ and $(X_\lambda, p_{\lambda\mu}, \Lambda')$ have isomorphic limits — i.e., their inverse limit spaces are homeomorphic.*

As an example of an inverse system, suppose the index set is the non-negative integers, $\mathbb{N}$, and let $X_0 = [0, 1]$, $X_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ and X_k be the level-k approximation to the middle-third Cantor set (c.f., Section 2.2). Since $X_{k+1} \subset X_k$, for all k, the bonding morphisms are just inclusion maps:

$$p_{kj} : X_j \hookrightarrow X_k \quad \text{when} \quad j > k.$$

It is easy to show that these maps satisfy the two conditions (3.5) and (3.6). The inverse system is represented by the diagram:

$$\cdots \xrightarrow{p_{k+1}} X_k \xrightarrow{p_k} X_{k-1} \cdots X_2 \xrightarrow{p_2} X_1 \xrightarrow{p_1} X_0.$$

The inverse limit space consists of sequences (x_k) such that $x_k \in X_k$ for all k, and $p_{kj}(x_j) = x_k$. Since the projections are inclusion maps, this means $x_k = x_j$. It follows that a sequence in the inverse limit space just repeats the same point, and this point must be in the Cantor set. Therefore, the inverse limit space is homeomorphic to the middle-third Cantor set.

The concept of inverse limit system also exists for the category of groups. In this case the bonding morphisms are group homomorphisms and the inverse limit is a subgroup of the direct sum $\bigoplus_\Lambda G_\lambda$ of the groups in the inverse system. We use this in the next section when we define Čech homology by relating an inverse system of topological spaces and the corresponding inverse system of homology groups.

3.2.4 Čech homology

Čech homology is a general homology theory that can capture the infinitely detailed structure of a fractal. It agrees with simplicial homology on finite simplicial complexes. For more background material, see Hocking and Young [35] or Mardesic and Segal [52].

The nerve of a cover

The Čech approach is based on a way to generate simplicial complexes by taking the *nerve of a cover*. Given a compact Hausdorff space, X, let $\Sigma(X)$ denote the family of all finite open

coverings of X (a covering, $\mathcal{U} \in \Sigma(X)$, is a collection of open sets whose union contains X). We construct an abstract simplicial complex, called the nerve of the cover, by associating each open set, $U \in \mathcal{U}$ with a vertex, also labelled U, in the complex. An edge exists between two vertices U and V if and only if the corresponding open sets intersect. Higher-dimensional simplices $[U_0 \dots U_k]$ are included if the intersection $\bigcap_{i=0}^{k} U_i$ is non-empty. Note that although the space X could be low dimensional, the nerve of a given covering could contain many high-dimensional simplices. In fact, given any finite simplicial complex K and a compact, perfect, Hausdorff space X, it is possible to construct a covering of X whose nerve is isomorphic to K [35]. This ambiguity is removed by considering an inverse limit system.

A partial order relation, $\succeq$, that makes $\Sigma(X)$ a directed set is given by the notion of *refinement*. The cover $\mathcal{V}$ is a refinement of $\mathcal{U}$, denoted $\mathcal{V} \succeq \mathcal{U}$, if for any set $V \in \mathcal{V}$, there is a set $U \in \mathcal{U}$ such that $V \subset U$. The family of all covers with this ordering is the index set for the Čech inverse system; we now define the bonding morphisms.

If $\mathcal{V}$ refines $\mathcal{U}$, we define a projection map $p_{\mathcal{U}\mathcal{V}} : \mathcal{V} \to \mathcal{U}$ by taking the image of a set $V \in \mathcal{V}$, $p_{\mathcal{U}\mathcal{V}}(V)$, to be any fixed element $U \in \mathcal{U}$ such that $V \subset U$. The definition of the projection map is not unique since there may be more than one choice of U for a given V. However, any two choices of projection from $\mathcal{V}$ into $\mathcal{U}$ are homotopic. Therefore, we define the bonding morphisms to be homotopy equivalence classes of projections $p_{\mathcal{U}\mathcal{V}}$, and these "H-maps" then satisfy the conditions (3.5) and (3.6). The inverse system $(\mathcal{U}, p_{\mathcal{U}\mathcal{V}}, \Sigma(X))$ is referred to as the Čech system. Next, we consider the corresponding inverse system of homology groups.

The Čech homology groups

The nerve of a finite cover $\mathcal{U}$ is a simplicial complex, so we can compute its homology groups by the usual techniques. There is one slight difference, however, the coefficient group G must be more general than the integers; for example, we could use the rational or real numbers. The homology groups of the nerves with coefficients in G are denoted $H_k(\mathcal{U}, G)$. The projection maps $p_{\mathcal{U}\mathcal{V}}$ are identified with simplicial maps of the nerves of $\mathcal{U}$ and $\mathcal{V}$. They therefore induce homomorphisms on the homology groups: $p_{\mathcal{U}\mathcal{V}*} : H_k(\mathcal{V}, G) \to H_k(\mathcal{U}, G)$. Theorem 8 implies that two elements of the homotopy class of $p_{\mathcal{U}\mathcal{V}}$ induce the same homomorphism on the homology groups. The inverse system $(H_k(\mathcal{U}, G), p_{\mathcal{U}\mathcal{V}*}, \Sigma(X))$ is called the kth Čech homology system. Finally, the kth Čech homology group with coefficients in G is denoted $\check{H}_k(X, G)$ and is defined to be the inverse limit of the kth Čech homology system.

Geometrically, suppose there is a non-bounding k-cycle z_k in $H_k(\mathcal{U}, G)$ for some $\mathcal{U} \in \Sigma(X)$. Then this k-cycle exists in the limit only if given any refinement $\mathcal{V}$ of $\mathcal{U}$, there is a k-cycle $z'_k \in H_k(\mathcal{V}, G)$ whose image under the projection homomorphism is z_k, i.e., $p_{\mathcal{U}\mathcal{V}*}(z'_k) = z_k$.

The Čech system for all finite covers of the space X is overwhelmingly large for computational purposes; this is why cofinal subsets are useful. A subcollection, $\Sigma'(X) \subset \Sigma(X)$, is cofinal in $\Sigma(X)$ if given any cover $\mathcal{U} \in \Sigma(X)$, there is a cover $\mathcal{V} \in \Sigma'(X)$ such that $\mathcal{V}$ is a refinement of $\mathcal{U}$. As an example, if we have a compact space $X \subset \mathbb{R}^d$, let $\mathcal{U}_n$ be a finite cover of X by open balls with diameter $1/n$. The collection of all such coverings for $n \in \mathbb{Z}$ forms a cofinal subset of $\Sigma(X)$. We can now set up a Čech system for this subfamily of coverings and compute the corresponding inverse system of homology groups. By Theorem 9, the inverse limit of this system is isomorphic to the inverse limit of the full Čech system. Thus, using a cofinal family of coverings greatly reduces the amount of work needed to "compute" the Čech homology groups. The use of cofinal families is also important in our computational work.

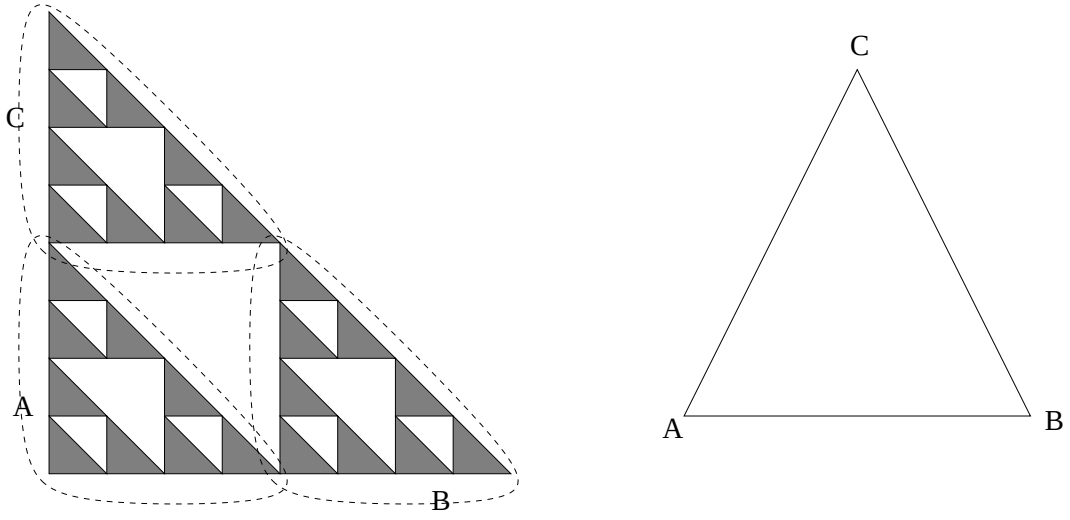


Figure 3.6: The level-1 cover of the Sierpinski triangle and the nerve of this cover. The higher level covers repeated this pattern on smaller scales.

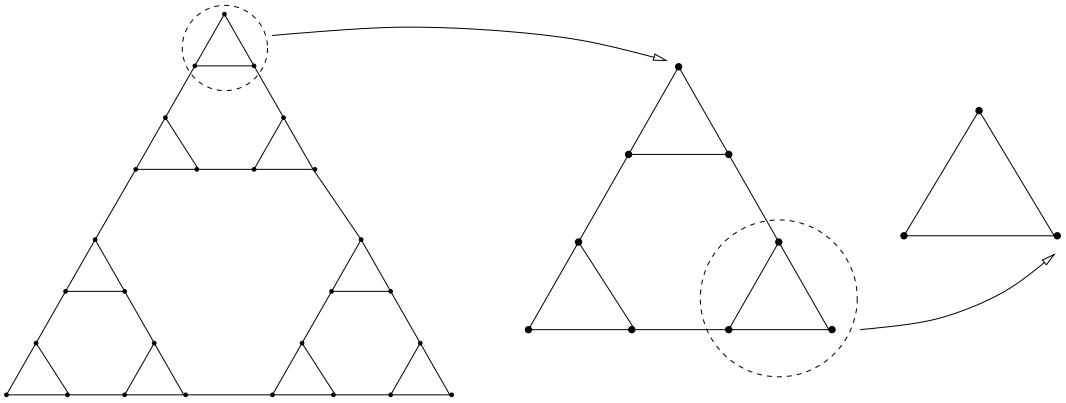


Figure 3.7: Nerves of the level-3, level-2, and level-1 covers of the Sierpinski triangle. The arrows show the action of the projection maps on the circled triples of points.

An example

We use a cofinal sequence of covers of the Sierpinski triangle, T, to compute its Čech homology groups. This example is considered further in Section 3.5.

For $i = 0, 1, 2, 3, \ldots$, let $\mathcal{U}_i$ be a cofinal sequence of covers of T with the following properties. The sets in each cover are identical open convex regions with different centers. There are 3^i such open sets in $\mathcal{U}_i$ such that for each $U_i \in \mathcal{U}_i$ there are exactly three sets $U_{i+1}, V_{i+1}, W_{i+1} \in \mathcal{U}_{i+1}$ that map into U_i, and each point of the Sierpinski triangle belongs to at most two of the sets in $\mathcal{U}_i$. See Figure 3.6 for an illustration of the level-1 cover, $\mathcal{U}_1$ and its nerve. The last condition implies that the nerve of $\mathcal{U}_i$ has only 0-simplices and 1-simplices. The nerves for $\mathcal{U}_1$, $\mathcal{U}_2$, and $\mathcal{U}_3$ are given in Figure 3.7. This figure also shows the action of the bonding morphisms, $p_{i,i+1} : \mathcal{U}_{i+1} \to \mathcal{U}_i$, which are defined by the inclusion of the three sets $U_{i+1}, V_{i+1}, W_{i+1} \in \mathcal{U}_{i+1}$ into $U_i \in \mathcal{U}_i$.

The simple, one-dimensional nature of the nerves means that we can write down the homology groups by inspection. The zeroth and second homology groups are the same for all

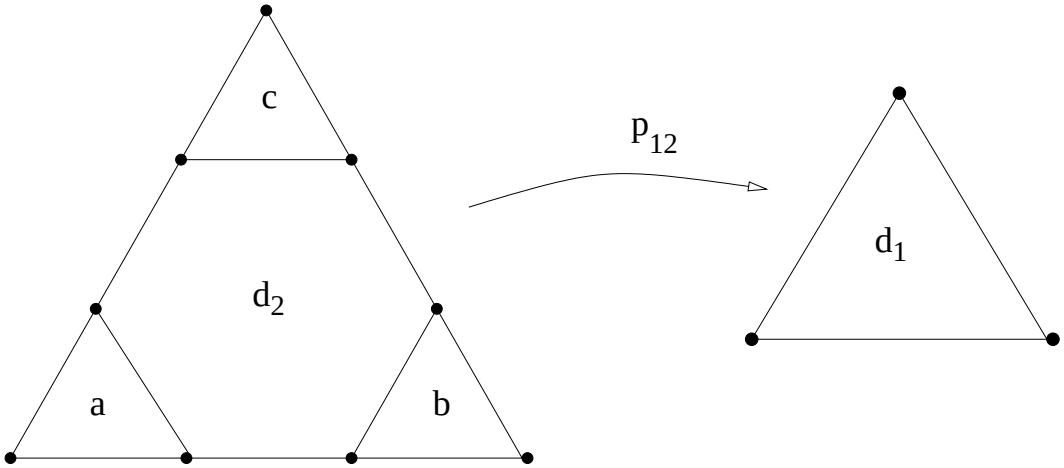


Figure 3.8: The projection map from $\mathcal{U}_2$ into $\mathcal{U}_1$.

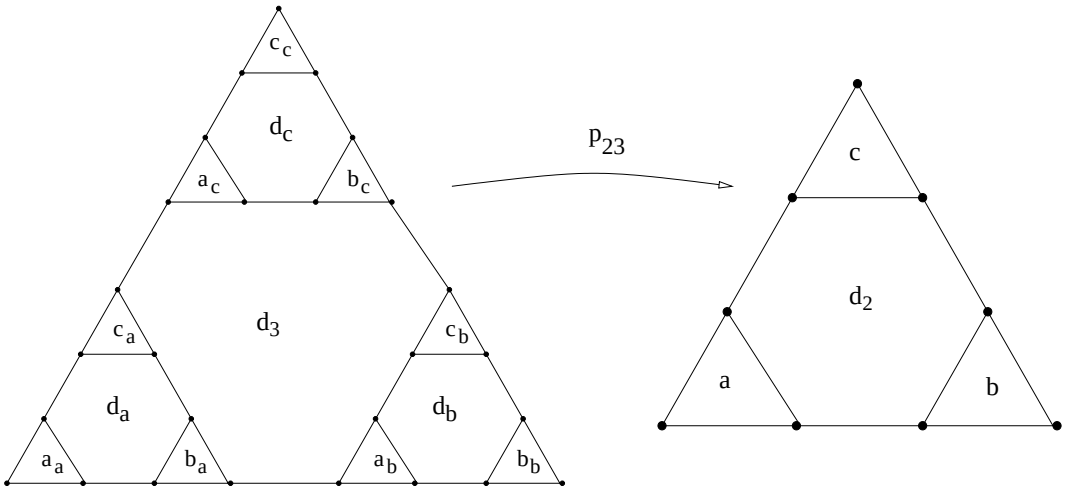


Figure 3.9: The projection map from $\mathcal{U}_3$ into $\mathcal{U}_2$.

i:

$$H_0(\mathcal{U}_i, G) \simeq G \quad \text{for} \quad i = 0, 1, 2, \ldots$$
$$H_2(\mathcal{U}_i, G) \simeq \{0\} \quad \text{for} \quad i = 0, 1, 2, \ldots$$

The first order homology groups have the following pattern:

$$\begin{aligned}
&H_1(\mathcal{U}_0, G) \simeq \{0\} \\
&H_1(\mathcal{U}_1, G) \simeq G \\
&H_1(\mathcal{U}_2, G) \simeq G \oplus G \oplus G \oplus G = G^4 \\
&H_1(\mathcal{U}_3, G) \simeq G^{13} \\
&H_1(\mathcal{U}_i, G) \simeq G^{n_i} \quad \text{where} \quad n_i = \sum_{k=0}^{i-1} 3^k = \frac{1}{2}(3^i - 1).
\end{aligned}$$

We consider the action of the inclusion maps on the first-order homology only, starting with p_{12*}. The action of this map is shown in Figure 3.8. The basis for $H_1(\mathcal{U}_1, G)$ is the loop d_1, and a basis for $H_1(\mathcal{U}_2, G)$ is given by the four loops, $\{a, b, c, d_2\}$. Since p_{12} always maps three points into one, the corner loops, a, b, c, are each collapsed to a single point. The action of $p_{12*} : H_1(\mathcal{U}_2, G) \to H_1(\mathcal{U}_1, G)$ is therefore

$$\begin{aligned}
a &\longmapsto 0 \\
b &\longmapsto 0 \\
c &\longmapsto 0 \\
d_2 &\longmapsto d_1.
\end{aligned}$$

A matrix representation of p_{12*} relative to the bases above is

$$P_{12} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}.$$

The above action is repeated in a threefold way for $p_{23*} : H_1(\mathcal{U}_3, G) \to H_1(\mathcal{U}_2, G)$, see Figure 3.9. We order the basis for $H_1(\mathcal{U}_3, G)$ as follows:

$$\{a_a, b_a, c_a, d_a, a_b, \ldots, d_b, a_c, \ldots, d_c, d_d\}.$$

The action of p_{23*} is

$$\begin{aligned}
d_a &\longmapsto a \\
d_b &\longmapsto b \\
d_c &\longmapsto c \\
d_3 &\longmapsto d_2.
\end{aligned}$$

All other loops in $H_1(\mathcal{U}_3, G)$ map to $\{0\}$. The matrix for this map is:

$$P_{23} = \begin{bmatrix} 0 & 0 & 0 & 1 & & & & & & & & & \\ & & & & 0 & 0 & 0 & 1 & & & & & \\ & & & & & & & & 0 & 0 & 0 & 1 & \\ & & & & & & & & & & & & 1 \end{bmatrix}.$$

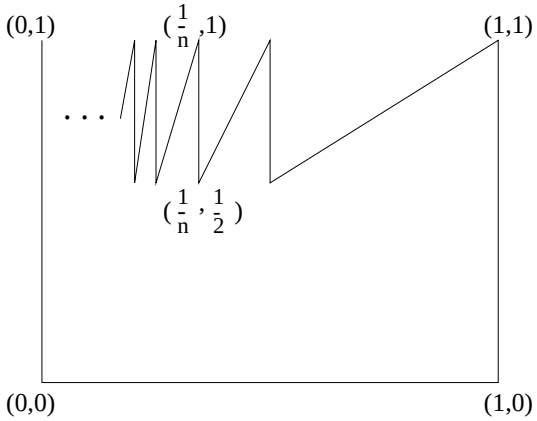


Figure 3.10: The Warsaw circle has trivial simplicial homology group, $H_1 = 0$, but its Čech homology group is $\check{H}_1 = G$. It has the same "shape" as the circle because they are both inverse limits of a sequence of annuli.

This pattern continues for higher levels, and for $p_{i,i+1*} : H_1(\mathcal{U}_{i+1}, G) \to H_1(\mathcal{U}_i, G)$ we have the recursive form for the matrix

$$P_{i,i+1} = \begin{bmatrix} P_{i-1,i} & & & \\ & P_{i-1,i} & & \\ & & P_{i-1,i} & \\ & & & 1 \end{bmatrix}.$$

This means that the first Čech homology group for the Sierpinski triangle, $\check{H}_1(T, G)$ is an infinitely generated abelian group. The elements of $\check{H}_1(T, G)$ consist of sequences, (z_i), of elements, $z_i \in H_1(\mathcal{U}_i, G)$, such that $p_{i,i+1*}(z_{i+1}) = z_i$ for all $i = 1, 2, 3, \ldots$. For example, the central hole in the Sierpinski triangle would be represented by the sequence $(d_i) = (d_1, d_2, d_3, \ldots)$, and the a-loop of $\mathcal{U}_2$ by $(a_i) = (0, a, d_a, \ldots)$. Smaller holes have a longer initial string of zeros.

The next section generalizes Čech homology by considering inverse limit systems of approximating spaces other than nerves of covers.

3.2.5 Shape theory

Shape theory generalizes the concept of homotopy equivalence by considering inverse systems of "nice" approximating spaces. As an example, the unit circle has the same shape as the Warsaw circle of Figure 3.10, because both are inverse limits of sequences of annuli, although they have different homotopy and simplicial homology groups.

The essential result we use from shape theory is that every compact metric space is homeomorphic to the limit of an inverse system in the category of finite polyhedra and homotopy equivalence classes of maps (H-maps). The Čech system for a compact space is an example of such an inverse system. Shape theory generalizes the Čech approach by allowing the approximating spaces to be homotopy equivalent to a finite polyhedron. The following theorem is a compilation of results from Mardesic and Segal [52, App.1]; it characterizes spaces that are homotopy equivalent to finite polyhedra.

Theorem 10. *A topological space with the homotopy type of a compact CW-complex or a compact ANR is homotopy equivalent to a compact polyhedron.*

CW-complexes are general cell-complexes, and ANR stands for absolute neighborhood retract; see [35] for definitions. ANRs are an important class of space; they have many useful properties — one of these is that the identity map extends to a neighborhood of the space (this is the neighborhood retract part of ANR). The unit cube in $\mathbb{R}^n$ and the n-sphere are examples of ANRs.

The algebraic side of shape theory associates an inverse system of homology groups with the inverse system of finite polyhedra, in much the same manner as Čech homology. That is, if $(P_\lambda, p_{\lambda\mu})$ is an inverse system of polyhedra, then $(H_k(P_\lambda), p_{\lambda\mu *})$ is the corresponding inverse system of homology groups. A third result we need from shape theory is:

Theorem 11. *If $(P_\lambda, p_{\lambda\mu})$ and $(Q_\lambda, q_{\lambda\mu})$ are inverse systems of polyhedra (or ANRs or CW-complexes) and H-maps whose inverse limit spaces are homotopy equivalent, then the corresponding inverse systems of homology groups have isomorphic limits.*

For a compact space, X, the Čech system is an inverse system of polyhedra. The above theorem therefore implies that any other inverse system of polyhedra yields an inverse system of homology groups whose limit isomorphic to Čech homology. This is the sense in which shape theory generalizes Čech homology.

In the following section we consider an inverse system of closed ϵ-neighborhoods and inclusion maps for a compact space X.

3.3 Foundations for computing homology

In this section we develop theoretical foundations for understanding how homology groups computed from data relate to the homology of the space they approximate. The setting for our analysis is as follows. We assume that the underlying space, X, is a compact subset of a metric space, (M, d), and that the finite set of points, $S \subset M$, approximates X in a metric sense, i.e., each point of X is within distance ρ of some point in S and vice versa. In other words, ρ is the Hausdorff distance between X and S: $d_H(X, S) = \rho$. In a given application, this assumption must be justified by physical or numerical arguments; we give a number of different examples in Chapter 4. A small value of ρ implies S is a good approximation to X. Typically, ρ depends on the number of points in S, as we demonstrated with examples in Chapter 2, and more points naturally result in a smaller value of ρ. In some applications, ρ could represent the magnitude of noise present in the data, or a discretization error.

To give the compact space, X, and its finite approximation, S, comparable topological structure, we form their closed ϵ-neighborhoods:

$$X_\epsilon = \{x \in M \mid d(x, X) \leq \epsilon\} \quad \text{and} \quad S_\epsilon = \{x \in M \mid d(x, S) \leq \epsilon\}.$$

Roughly speaking, since X and S are within ρ, their ϵ-neighborhoods should have similar properties for $\epsilon > \rho$. We make this precise in Section 3.3.4 using an inverse system framework from shape theory. Since the homology of X_ϵ converges to the homology of X as $\epsilon \to 0$ in an inverse limit sense, we hope to extrapolate the homology of X from the homology of the ϵ-neighborhoods of the data, S, for $\epsilon > \rho$. Of course, extrapolation is never guaranteed to give the correct answer, and we are always restricted by the inherent accuracy of the finite data.

3.3.1 The inverse system of ϵ-neighborhoods

We begin by describing the inverse system of ϵ-neighborhoods for the underlying compact set X. The spaces for the inverse system are the closed ϵ-neighborhoods, $X_\epsilon = \{x \mid d(x, X) \leq \epsilon\}$,

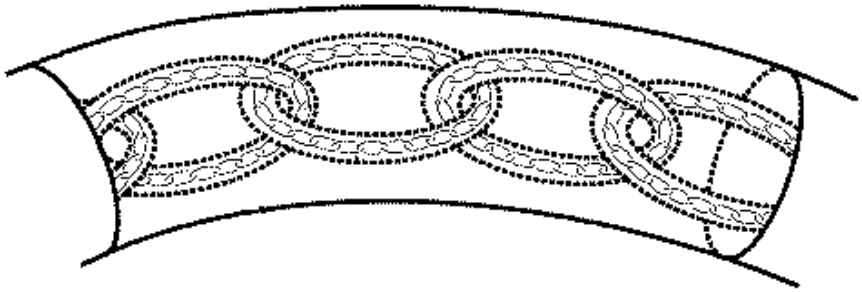


Figure 3.11: A segment of Antoine's necklace. This picture is from Eric Weisstein's World of Mathematics [85]. http://mathworld.wolfram.com/AntoinesNecklace.html

indexed by $0 < \epsilon \leq \epsilon_0$. Since we are interested in the limit as $\epsilon \to 0$, the order relation is an inverted one: $\lambda \succeq \epsilon$ when $\lambda \leq \epsilon$. Since $X_\lambda \subset X_\epsilon$ when $\lambda \leq \epsilon$, the bonding morphisms are simply inclusion maps $p_{\epsilon\lambda} : X_\lambda \to X_\epsilon$, which easily satisfy the conditions (3.5) and (3.6). The inverse limit space is homeomorphic to X and the projections $p_\epsilon : X \to X_\epsilon$ are again inclusion maps.

Each ϵ-neighborhood of X is an ANR and therefore has the homotopy type of a finite polyhedron. This means the kth simplicial homology groups $H_k(X_\epsilon)$ are well-defined for $\epsilon > 0$. The inclusion maps $p_{\epsilon\lambda} : X_\lambda \to X_\epsilon$ induce homomorphisms on the homology groups in the standard way described in Section 3.2.2. We write $p_{\epsilon\lambda *} : H_k(X_\lambda) \to H_k(X_\epsilon)$ for these induced homomorphisms. The homology groups, together with the inclusion-induced homomorphisms, yield inverse systems of groups, denoted by $H_k(\mathbf{X})$. Results from shape theory show that the inverse limit of $H_k(\mathbf{X})$ is isomorphic to the kth Čech homology group $\check{H}_k(X)$; for details see [52, p.121].

For computational purposes, we typically use a cofinal sequence of ϵ-neighborhoods, X_{ϵ_i}, where $\epsilon_1 \geq \epsilon_2 \geq \ldots$ is a decreasing sequence of ϵ-values with $\epsilon_i \to 0$.

3.3.2 Persistent Betti numbers

We would like to quantify the structure of X by looking at the Betti numbers $\beta_k(X_\epsilon) = \operatorname{rank} H_k(X_\epsilon)$ as $\epsilon \to 0$. In general though, it is not the case that $\beta_k(X_\epsilon) \to \beta_k(X)$ as $\epsilon \to 0$, i.e.,

$$\lim_{\epsilon \to 0} \beta_k(X_\epsilon) \neq \beta_k(\lim_{\epsilon \to 0} X_\epsilon).$$

As an example, consider Antoine's necklace, Figure 3.11. This Cantor set, A, is constructed by taking the intersection of a sequence of nested and linked solid tori. The zeroth level, A_0, is a single solid torus, which is homotopy equivalent to a circle and therefore has first Betti number, $\beta_1(A_0) = 1$. A chain of N linked solid tori is embedded in A_0 to give the first level approximation, A_1. This process is repeated inside each one of these tori so that the ith term in the sequence, A_i, consists of N^i links. Antoine's necklace is then $A = \bigcap A_i$. It is possible to choose a sequence of ϵ-values so that $A_{\epsilon_i} \simeq A_i$. We then have $\beta_1(A_i) = N^i \to \infty$. However, since the limit A is a Cantor set, $\beta_1(A) = 0$. The problem stems from ignoring the role of the bonding morphisms. We now describe how to incorporate this information.

The essential point is that we only want to count holes in an ϵ-neighborhood that are generated by a hole in the underlying space, not holes that are caused after fattening the set to its

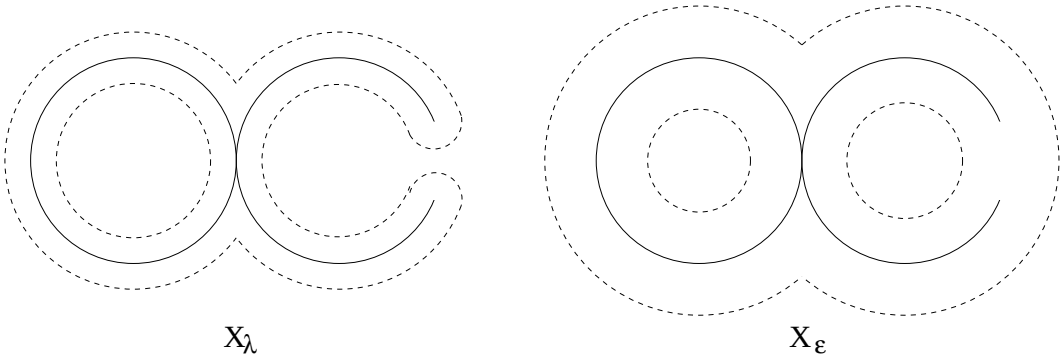


Figure 3.12: An illustration of the difference between persistent and non-persistent holes. The underlying space, X, is made of an 'O' and a 'C' that touch at a point. In the larger ϵ-neighborhood on the right, the hole from the 'O' has a preimage in the smaller λ-neighborhood. The hole from the 'C' is not the image of a hole in X, or X_λ.

ϵ-neighborhood. This idea is illustrated in Figure 3.12. The smaller λ-neighborhood captures the homology of the underlying space, which has a single hole (i.e., $\beta_1 = 1$). For the larger ϵ-value, the neighborhood has two holes; one corresponds to a genuine hole in X, and the other does not. In terms of homology, we want to count only those cycles in $H_k(X_\epsilon)$ that are the image of some cycle in $H_k(X_\lambda)$, for all $\lambda < \epsilon$.

More formally, for $\lambda < \epsilon$, we say that an equivalence class of cycles $[z_\epsilon] \in H_k(X_\epsilon)$ *persists in* $H_k(X_\lambda)$ if it is in the image of the bonding homomorphism $[z_\epsilon] \in p_{\epsilon\lambda *}(H_k(X_\lambda))$. The number of holes in X_ϵ that persist in X_λ is therefore just the rank of the image subgroup:

$$\beta_k^\lambda(X_\epsilon) = \mathrm{rank}(p_{\epsilon\lambda *}(H_k(X_\lambda))). \tag{3.8}$$

Since we want to know the topology of X, we are also interested in the quantity

$$\beta_k^0(X_\epsilon) = \mathrm{rank}(p_{\epsilon *}(H_k(X))). \tag{3.9}$$

We refer to $\beta_k^\lambda(X_\epsilon)$ for $\lambda \geq 0$, as the *persistent Betti numbers*.

For Antoine's Necklace, we have that $\beta_1^{i+1}(A_i) = 0$ for all i, because there is no cycle in A_{i+1} that maps onto a link in A_i. The projection from A into A_i gives $\beta_1^0(A_i) = 0$ for all i, so $\lim_{i\to\infty} \beta_1^0(A_i) = \beta_1(A) = 0$.

The persistent Betti number is an integer-valued function of two real numbers, $\lambda \leq \epsilon$. In order to understand the properties of this function, we give some elementary bounds on $\beta_k^\lambda(X_\epsilon)$. From the definition, it follows that the persistent Betti number is less than the regular Betti number for both X_λ and X_ϵ:

$$\beta_k^\lambda(X_\epsilon) \leq \beta_k(X_\epsilon) \qquad \text{and} \tag{3.10}$$

$$\beta_k^\lambda(X_\epsilon) \leq \beta_k(X_\lambda). \tag{3.11}$$

The next two inequalities say that for a fixed ϵ-neighborhood, $\beta_k^\lambda(X_\epsilon)$ is a monotonic non-decreasing function of λ; while for a fixed λ, the persistent Betti number is a non-increasing function of ϵ.

$$\text{For} \quad \nu < \lambda < \epsilon, \quad \beta_k^\nu(X_\epsilon) \leq \beta_k^\lambda(X_\epsilon). \tag{3.12}$$

$$\text{For} \quad \lambda < \epsilon < \mu, \quad \beta_k^\lambda(X_\mu) \leq \beta_k^\lambda(X_\epsilon). \tag{3.13}$$

Roughly speaking, increasing the difference between λ and ϵ decreases the number of persistent holes.

The above inequalities are a first step towards understanding the continuity properties of $\beta_k^\lambda(X_\epsilon)$, but more work remains to be done. Most importantly, we want conditions on X that guarantee

$$\beta_k^0(X_\epsilon) \to \beta_k(X) \quad \text{as} \quad \epsilon \to 0. \tag{3.14}$$

If $\beta_k(X)$ is finite and (3.14) holds, then there must be an $\epsilon_0 > 0$ such that $\beta_k^0(X_\epsilon) = \beta_k(X)$ for $\epsilon \in [0, \epsilon_0)$. In addition to this, our computational work will be most effective for spaces where there is a λ_0 such that $\beta_k^\lambda(X_\epsilon) = \beta_k^0(X_\epsilon)$ for $\lambda \in [0, \lambda_0]$. We do not expect these conditions to hold for an arbitrary compact space. However, for special cases such as the iterated function system attractors in Section 3.5, it may be possible to say something more concrete about the continuity of persistent Betti numbers as λ and ϵ tend to zero.

3.3.3 Growth rates for persistent Betti numbers

In Chapter 2, we quantified the rate of growth in the number of connected components, $C(\epsilon)$ by the disconnectedness index, γ. We can do the same thing for the k-dimensional holes, as counted by the persistent Betti numbers. If $\beta_k^0(X_\epsilon) \to \infty$ as $\epsilon \to 0$, we quantify the rate of divergence by assuming an asymptotic power law, $\beta_k^0(X_\epsilon) \sim \epsilon^{\gamma_k}$. The exponent γ_k can be computed as the following limit (when it exists)

$$\gamma_k = \lim_{\epsilon \to 0} \frac{\log \beta_k^0(X_\epsilon)}{\log(1/\epsilon)}. \tag{3.15}$$

If the limit does not exist, then we use the limsup or liminf.

Recall that for $k = 0$, the Betti number is just the number of connected components, so the definition of γ_0 agrees with that for the disconnectedness index, γ. The 1-dimensional holes counted by β_1^0 are really loops, and the 2-dimensional holes are spherical voids like those in Swiss cheese, so we might call γ_1 the loopiness index and γ_2 the holiness. Results in Chapter 5 show that for subsets of $\mathbb{R}^2$, there is a relationship between γ_1 and the fractal dimension; the examples in Section 3.5 confirm this.

Simple Examples

We can determine the growth rates, γ_k, analytically for simple, self-similar fractal such as the Sierpinski triangle, Sierpinski curve (or carpet), Menger sponge and so on. We already considered the homology groups for the Sierpinski triangle, T, in Section 3.2.4. It is possible to choose a sequence of ϵ-values so that the homology groups $H_1(T_{\epsilon_i}) = H_1(\mathcal{U}_i)$. Specifically, let r be the radius of the largest circle inscribed by the triangular hole with vertices at $(\frac{1}{2}, 0)$, $(0, \frac{1}{2})$ and $(\frac{1}{2}, \frac{1}{2})$, so that $r = \frac{1}{2} - \frac{\sqrt{2}}{4} \approx 0.14645$. Since the next largest hole has radius $r/2$, we choose $r/2 < \epsilon_1 < r$ and $\epsilon_{i+1} = \epsilon_i/2$. Every element, $[z_1] \in H_1(T_{\epsilon_i})$, has a preimage under the projection map, $p_{\epsilon_i *} : \check{H}_1(T) \to H_1(T_{\epsilon_i})$, so the persistent Betti number is the same as the regular Betti number. From our previous calculations, then, $\beta_1^0(T_{\epsilon_i}) = n_i = (3^i - 1)/2$. Thus,

$$\gamma_1 = \lim_{\epsilon \to 0} \frac{\log(\beta_1^0(T_{\epsilon_i}))}{\log(1/\epsilon_i)} = \lim_{i \to \infty} \frac{\log((3^i - 1)/2)}{\log 2^i/\epsilon_1} = \frac{\log 3}{\log 2}. \tag{3.16}$$

Not surprisingly, this number is the same as the similarity dimension.

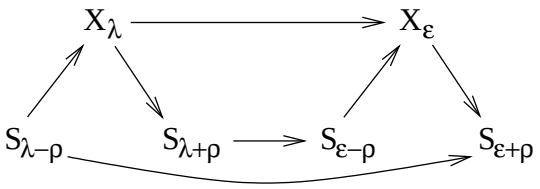
3.3.4 Finite approximations

We now analyze the finite approximation, S, and its relationship to the compact set X. As described in the introduction to this section, we assume that the Hausdorff distance, $d_H(X, S) = \rho$, i.e., that $X \subset S_\rho$ and $S \subset X_\rho$. We keep the notation, $(X_\epsilon, p_{\lambda\epsilon})$, for the inverse systems of ϵ-neighborhoods of X and write $(S_\epsilon, q_{\lambda\epsilon})$ for the system of neighborhoods of S. The inclusions of S into X_ρ and X into S_ρ give a formal relationship between these inverse systems and allow us to derive bounds on the persistent Betti numbers of X_ϵ in terms of ρ-persistent Betti numbers of S_ϵ.

Since $X \subset S_\rho$ and $S \subset X_\rho$, we have the following inclusion maps for any $\epsilon > 0$,

$$i_\epsilon : S_\epsilon \to X_{\epsilon+\rho} \quad \text{and} \quad j_\epsilon : X_\epsilon \to S_{\epsilon+\rho}.$$

With appropriate bonding morphisms, this gives us a commuting diagram for any $\lambda \geq \rho$ and $\epsilon \geq \lambda + 2\rho$:



(3.17)

This diagram implies that if an element of $H_k(X_\epsilon)$ has a preimage in $H_k(X_\lambda)$, there must be an element of $H_k(S_{\epsilon-\rho})$ with a preimage in $H_k(S_{\lambda+\rho})$. Similarly, if $S_{\epsilon+\rho}$ has a hole that persists in $S_{\lambda-\rho}$, there must be a corresponding hole in X_ϵ that persists in X_λ. In terms of Betti numbers, we have that for $\lambda \geq \rho$ and $\epsilon \geq \lambda + 2\rho$,

$$\beta_k^{\lambda-\rho}(S_{\epsilon+\rho}) \leq \beta_k^{\lambda}(X_\epsilon) \leq \beta_k^{\lambda+\rho}(S_{\epsilon-\rho}). \tag{3.18}$$

We can swap the roles of X and S in (3.17) to obtain analogous bounds on $\beta_k^\lambda(S_\epsilon)$:

$$\beta_k^{\lambda-\rho}(X_{\epsilon+\rho}) \leq \beta_k^{\lambda}(S_\epsilon) \leq \beta_k^{\lambda+\rho}(X_{\epsilon-\rho}). \tag{3.19}$$

Since X maps into S_ρ, we can only hope to get information about X from holes in S_ϵ that persist in S_ρ. Setting $\lambda = \rho$ in (3.19), we have that for $\epsilon \geq \rho$,

$$\beta_k^0(X_{\epsilon+\rho}) \leq \beta_k^\rho(S_\epsilon). \tag{3.20}$$

Thus, if $\beta_k^0(X_\epsilon) \sim \epsilon^{\gamma_k}$ as $\epsilon \to 0$ and ρ is small enough, we should see at least that order of growth in $\beta_k^\rho(S_\epsilon)$.

In one sense, ρ is the optimal resolution for coarse-graining the data to estimate the topological structure of the underlying space. This could imply that $\beta_k(S_\rho)$ is the best approximation to $\beta_k(X)$, which would render our multiresolution approach redundant. For very simple spaces this may be the case. However, we are interested in more complicated settings. In general, the inverse system of ϵ-neighborhoods and the persistent Betti numbers offer two advantages. First, the cutoff resolution ρ is typically not known in advance and must be determined from the data; examining S_ϵ at many ϵ-values helps us estimate ρ. Second, computing the Betti numbers $\beta_k(S_\rho)$ at a single resolution does not distinguish between holes due to the topology of X and holes in S_ρ induced by the geometry. The sequence of persistent Betti numbers $\beta_k^\rho(S_\epsilon)$ for $\epsilon > \rho$ give a more accurate basis than $\beta_k(S_\rho)$ from which to extrapolate topological information about X (c.f., the example of Antoine's necklace). Such an extrapolation must always be given with respect to the cutoff resolution ρ, however, since it is possible that the topological properties of X change at resolutions below ρ.

3.3.5 Computing persistent Betti numbers

We want to compute the number of holes in S_ϵ that persist in S_ρ, that is, $\beta_k^\rho(S_\epsilon) = \operatorname{rank} q_{\epsilon\rho *}(H_k(S_\rho))$. This definition is in terms of the rank of a linear operator, $q_{\epsilon\rho *}$, on the two quotient spaces, $H_k(S_\rho)$ and $H_k(S_\epsilon)$. Below, we derive an equivalent formula that is given in terms of the linear map on the chain groups, $q_{\epsilon\rho\sharp} : C_k(S_\rho) \to C_k(S_\epsilon)$. Recall that the k-simplices form a basis for the kth chain group, so it is easy to write down a matrix representation for this map.

To simplify notation a little, we write:

$$\begin{aligned}
C_\rho &\quad \text{for the chain group} \quad C_k(S_\rho),\\
C_\epsilon &\quad \text{for} \quad C_k(S_\epsilon),\\
q_\sharp &\quad \text{for the chain map} \quad q_{\epsilon\rho\sharp},\\
q_* &\quad \text{for the homology homomorphism} \quad q_{\epsilon\rho *},\\
Z_\rho, Z_\epsilon &\quad \text{for the cycle groups}\\
B_\rho, B_\epsilon &\quad \text{for the boundary groups, and}\\
H_\rho = Z_\rho/B_\rho &\quad \text{and}\\
H_\epsilon = Z_\epsilon/B_\epsilon &\quad \text{for the homology groups.}
\end{aligned}$$

Our starting point is the image group, $q_*(H_\rho) \subset H_\epsilon$. One of the fundamental theorems about homomorphisms of groups [34] is that the image of a group, G, under a homomorphism, ϕ, is isomorphic to the quotient, $G/\ker\phi$. Thus,

$$q_*(H_\rho) \simeq H_\rho/\ker q_*.$$

The kernel, $\ker q_*$, contains all elements of H_ρ that map to the zero element of H_ϵ. That is,

$$\ker q_* = \{[z] \in H_\rho \mid [q_\sharp(z)] = [0] \in H_\epsilon\}.$$

But a cycle is homologous to 0 if and only if it is in the boundary group, so

$$\ker q_* \simeq Z_\rho \cap q_\sharp^{-1}(B_\epsilon).$$

Now, recall from Section 3.2.2 that a homomorphism induced by a simplicial map commutes with the boundary operator. It follows that cycles map to cycles and boundaries to boundaries; i.e.,

$$q_\sharp(Z_\rho) \subset Z_\epsilon \quad \text{and} \quad q_\sharp(B_\rho) \subset B_\epsilon.$$

The second expression implies that $B_\rho \subset q_\sharp^{-1}(B_\epsilon)$, so $B_\rho \subset Z_\rho \cap q_\sharp^{-1}(B_\epsilon)$. It follows that the quotient

$$q_*(H_\rho) \simeq H_\rho/\ker q_* \simeq [Z_\rho/B_\rho]/[Z_\rho \cap q_\sharp^{-1}(B_\epsilon)] \simeq Z_\rho/[Z_\rho \cap q_\sharp^{-1}(B_\epsilon)].$$

Finally, this means that the persistent Betti number is

$$\beta_k^\rho(\epsilon) = \operatorname{rank}[Z_\rho] - \operatorname{rank}[Z_\rho \cap q_\sharp^{-1}(B_\epsilon)]. \tag{3.21}$$

In terms of actual computation, then, we can find the persistent Betti number from the dimensions of null spaces and ranges of matrix representations for the boundary operator and the inclusion maps. Specifically, $\operatorname{rank} Z_k(\rho)$ is the dimension of the null space of the matrix for

$\partial_k(\rho) : C_k(\rho) \to C_{k-1}(\rho)$. For the second term in (3.21) we need to find the intersection of two spaces: $q_\sharp[\text{null}(\partial_k(\rho))]$, and B_ϵ, which is the range of $\partial_{k+1}(\epsilon) : C_{k+1}(\epsilon) \to C_k(\epsilon)$. Finding the intersection of two linear subspaces requires some information about their bases, and is therefore a more difficult problem than computing dimensions. The standard algorithms for determining intersections typically run in $O(n^3)$ time, where n is the largest dimension of the matrices involved [29]. As we discuss in the following section, we have not yet implemented these algorithms. Instead, for the examples of Section 3.5 we use pre-existing software for computing the regular Betti numbers from subsets of a triangulation described in Section 3.4.1. These examples illustrate why the regular Betti numbers are insufficient for extrapolation.

3.4 Implementation

Our goal is to take a finite cloud of points S as input, and compute persistent Betti numbers as a function of a resolution parameter. There are four parts to the overall process:

1. For a sequence of ϵ-values, generate simplicial complexes that triangulate the ϵ-neighborhoods of the data.

2. Estimate the cutoff resolution ρ.

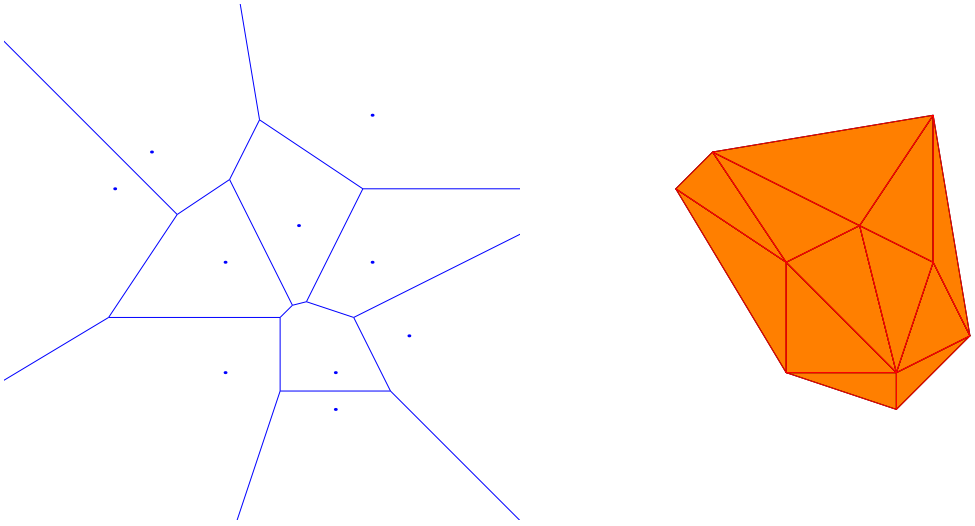
3. Compute persistent Betti numbers, $\beta_k^\rho(\epsilon)$, for $\epsilon > \rho$.

4. If appropriate, compute the growth rate, γ_k.

This is essentially the same approach we used in Chapter 2 in computing the number of connected components. For that case, however, a simplicial complex is unnecessary — all the information about ϵ-connected components of S is encoded in the Euclidean minimal spanning tree of the data.

In the present context, steps 1 and 3 are the most computationally intensive. The two steps are also closely related; efficient algorithms for computing Betti numbers make explicit use of the data structures involved in building the complexes. Thus, given an ϵ-neighborhood, S_ϵ, the first problem is to generate a simplicial complex, $\mathcal{C}_\epsilon$, whose underlying space is at least homotopy equivalent to S_ϵ. Since we are interested in the inverse system of ϵ-neighborhoods, we need simplicial complexes for a sequence of numbers $\epsilon_i \to 0$. In order to have inclusion maps that are well defined, we need $\mathcal{C}_{\epsilon_i}$ to be either a *subcomplex* or a *subdivision* of $\mathcal{C}_{\epsilon_j}$ when $\epsilon_i < \epsilon_j$. A subcomplex approach to this problem due to Edelsbrunner *et al.* [18], is described in detail in Section 3.4.1. This group has also developed a fast incremental algorithm for computing Betti numbers of complexes in $\mathbb{R}^2$ or $\mathbb{R}^3$. We use their implementations for the examples in Section 3.5.

We use the same criterion for Step 2 that we derived in Chapter 2 for approximations to perfect spaces. Since a perfect space has no isolated points, we estimate the cutoff resolution as the largest value of ϵ for which S_ϵ has at least one isolated point. This underestimates the value for which $S_\rho \supset X$, but the examples in Section 3.5 show it to be a reasonably good approximation. Recall that isolated points are straightforward to detect numerically — a point is ϵ-isolated if the distance from it to every other point in the set is greater than ϵ. The computation of growth rates in Step 4 is straightforward once the persistent Betti numbers are found.



(a) (b)

Figure 3.13: (a) The Voronoi diagram. (b) The Delaunay triangulation.

3.4.1 Alpha shapes

As we mentioned above, alpha shapes are a sequence of simplicial complexes derived from the Voronoi diagram and the Delaunay triangulation (DT) — fundamental constructions in computational geometry. Alpha shapes were first defined by Edelsbrunner *et al.* for a finite set of points in the plane [17] and later generalized to higher dimensions [18].

The Voronoi Diagram and the Delaunay Triangulation

The Voronoi diagram of a finite set of points represents the region of influence of each point; an example is given in Figure 3.13(a). The graph has a wide range of applications in computational geometry (finding nearest neighbors, for example), in biology (assigning areas of influence to individual trees), and modelling crystal growth, to name a few. The Delaunay complex is closely related to the Voronoi diagram, and is equally important in computational geometry. We give the basic definitions of these graphs here. For more details about their properties, and algorithms for computing them, see [67]

Given a finite set of points, $S \subset \mathbb{R}^d$, we define the *Voronoi cell*, $V(p)$, of $p \in S$ to be the set of all points closer to p than any other point in S:

$$V(p) = \{x \in \mathbb{R}^d \mid d(x,p) \leq d(x,q), \quad \text{for} \quad q \in S - p\}. \tag{3.22}$$

The collection of Voronoi cells is the *Voronoi diagram*. The cells are closed convex regions, and the union $\bigcup_{p \in S} V(p) = \mathbb{R}^d$. If $V(p) \cap V(q) \neq \emptyset$, then their intersection is a subset of the hyperplane that is perpendicular to, and bisects the edge pq; the interiors of the Voronoi cells are disjoint.

The *Delaunay complex*, $\mathcal{T}$, is defined to be the geometric dual of the Voronoi diagram; see Figure 3.13(b) for an example in the plane. The *geometric dual* is a similar construction to the nerve of a cover: if the cells $V(p_0), \ldots, V(p_k)$ have non-empty intersection, then the convex hull of $p_0, \ldots, p_k$, is their dual. The nerve of $V(p_0), \ldots, V(p_k)$, however, is the k-simplex

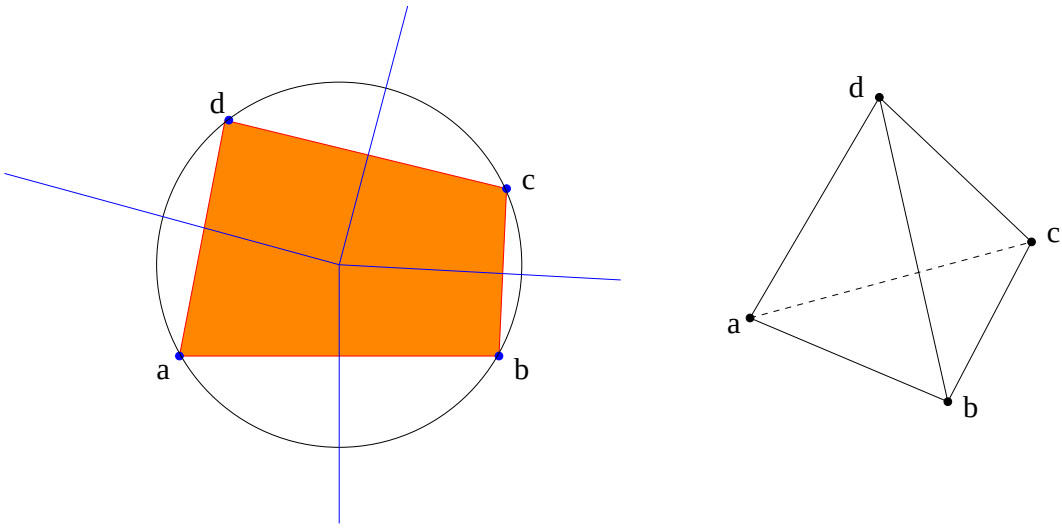


Figure 3.14: The points, a, b, c, d are not in general position because they simultaneously lie on a circle. The four Voronoi cells intersect at the center of this circle which means the Delaunay complex contains the quadrilateral $abcd$. The nerve of the Voronoi complex contains the tetrahedron $[abcd]$.

$[p_0, \ldots, p_k]$ — the difference is illustrated in Figure 3.14. In this example, four cells intersect at a point, so the geometric dual is a quadrilateral, whereas the nerve contains a tetrahedron. This situation is a degenerate one; it only occurs when four points simultaneously lie on a circle. To exclude this degeneracy, the points in S are typically assumed to be in *general position.* This imposes the condition that no $(d+2)$ points in S lie on the boundary of a d-sphere (we discuss the practicality of this assumption later). When S satisfies the general position condition, a point in $\mathbb{R}^d$ can belong to at most $d+1$ different Voronoi cells and it follows that the Delaunay complex is a simplicial complex. This is why the Delaunay complex is usually referred to as the *Delaunay triangulation.* Note that the underlying space of the Delaunay complex, $|\mathcal{T}|$, is the convex hull of S.

A consequence of the above discussion is that when the points are in general position, the Delaunay triangulation is a geometric realization of the nerve of the Voronoi diagram. This correspondence is used to define the alpha complexes.

Alpha complexes

We now describe how Edelsbrunner introduces the resolution parameter α into the Voronoi diagram and Delaunay triangulation. This operation generates a sequence of simplicial complexes that triangulate the α-neighborhoods of the data set S. The parameter α is exactly the same as our parameter ϵ; we switch to α in this section to be consistent with the original papers.

The closed α-neighborhood S_α is just the union of all closed balls of radius α with centers in S:

$$S_\alpha = \bigcup_{p \in S} B_\alpha(p), \quad \text{where} \quad B_\alpha(p) = \{x \mid d(x,p) \leq \alpha\}.$$

These α-balls therefore form a cover of S_α. We could take the nerve of this cover, but the corresponding simplicial complex is likely to have simplices of a higher dimension than the

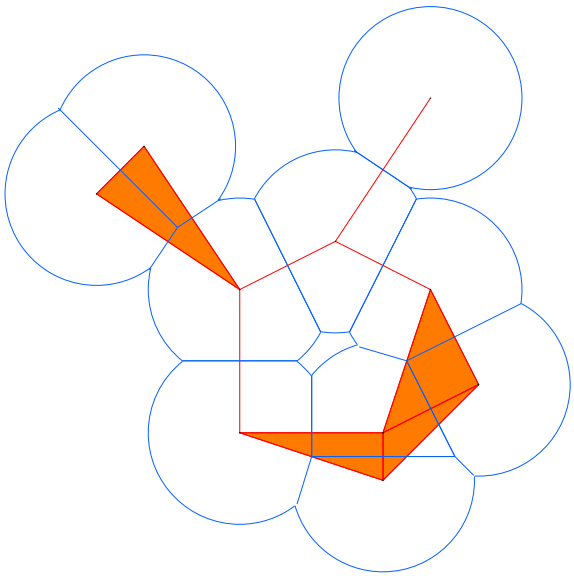


Figure 3.15: The union of alpha balls is partitioned by the Voronoi cells (in blue). The alpha complex (in orange) is a subset of the Delaunay triangulation that is homotopy equivalent to the α-neighborhood.

ambient space of S. A better approach is to take the intersection of the α-balls with the Voronoi cells:

$$V_\alpha(p) = V(p) \cap B_\alpha(p).$$

This gives us a cover of S_α by closed convex regions, called the *alpha diagram*:

$$S_\alpha = \bigcup_{p \in S} V_\alpha(p).$$

If the points of S are in general position, then the nerve of the alpha diagram is a subset of the Delaunay triangulation, called the *alpha complex*, $\mathcal{C}_\alpha$. This subset relation holds since there is a k-simplex $[p_0, \ldots, p_k]$ in the alpha complex only if $V_\alpha(p_0) \cap \cdots \cap V_\alpha(p_k) \neq \emptyset$. This implies $V(p_0) \cap \cdots \cap V(p_k) \neq \emptyset$, so the simplex $[p_0, \ldots, p_k]$ is in $\mathcal{T}$. See Figure 3.15 for an illustration of the above constructions.

We now give some properties of the alpha complexes. The most important for our purposes is that the underlying space of the alpha complex, $|\mathcal{C}_\alpha|$, is homotopy equivalent to the α-neighborhood, S_α. This means the two spaces have isomorphic homology groups (recall Theorem 7), and therefore the same Betti numbers. A direct proof of this property is given by Edelsbrunner in [16].

The construction of the alpha complexes implies a sequential ordering of simplices that is very useful in developing efficient algorithms. First, by the same argument that $\mathcal{C}_\alpha \subset \mathcal{T}$, we see that if $\alpha < \alpha'$, then $\mathcal{C}_\alpha \subset \mathcal{C}_{\alpha'}$. Also, since the Delaunay triangulation is finite, there are only a finite number of distinct alpha complexes:

$$\emptyset = \mathcal{C}_{\alpha_0}, \mathcal{C}_{\alpha_1}, \ldots, \mathcal{C}_{\alpha_n} = \mathcal{T}.$$

These are ordered by increasing α, and by convention, $\alpha_0 = 0$ and $\mathcal{C}_0 = \emptyset$. This ordering of the alpha complexes induces an ordering of the simplices in $\mathcal{T}$. If $\sigma_0, \sigma_1, \ldots, \sigma_{r'}$ is the ordered list of simplices in $\mathcal{C}_{\alpha_r}$, then $\sigma_{r'+1}, \ldots, \sigma_{r'+j}$ are the j simplices in $\mathcal{C}_{\alpha_{r+1}} - \mathcal{C}_{\alpha_r}$, listed in order of increasing dimension. This sequential ordering of simplices is called a *filter* and the corresponding sequence of complexes is a *filtration.* The algorithm for determining this ordering is based on computing the radius of the circumsphere of each simplex in $\mathcal{T}$; see [18] for details.

Delfinado and Edelsbrunner's algorithm [11] for computing the Betti numbers of a simplicial complex relies on this sequential ordering of simplices. The theoretical underpinnings of their algorithm come from central results in homology theory such as the Mayer-Vietoris sequence and Poincaré duality. The Betti numbers are computed incrementally as each simplex is added to the complex. This process depends on a test to determine whether the new simplex belongs to a k-cycle of the new complex. There are efficient algorithms for testing 1-cycles, and homology-cohomology duality theorems transform the $(d-1)$-cycles into 1-cocycles that are equally easy to test for. However, there is no test for other k-cycles, so Delfinado and Edelsbrunner's algorithm applies only to subcomplexes of $\mathbb{R}^2$ or $\mathbb{R}^3$.

Remarks

The NCSA ftp site provides software that implements all of the above alpha shape constructions in $\mathbb{R}^2$ and $\mathbb{R}^3$ [1]. We generate the Betti number data for the examples in Section 3.5 using this software. The NCSA alpha shape software requires the input data, S, to be in integer format. This reduces some of the standard problems with data structures in computational geometry. The implementation uses a technique of simulated perturbations to cope with any degeneracies, so the restriction that the points of S be in general position is removed [18]. Complexity bounds for the algorithms involved are at worst quadratic in the number of points, n, for both time and storage. The Delaunay triangulation of a set of points in the plane can be found in $O(n \log n)$ time and $O(n)$ storage. For a subset of $\mathbb{R}^3$, the NCSA software uses an incremental flip algorithm which builds the complex in $O(n^2)$ time and storage. The simplices of the Delaunay triangulation are then sorted by the radius of their circumsphere; this process requires $O(m \log m)$ time, where m is the number of simplices. The incremental algorithm for computing the Betti number takes $O(m\alpha(m))$ to find all Betti numbers for all the alpha complexes of a subset of $\mathbb{R}^3$, and is slightly faster for subsets of $\mathbb{R}^2$. The function $\alpha(m)$ is the inverse of Ackermann's function (which is defined by repeated exponentiation) and it therefore has extremely slow growth. This time estimate is common in algorithms involving set operations; see [9].

The NCSA alpha shape implementation goes a long way towards carrying out our desired program. It is not clear how to easily incorporate the computation of persistent Betti numbers. It is possible that an incremental algorithm for finding the persistent Betti numbers exists, as is the case for the regular Betti numbers. However, finding the persistent Betti numbers requires some explicit information about the cycles and boundaries, and the alpha shape algorithm does not generate or record this information. This problem clearly requires further work.

Another drawback of the alpha shape algorithm is a large degree of redundancy in the triangulations for the type of data we are interested in. The Delaunay triangulation builds simplices that involve every single data point, and this generates a much finer complex than is necessary for resolutions $\epsilon \geq \rho$. This redundancy is likely to occur whenever we construct $\mathcal{C}_{\epsilon_i}$ as a subcomplex of $\mathcal{C}_{\epsilon_j}$ for $\epsilon_i < \epsilon_j$. In Section 3.4.3, we discuss a possible alternative approach to building complexes on multiple scales based on subdivisions. In the following section, we outline some other fast algorithms for computing Betti numbers of a single complex.

3.4.2 Other algorithms for computational homology

The development of fast algorithms for computational homology is an active area of research. Many different approaches exist for various specific applications, mostly for complexes in $\mathbb{R}^3$. To the best of our knowledge, the alpha shape algorithm is the only one that gives information at multiple resolutions. In this section, we give a very brief overview of some of the recent literature.

Dey and Guha [14] describe an efficient algorithm for computing Betti numbers and geometric representations of the non-bounding cycles from 3-manifolds in $\mathbb{R}^3$. They then generalize this to include any simplicial complex in $\mathbb{R}^3$ through a process of thickening. They are interested in applications to solid modelling, molecular biology and computer-aided manufacturing. This algorithm finds the Betti numbers with time and storage complexity that are linear in the size of the complex, and the generators for the first and second homology groups with a cost of order $O(n^2 g)$ in time, where g is the maximum genus of the boundary surfaces.

In [37] Kalies *et al.* develop an algorithm for computing the Betti numbers from cubical cell complexes. Their approach is based on a local reduction of the complex to simpler form. In $\mathbb{R}^2$ this reduction is a homotopy equivalence and the result of the reductions is a minimal complex consisting of loops. The number of loops gives the first Betti number. In higher dimensions, the reduced complex may not be minimal and the reduction steps are no longer simple homotopies. Kalies *et al.* conjecture that the Betti numbers can be computed with $O(n \log^3 n)$ operations, where n is the number of cubes in the complex. This code was developed for applications in dynamical systems — specifically, for computing the Conley index of isolating neighborhoods of invariant sets for flows generated by ODEs; see [58] for an example. The generation of cubical covers of such sets is part of the GAIO (global analysis of invariant objects) project [12] and is an efficient way to represent successive approximations to attractors or unstable manifolds, for example.

Finally, we describe an approach due to Friedman [26] which computes the Betti numbers of arbitrary simplicial complexes in $\mathbb{R}^d$. This method is based on a fundamental result of Hodge theory which says that the homology groups (with real or rational coefficients) are isomorphic to the null space of a Laplacian operator on the chain complexes. The Laplacian, $\Delta_k : C_k \to C_k$ is formed from the boundary operators and their transposes:

$$\Delta_k = \partial_{k+1}\partial_{k+1}^* + \partial_k^*\partial_k.$$

Hodge theory implies that the kth Betti number is the dimension of the null space of Δ_k. For simplicial complexes, Friedman constructs a matrix representation of the Laplacian directly, without using the boundary operators explicitly. The matrix for Δ_k is positive semidefinite and symmetric, and typically quite sparse, so it is amenable to fast algorithms for computing ranks and null spaces. To compute the dimension of the null space, Friedman makes a careful application of the power method for finding eigenvalues and eigenvectors. The power method can be inaccurate for large matrices with repeated eigenvalues; Friedman's method attempts to rigorously verify the correctness of the computed Betti numbers. This algorithm holds for simplicial complexes of any finite dimension and the running time is approximately quadratic in the number of simplices. See [26] for detailed complexity bounds.

3.4.3 A better way?

The two main drawbacks to the alpha shape implementation are (1) that the simplicial complexes are finer than they need to be, and (2) the fast algorithm for computing Betti numbers

only holds in $\mathbb{R}^2$ and $\mathbb{R}^3$. The excessive number of simplices in the alpha complexes means that computing the persistent Betti numbers from the formula in Section 3.3.5 is unrealistic. Any subcomplex approach to generating a sequence of complexes at different resolutions will encounter the same problem of having an unnecessarily large number of simplices at coarser resolutions. The remainder of this section sketches an alternative approach to generating complexes at multiple resolutions based on subdivisions.

The advantage of subdivisions is that complexes at coarse resolutions have fewer simplices than those at fine resolutions. Subdivision of simplicial complexes is a common process, both in homology theory and in computational geometry. For applications to general scattered point data, however, cubical complexes are a more natural construction. The cubical complexes we have in mind are essentially regular meshes in $\mathbb{R}^n$. The level-0 complex is a single cube $|\mathcal{C}_0| \supset S$, with side c. This cube is then subdivided into 2^n equal cubes of side $c/2$, and the level-1 complex, $\mathcal{C}_1$, consists of those cubes that contain a point from S. This process is repeated on the cubes of $\mathcal{C}_1$ to get $\mathcal{C}_2$, and so on. The sequence of complexes can be organized into a tree structure; such a multiresolution cubical complex is usually referred to as a quadtree in $\mathbb{R}^2$, and an octree $\mathbb{R}^3$.

This sequence of cubical complexes does not have as close a correspondence with ϵ-neighborhoods as the alpha complexes do, so we need to slightly modify the inverse systems of Section 3.3.1. Given a compact space $X \subset \mathbb{R}^n$, it is possible to construct a sequence of cubical complexes as we described in the previous section. The underlying space, $|\mathcal{C}_i|$, of such a cubical complex still has the homotopy type of a finite polyhedron; $|\mathcal{C}_i| \to X$ in the Hausdorff metric; and for $i > j$, $\mathcal{C}_i$ is a refinement of $\mathcal{C}_j$. These properties should be enough to show that the resulting inverse system of homology groups is isomorphic to Čech homology.

Another substantial difference between the subdivision and subcomplex approaches is in the chain maps induced by inclusion on the underlying spaces. For the sequence of alpha complexes, these chain maps are one-to-one inclusion maps on the complexes. For the cubical complexes, if $\mathcal{C}_i$ is a subdivision of $\mathcal{C}_j$, we still have that $|\mathcal{C}_i| \subset |\mathcal{C}_j|$. The chain maps $q_\sharp$ are no longer one-to-one, however, since each cube in $\mathcal{C}_j$ is subdivided into 2^n cubes, and all of these are possibly in $\mathcal{C}_i$. The inclusion-induced chain map therefore maps all of these cubes onto the larger one. This has implications for the numerical implementation of computing persistent Betti numbers.

Of all the fast Betti number computations described earlier, it seems that Friedman's approach using the Laplacian has the most potential for adaptation to our proposed subcomplex approach. The isomorphism between homology groups and the null space of a Laplacian matrix suggests that in order to compute persistent Betti numbers, we need only find intersections of null spaces of the appropriate Laplacian matrices. This is an easier numerical linear algebra problem than the one implied by the formula in Section 3.3.5. There are still problems to be worked through here. In particular, how to build the Laplacian matrix from cubical complexes (rather than simplicial ones), and how to incorporate the inclusion-induced chain maps between the Laplacian null spaces.

The above ideas are just one possible direction for the development of more efficient algorithms to compute persistent Betti numbers from data. This an open problem that needs a substantial amount of further work.

Table 3.1: Computed values of γ_0 and γ_1. The results are slopes of linear least-squares fits to the data; error bounds are estimated by varying the scaling range. The exact values for the non-zero exponents are the same as the similarity dimension, $\log 3/\log 2 \approx 1.585$.

Data Set	Figure	γ_0	γ_1
Sierpinski triangle	3.16	0	1.59 ± 0.02
Cantor set	3.18	1.40 ± 0.05	0
Simply connected set	3.20	0	0
Disconnected set	3.22	1.42 ± 0.05	0

3.5 Examples

In this section we examine the ϵ-neighborhoods of some simple fractal examples generated by iterated function systems. We use the same relatives of the Sierpinski triangle as in Chapter 2. These examples are chosen because they have well understood topological structure, and it is easy to generate finite point-set approximations to them. We use the alpha shape software described in Section 3.4.1 to compute the regular Betti numbers as a function of resolution. The results demonstrate that growth in the regular Betti numbers can be misleading and that for truly topological information, the persistent Betti numbers are necessary. Since we have not implemented an algorithm for computing the persistent Betti number, we do not have the data to make a comparison. For these simple examples, however, we do know what the persistent Betti numbers should be. This is a very preliminary set of numerical experiments. Further investigations require a more efficient implementation to compute persistent Betti numbers.

3.5.1 Sierpinski triangle relatives revisited

Recall from Section 2.4 that the Sierpinski triangle relatives are attracting fixed sets of a family of iterated function systems:

$$S = f[S] = f_1[S] \cup f_2[S] \cup f_3[S].$$

The functions f_i are similarity transformations of the unit square with a contraction ratio of one half, and they can involve rotations or reflections. The topology of these fractals falls into four classes: simply connected (Figure 3.20), connected (Figure 3.16), totally disconnected (Figure 3.18), and disconnected (Figure 2.13). The techniques of Chapter 2 distinguish between connected and disconnected examples. With the mathematical machinery developed in this chapter, we are now equipped to distinguish between simply connected sets and connected sets with holes.

The points on the fractals are generated in the same manner as in Chapter 2, by applying the transformations f_1, f_2, f_3 in random order to any initial point. For each of the following fractals, we compute the number of components, $\beta_0(\epsilon)$, the number of holes, $\beta_1(\epsilon)$, and the number of isolated points, $I(\epsilon)$, for an approximation with 10^4 points. Where appropriate, we calculate the growth rates, γ_0 and γ_1; the results are compiled in table 3.1. Since $\beta_0(\epsilon)$ counts the number of connected components in the ϵ-neighborhood, rather that the number of ϵ-connected components, we have that $\beta_0(\epsilon) = C(\epsilon/2)$. With this change, the data here agree with the results in Chapter 2.

The Sierpinski triangle

The generating functions for the Sierpinski triangle are:

$$\begin{aligned} f_1(x,y) &= \tfrac{1}{2}(x,y) \\ f_2(x,y) &= \tfrac{1}{2}(x+1,y) \\ f_3(x,y) &= \tfrac{1}{2}(x,y+1). \end{aligned} \tag{3.23}$$

A finite point-set approximation to the triangle, an ϵ-neighborhood and the corresponding subset of the Delaunay triangulation are shown in Figure 3.16. The underlying set is perfect and connected with infinitely many holes, so we should see $\beta_0(\epsilon) = 1$ and $\beta_1(\epsilon) \to \infty$ as $\epsilon \to 0$. As we derived in Section 3.2.4, for $r \approx 0.146$,

$$\begin{aligned} r/2^n < \epsilon_n < r/2^{n-1}, \\ \beta_1(\epsilon_n) = \sum_{k=1}^{n} 3^k = \tfrac{1}{2}(3^n - 1). \end{aligned}$$

The growth rate, $\gamma_1 = \log 3 / \log 2 \approx 1.585$, is the same as the similarity dimension.

These expected results are reflected by the computations of $\beta_0(\epsilon)$, $\beta_1(\epsilon)$ and $I(\epsilon)$ (graphed in Figure 3.17) for the 10^4 point approximation to the triangle. We see that for ϵ above a threshold value, the computed values of $\beta_0(\epsilon)$ and $\beta_1(\epsilon)$ are in close agreement with the theory. The point at which $\beta_0(\epsilon)$ and $\beta_1(\epsilon)$ "blur" is approximately $\epsilon = 0.003$, close to the value at which the number of isolated points, $I(\epsilon)$, becomes positive. This ϵ value is, of course, the cutoff resolution ρ. As we saw in Chapter 2, at finer resolutions — i.e., $\epsilon < \rho$ — there is a sharp transition in the graph of β_0 from one to the number of points in the set, as each point becomes isolated. The graph of β_1 shows that the holes are destroyed as ϵ decreases. This is because the edges that form the loops are eventually deleted from the triangulation. We estimate the slope of the staircase by a linear, least-squares fit and find $\gamma_1 \approx 1.59$. This is very close to the value derived above.

A Cantor set relative

Figure 3.18 shows the attractor for the iterated function system generated by

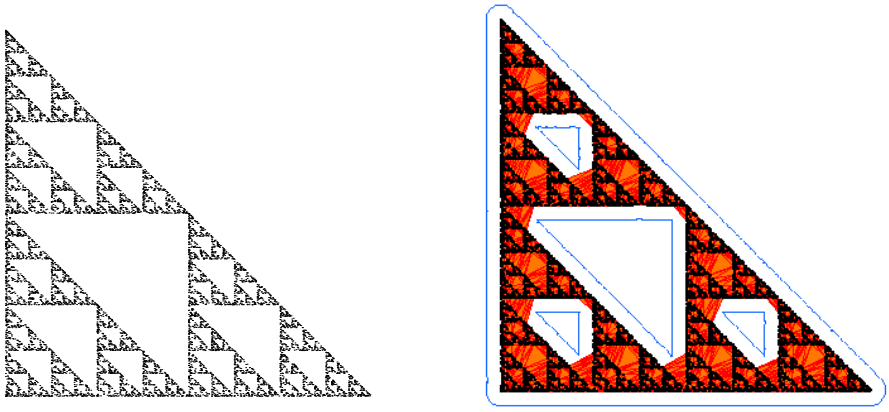
$$\begin{aligned} f_1(x,y) &= \tfrac{1}{2}(-y+1,x) \\ f_2(x,y) &= \tfrac{1}{2}(y+1,x) \\ f_3(x,y) &= \tfrac{1}{2}(y,-x+2). \end{aligned} \tag{3.24}$$

This fractal is a Cantor set, and therefore perfect and totally disconnected, so we expect to see $\beta_0(\epsilon) \to \infty$ as $\epsilon \to 0$ and $\beta_1^0(\epsilon) = 0$. In Chapter 2 we derived the following form for $\beta_0(\epsilon)$:

$$\begin{aligned} \epsilon_0/2^{n+1} < \epsilon_n < \epsilon_0/2^n \\ \beta_0(\epsilon_n) = \begin{cases} 3^n + 2 \cdot 3^{(n-1)/2} & \text{if } n \text{ is odd} \\ 3^n + 3^{n/2} & \text{if } n \text{ is even.} \end{cases} \end{aligned}$$

Here, ϵ_0 is the smallest value of ϵ for which S_ϵ is connected. Recall that the value of γ_0 is $\log 3 / \log 2$, which is the similarity dimension again.

We saw in Chapter 2 that the numerical computations of $\beta_0(\epsilon)$ agree very well with the theoretical values above when ϵ is greater than the cutoff resolution ρ. Here, the value of



(a) (b)

Figure 3.16: (a) 10^4 points on the Sierpinski triangle. (b) An ϵ-neighborhood (blue outline) and corresponding subset of the Delaunay triangulation (orange).

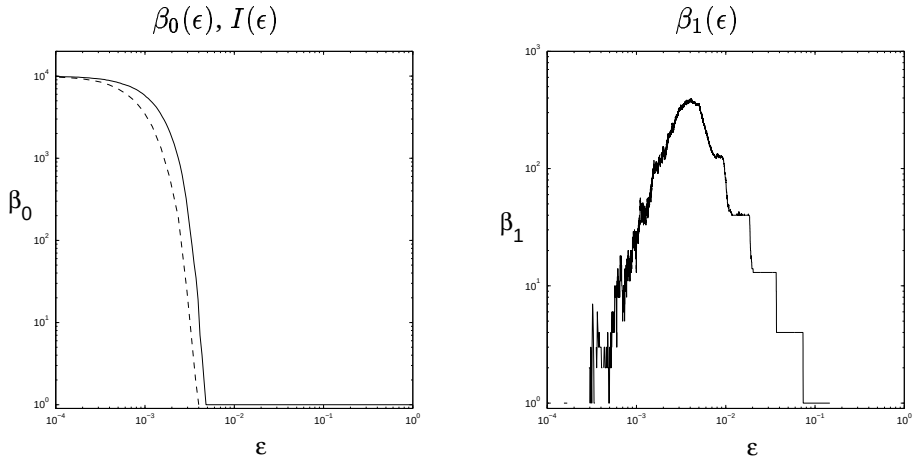
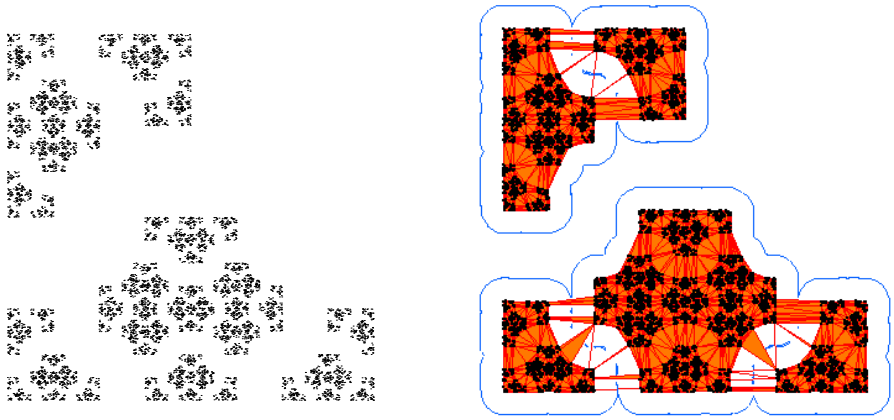


Figure 3.17: Number of components, $\beta_0(\epsilon)$, and number of holes, $\beta_1(\epsilon)$, for 10^4 points uniformly distributed over the Sierpinski triangle. The dashed line in the graph of $\beta_0(\epsilon)$ is the number of isolated points, $I(\epsilon)$. All axes are logarithmic. The horizontal axis range is $10^{-4} < \epsilon < 1$.



(a) (b)

Figure 3.18: (a) 10^4 points on the Cantor set relative. (b) An ϵ-neighborhood and corresponding subset of the Delaunay triangulation.

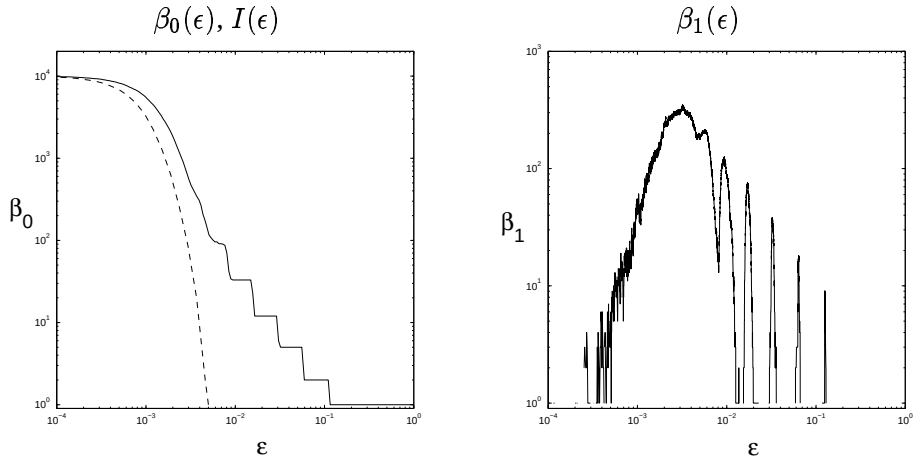


Figure 3.19: Number of components, $\beta_0(\epsilon)$, and number of holes, $\beta_1(\epsilon)$, for 10^4 points uniformly distributed over the Cantor set relative. The dashed line in the graph of $\beta_0(\epsilon)$ is the number of isolated points, $I(\epsilon)$. All axes are logarithmic. The horizontal axis range is $10^{-4} < \epsilon < 1$.

$\rho \approx 0.005$. The graph of $\beta_1(\epsilon)$ has regularly spaced spikes — the reason for this is seen in Figure 3.18(b). The ϵ-neighborhood drawn there has two connected components. Data points in the two components are separated by a distance of at least $2\epsilon_0$. We set the value of ϵ in Figure 3.18(b) to be slightly larger than $\epsilon_0/2$ — the value at which the ϵ-neighborhood breaks into five components. Holes appear in the triangulation at the value of ϵ displayed because of the few remaining edges that bridge the "gap". These edges are deleted at slightly smaller values of ϵ and the holes disappear. A graph of the persistent Betti number, $\beta_1^\rho(\epsilon)$, would not contain these spikes.

A Simply connected relative

A simply connected relative of the Sierpinski triangle, shown in Figure 3.20, is generated by:

$$\begin{aligned} f_1(x,y) &= \tfrac{1}{2}(-x+1,-y+1) \\ f_2(x,y) &= \tfrac{1}{2}(-y+2,x) \\ f_3(x,y) &= \tfrac{1}{2}(x,y+1). \end{aligned} \tag{3.25}$$

We expect to see a single connected component and no persistent holes. However, holes do appear in the ϵ-neighborhoods, as recorded in Figure 3.21. These holes are due to the geometry of the fractal, not its topology. Again, these spikes in the graph of $\beta_1(\epsilon)$ would not appear in a graph of the persistent Betti number.

A relative with infinitely many connected components

A fourth triangle relative, shown in Figure 3.22, is generated by the following similarities:

$$\begin{aligned} f_1(x,y) &= \tfrac{1}{2}(x,y) \\ f_2(x,y) &= \tfrac{1}{2}(y+1,-x+1) \\ f_3(x,y) &= \tfrac{1}{2}(x,y+1). \end{aligned} \tag{3.26}$$

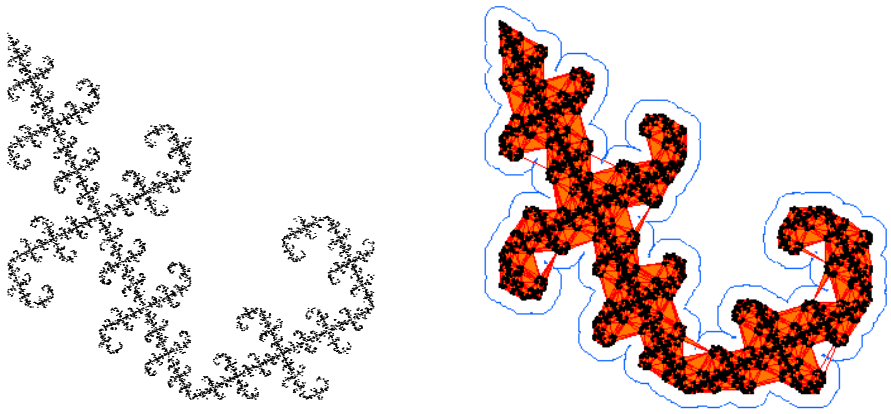
The attractor for this system has infinitely many connected components, yet is not totally disconnected because the components have positive diameters. Thus, we expect $\beta_0(\epsilon) \to \infty$ and $\beta_1^0(\epsilon) = 0$. Again, self-similarity means that for:

$$\begin{aligned} \epsilon_0/2^{n+1} < \epsilon_n < \epsilon_0/2^n, \\ \beta_0(\epsilon_n) = \tfrac{1}{2}(3^{n+1}+1), \end{aligned}$$

giving a growth rate of $\gamma_0 = \log 3/\log 2$. We estimate the slope of the graph of $\beta_0(\epsilon)$ as before and find $\gamma_0 \approx 1.42$. This is lower than the limiting value because of the small range of ϵ for which the computed values of $\beta_0(\epsilon)$ reflect those of the underlying fractal.

This example is another good illustration of why the computation of persistent Betti numbers is important. Topologically, the set is composed of disconnected line segments, so there can be no non-bounding cycles in the first Čech homology group. However, the geometry of the set creates holes in the ϵ-neighborhoods, as seen in Figure 3.22(b). In the previous two examples, it is clear that the holes do not persist as ϵ decreases because $\beta_1(\epsilon) = 0$ between the spikes. In Figure 3.23 we see an apparent growth in $\beta_1(\epsilon)$. The difference is that smaller holes appear before the larger ones disappear, resulting in an accumulation.

Notice that this set has the same values of γ_0 and γ_1 as the Cantor set relative. This implies that the Betti numbers are not enough to distinguish their different topological structure. Recall



(a) (b)

Figure 3.20: (a) 10^4 points on a simply connected set. (b) An ϵ-neighborhood and corresponding subset of the Delaunay triangulation.

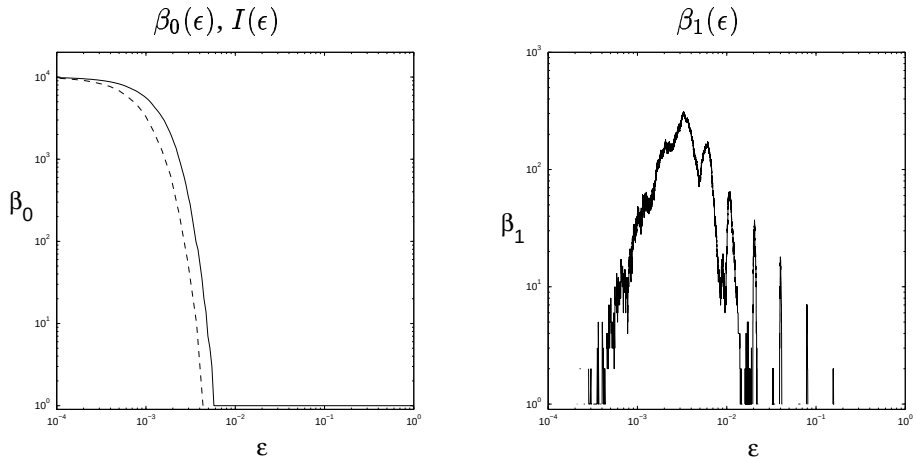
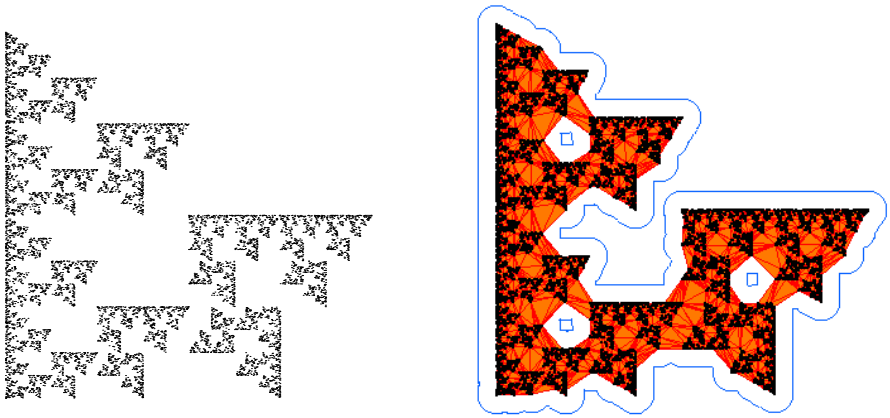


Figure 3.21: Number of components, $\beta_0(\epsilon)$, and number of holes, $\beta_1(\epsilon)$, for 10^4 points uniformly distributed over a simply connected fractal. The dashed line in the graph of $\beta_0(\epsilon)$ is the number of isolated points, $I(\epsilon)$. All axes are logarithmic. The horizontal axis range is $10^{-4} < \epsilon < 1$.



(a) (b)

Figure 3.22: (a) 10^4 points on a disconnected set. (b) An ϵ-neighborhood and corresponding subset of the Delaunay triangulation.

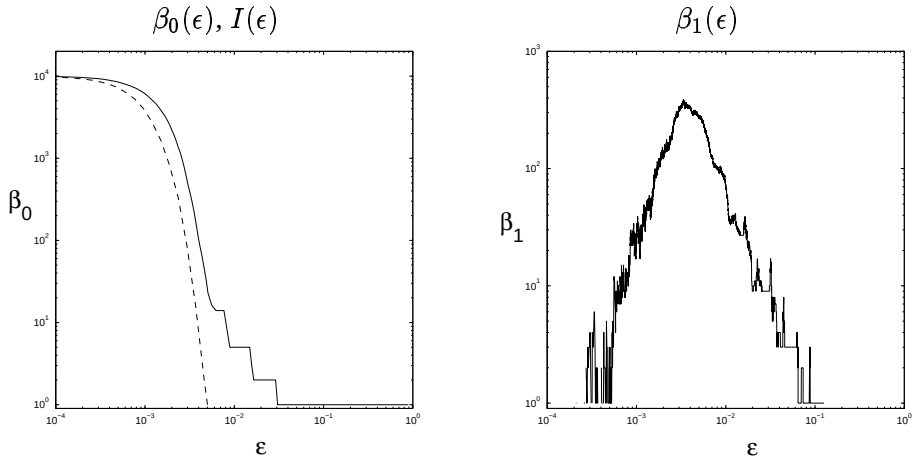


Figure 3.23: Number of components, $\beta_0(\epsilon)$, number of holes, $\beta_1(\epsilon)$, for 10^4 points uniformly distributed over the disconnected relative. The dashed line in the graph of $\beta_0(\epsilon)$ is the number of isolated points, $I(\epsilon)$. All axes are logarithmic. The horizontal axis range is $10^{-4} < \epsilon < 1$.

that this was addressed in Chapter 2 by examining the diameters of the connected components. For a Cantor set, the diameters go to zero, while for the example in Figure 3.22, the largest diameter converges to one.

3.6 Concluding remarks

We have demonstrated that it is possible to extract information about the topological structure of a compact space from a finite approximation to it. The examples given here are subsets of the plane where it is easy to see the structure of the data. In higher-dimensional spaces, it is extremely difficult visualize the underlying topology of a cloud of points, so computational techniques become even more important tools.

In Section 3.5, we computed only the regular Betti numbers, $\beta_k(\epsilon)$, of the ϵ-neighborhoods. The examples given highlight the need for computation of the ρ-persistent Betti numbers, $\beta_k^\rho(\epsilon)$. The latter give a better characterization of the topology of the underlying space. In applications, both sets of numbers may prove useful, since the Betti numbers of the ϵ-neighborhoods give geometric information about how the underlying space is embedded.

Many open problems remain. For the theory of persistent Betti numbers, we need further results relating the limit of $\beta_k^\lambda(X_\epsilon)$ as λ and ϵ tend to zero to the Betti numbers of X. In terms of computation, more work can be done on developing and implementing efficient algorithms for building complexes and computing persistent Betti numbers. Faster algorithms are essential for ease in studying the large data sets typically encountered in the study of dynamical systems. We anticipate that such computational tools will assist in the numerical investigation of chaotic systems by giving a deeper understanding of the structure of attractors and other invariant sets.

Chapter 4

Applications in Dynamical Systems

4.1 Introduction

In this chapter we examine data from numerical simulations of some discrete dynamical systems. The goal is to illustrate how our computational topology tools apply in this context. Previous applications of computational topology in dynamical systems focus on flows and chaotic time series. Muldoon *et al.* [60] compute homology groups for embedded time series data; Mischaikow *et al.* [37, 59] also use homology computation in a numerical implementation of Conley index theory to experimental data; a number of groups [28, 57, 82] have used knot-theoretic ideas to model the dynamics of attractors in $\mathbb{R}^3$. We focus on discrete dynamics rather than flows because there is less numerical error and greater efficiency in iterating a map than in solving a differential equation. Maps arise naturally from flows via Poincaré sections, for example, or through methods for their numerical solution, so their properties are closely related. While our techniques apply equally well to flows, every flow trajectory is connected, so only the higher order homology is interesting. There is a greater variety of topological structures in the phase space of a discrete map since an orbit can cover a disconnected set.

The three examples we study in detail are the Hénon attractor, the transition from invariant circle to Cantor set in the standard map, and cantori in a four-dimensional sawtooth map near the anti-integrable limit. These examples have well understood structure, so we are able to evaluate the effectiveness of our computational tools. We also show how our techniques could be applied to investigate the breakup of invariant tori in higher-dimensional symplectic maps, the structure of the chaotic region of the standard map, and for pruning outliers from embedded noisy time-series data. The examples are chosen to cover a wide range of dynamical and topological phenomena. They also require different approaches to approximating the underlying set and thereby convey the general applicability of the ideas from Chapters 2 and 3.

The Hénon map is the canonical two-dimensional quadratic map with constant Jacobian, in the sense that any other map in this class is conjugate to the Hénon map with some choice of parameters. In Section 4.2 we study the map at parameter values for which it has an attractor and use the minimal spanning tree techniques from Chapter 2 to study its connectedness properties. Our results give strong support to the common intuition that the Hénon attractor has Cantor-set cross-sections.

The standard map is an area-preserving twist map of the cylinder. It is the most commonly studied example of this class because it models a number of different physical problems, has an easily controlled perturbation from integrability, and exhibits most of the dynamical phenomena common to area-preserving twist maps. In Section 4.3 we investigate the topological change in

a quasiperiodic orbit as the perturbation parameter, k, is increased. Each quasiperiodic orbit has a critical k-value below which the orbit covers a circle, and above it, a Cantor set. The existence of invariant circles is important because they trap chaotic orbits and imply some degree of stability in the system. The transition from circle to Cantor set has been studied extensively, but this is the first time that the topology of these sets has been examined numerically. The minimal spanning tree tools of Chapter 2 successfully show the change in topology on a coarse k scale. A more precise criterion for determining the critical k-value is suggested by a closer analysis of how the longest MST edge scales with the number of points in the approximating orbit.

Symplectic maps are higher-dimensional versions of area-preserving maps and are discrete versions of Hamiltonian flows. The problem of torus breakup in nearly integrable symplectic maps is not well understood; this was one of the original motivating questions for the thesis. As a first step towards this goal, we examine a four-dimensional piecewise linear symplectic map near its anti-integrable limit in Section 4.4. It is known that at this extreme, orbits with incommensurate rotation frequencies cover Cantor sets called cantori since they are the remnants of invariant tori. We show that the cantori exhibit logarithmic rather than polynomial growth in the number of components, which is related to the fact that their Hausdorff dimension is zero. We discuss the potential of the computational topology tools from Chapter 3 to help understand torus breakup of nearly-integrable systems in Section 4.5.1.

The material in Sections 4.2 and 4.4 is published in [72].

4.2 The Hénon attractor

Figure 4.1 shows the much-studied Hénon attractor Λ, for the map H:

$$\begin{aligned} x_{n+1} &= y_n + 1 - a x_n^2 \\ y_{n+1} &= b x_n \end{aligned} \tag{4.1}$$

with parameter values $a = 1.4$ and $b = 0.3$. See Robinson [73] for a review of the properties of this map and its attractor. The first observation we make is that the set has a topological dimension of one and must be connected. This follows from the fact that the attractor is the closure of the unstable manifold of a fixed point for H. The attractor has fractal structure nonetheless, and is often described as having a Cantor set cross-section [73]. We use the minimal spanning tree techniques from Chapter 2 to investigate the above topological properties.

To generate finite-point approximations to the attractor, we compute the orbit, S, of a single initial point (x_0, y_0), from the trapping region for the Hénon attractor. For our connectedness data to be valid, we need some guarantee that this gives a good approximation, i.e., that the Hausdorff distance, $d_H(S, \Lambda) < \rho$. Firstly, this orbit must converge to the attractor, so given $\rho > 0$, there is an integer n, such that for $j > n$, $d(H^j(x_0, y_0), \Lambda) < \rho$. This means we have to iterate the map a few hundred times before we start recording the points visited by the orbit. The length of these transients depends on the strength of contraction along the stable manifolds of the attractor. If an attractor is topologically transitive, then a typical orbit will fill out the entire attractor. There is no proof that this is the case for the Hénon attractor with the parameter values $a = 1.4$ and $b = -0.3$, but numerical experiments suggest that there is a dense orbit; see [73] for more discussion.

The graphs in Figure 4.2 show the number of connected components, $C(\epsilon)$, number of isolated points, $I(\epsilon)$, and the largest component diameter, $D(\epsilon)$, for two orbits, one with 10^4 points and one with 5×10^4. These graphs are exactly what we expect to see for data that cover a connected set in a slightly nonuniform fashion. For the orbit with 10^4 iterates, isolated points

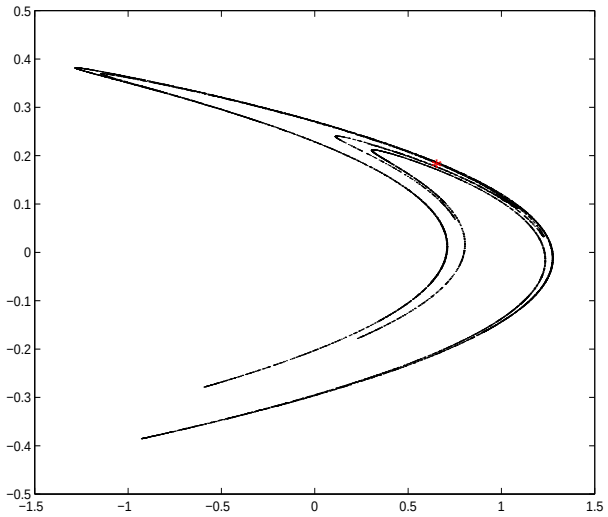


Figure 4.1: An orbit on the Hénon attractor. The red star is the unstable fixed point of H with coordinates $(x, y) \approx (0.631, 0.189)$.

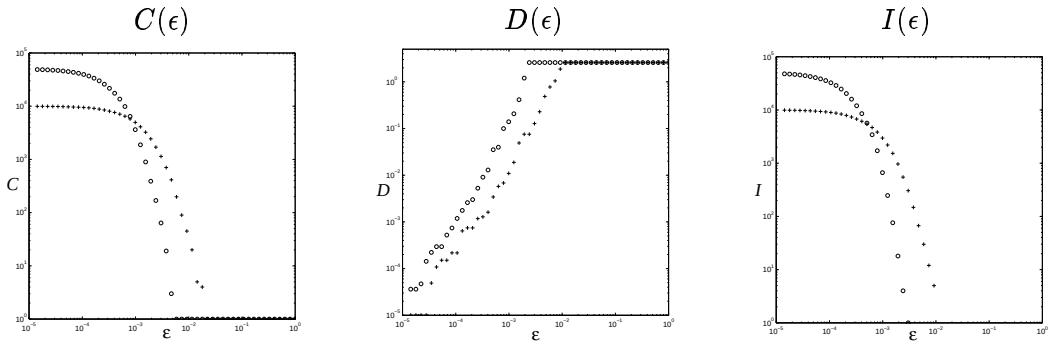
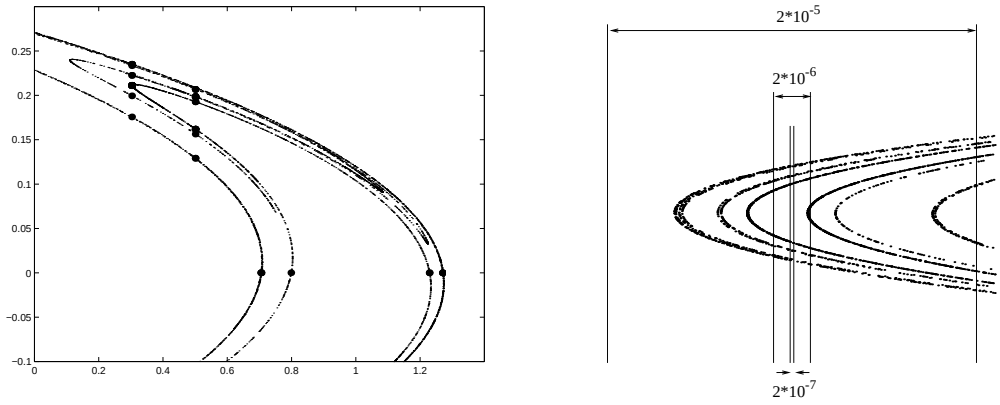


Figure 4.2: $C(\epsilon), D(\epsilon)$, and $I(\epsilon)$ data for two orbits on the Hénon attractor. The crosses, $+$, represent calculations for the orbit of 10^4 iterates and the circles, $\circ$, are for an orbit with 5×10^4 points. All axes are logarithmic. The horizontal axis range is $10^{-5} < \epsilon < 1$.



(a) (b)

Figure 4.3: (a) A close-up of the Hénon attractor. The bold face dots are points in the three cross-sections considered in the text: slices at $x = 0.302435$, $x = 0.5$ and $y = 0$. (b) A small part of the slice at $x = 0.302435$, $y = 0.22$ that shows the folding of the attractor. The pairs of vertical lines in (b) are the boundaries of the different sub-slices of widths 2×10^{-5}, 2×10^{-6} and 2×10^{-7}.

are resolved at $\rho \approx 10^{-2}$. The number of components at this resolution is actually around ten, so this value of ρ underestimates the true cutoff resolution. The graph of $D(\epsilon)$ shows that the largest component diameter does not change until $\epsilon < \rho$, and then it decreases sharply to zero, which is what we see for other connected sets. The orbit with 10^5 points has a smaller cutoff resolution, $\rho \approx 3 \times 10^{-3}$, and qualitatively similar graphs, which increases our confidence in the numerical results.

To investigate the Cantor set cross-sections, we compute $C(\epsilon)$, $D(\epsilon)$, and $I(\epsilon)$ for thin slices taken through the attractor at three different places: $x = 0.302435$, $x = 0.5$ and $y = 0$ (the bold dots in Figure 4.3(a)). This process is related to taking the Poincaré section — a common technique for visualizing the structure of attractors from flows. To create a Poincaré section, a surface of codimension one is chosen and points on the section are recorded whenever the trajectory pierces the surface. The Hénon attractor is generated by iterating a map, so it is not possible to find many points on a given section. Instead, the sections are generated by recording points that fall within an interval of the given section coordinate. This means that the slices have a finite width and the data are still two-dimensional. To be confident that the observed scaling behavior is approximating that of a one-dimensional Cantor set, we compute $C(\epsilon)$, $D(\epsilon)$, and $I(\epsilon)$ for four successively narrower slices at each cross-section. The thinnest slice in each case has a width of 2×10^{-7}.

The section at $x = 0.302435$ is interesting because it cuts through a fold in the attractor. This folding of the Hénon attractor is the source of its nonuniform hyperbolic structure. Note that a one-dimensional cross-section that touches a fold will have an isolated point and therefore cannot be a Cantor set. A close-up of such a fold is shown in Figure 4.3(b). This figure shows that slices of different widths taken at this x value capture different folding structure. This is reflected in the $C(\epsilon)$ data in figure 4.4. The data for different slices at $x = 0.302435$ does not coincide exactly for $\epsilon > \rho$, as it does for the other two sections at $x = 0.5$ and $y = 0$, which show no folding at these resolutions. Note that we can reverse this observation and use the inconsistency to detect cross-sections that touch a fold. The sections at $x = 0.5$ and $y = 0$ have simpler structure. The graphs of $D(\epsilon)$ show the now-familiar staircase structure of a Cantor set.

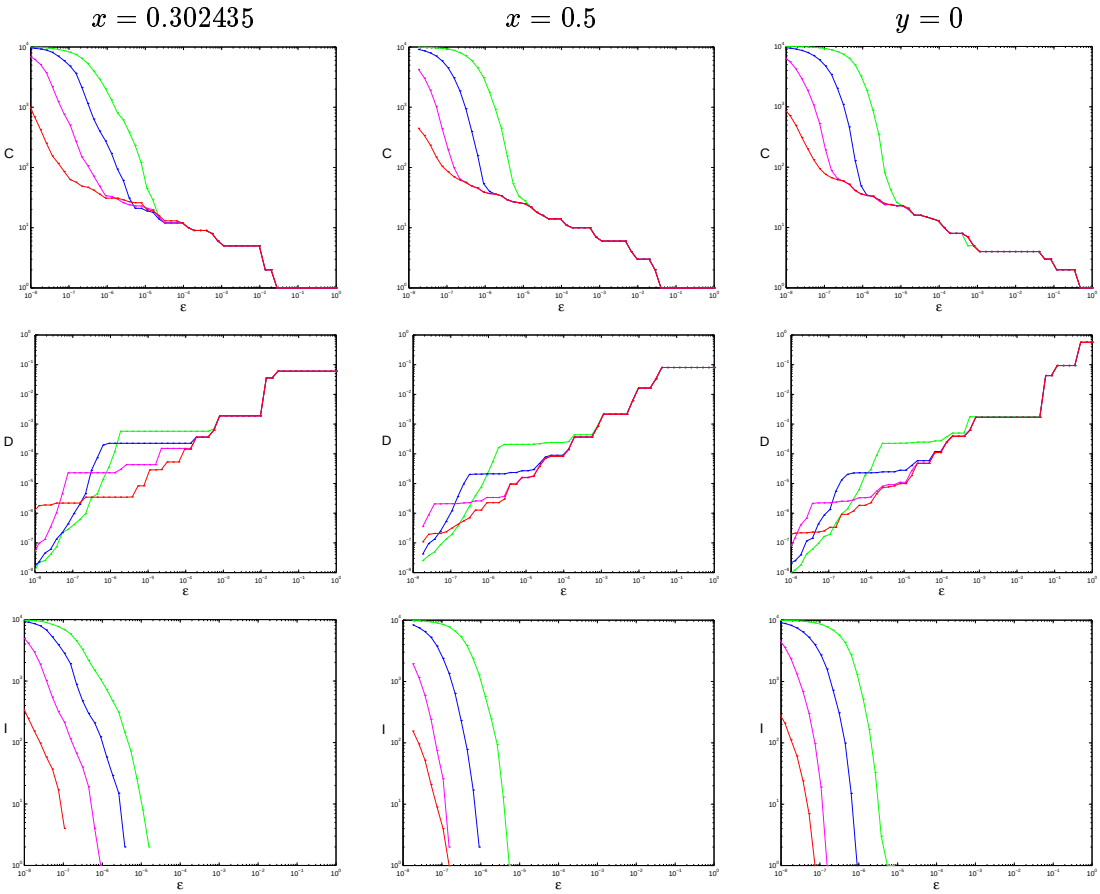


Figure 4.4: $C(\epsilon)$, $D(\epsilon)$, and $I(\epsilon)$ data for three sections of the Hénon attractor. The left column shows data for $x = 0.32435$, the middle column for $x = 0.5$, and the right for $y = 0$. The curves are colored according to the width of the slice: 2×10^{-4} is green, 2×10^{-5} is blue, 2×10^{-6} is magenta, and 2×10^{-7} is red. All axes are logarithmic. The horizontal axis range is $10^{-8} < \epsilon < 1$.

Table 4.1: Values of γ and δ for the three sections of the Hénon attractor shown in Figure 4.3.

section	γ	δ
$x = 0.302435$	0.25 ± 0.01	0.9 ± 0.1
$x = 0.5$	0.26 ± 0.01	0.85 ± 0.04
$y = 0$	0.27 ± 0.01	0.88 ± 0.02

The flat segments in each graph of $D(\epsilon)$ are due to the finite width of each slice, making the data a Cantor set of line segments.

Values of γ and δ are calculated from the $C(\epsilon)$ and $D(\epsilon)$ data for the thinnest slice at each section. The results are summarized in Table 4.1. The multifractal nature of the Hénon attractor [86] means that we expect the dimension to vary for different cross-sections. For the three examples given here, though, the variation is not significant. It is possible that the multifractal nature would become apparent if we computed γ from thinner slices with more data. However, the multifractal spectrum is determined by choosing points on the attractor and computing local scaling rates. The sections are not local and this may obscure the multifractal nature of the attractor.

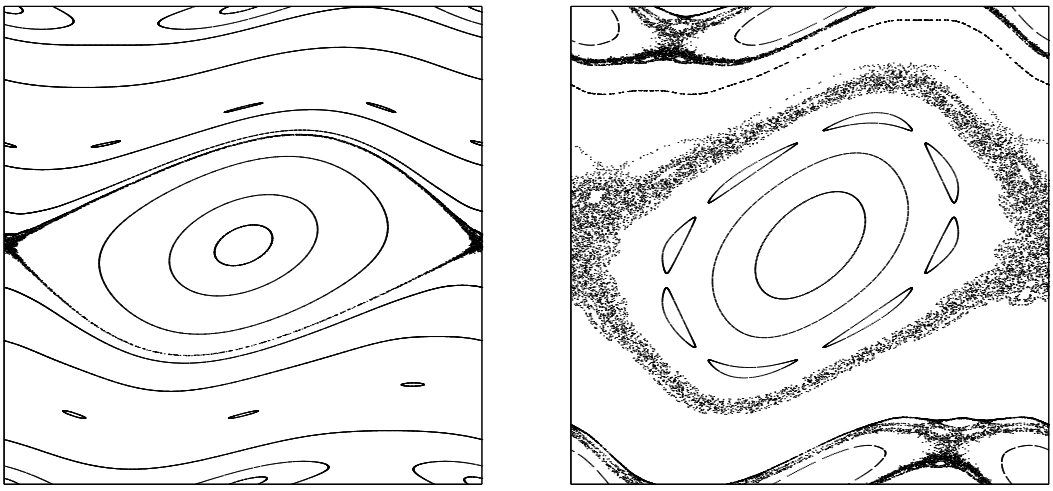
The above results give strong numerical support for the common belief that cross-sections of the Hénon attractor are Cantor sets. The box-counting dimension of the Hénon attractor is estimated to be about 1.27 [86, 63]. Results on the dimension of intersections of sets [23] imply that the dimension of a cross section through the Hénon attractor should be $1.27 - 1 = 0.27$. The values of γ given in Table 4.1 are in close agreement with this value, providing further support for our conjecture that Cantor sets of zero measure have γ equal to the box-counting dimension.

4.3 Circle breakup in the standard map

In this section we examine orbits from the standard map as a parameter controlling the non-linear perturbation is increased. This map is a popular example of the class of area-preserving twist maps of the cylinder, which are closely related to Hamiltonian flows with two degrees of freedom. The properties of these maps are reviewed in [55].

Our goal is to see how the computational topology tools from Chapter 2 perform in detecting a transition from circle to Cantor set. This transition is interesting dynamically because the invariant circles trap chaotic orbits and therefore imply some degree of stability. When an invariant circle is destroyed, it is possible for nearby chaotic orbits to diffuse through the gaps in the remaining Cantor set (see Figure 4.17). If no circles exist, then it is possible for a single chaotic orbit to access most of the phase space. Many approaches to detecting the existence of invariant circles have been explored in the last twenty or so years, [8, 31, 44, 49, 53]. These techniques exploit properties such as the stability, flux, and frequency of orbits; our work is the first time the topology has been used directly.

For our computational tools to give valid results, we need good finite point approximations to the underlying circle or Cantor set. The quasiperiodic orbits which cover these sets are very difficult to find numerically because their position in phase space is uncertain and they do not attract other orbits. Instead, we use long periodic orbits that are close to quasiperiodic ones, since symmetry properties of the standard map make these relatively easy to find. This method



$k = 0.5$ $k = 0.976135406$

Figure 4.5: Phase space portraits for the standard map at two different k values.

of approximation was pioneered by Greene [31], and is justified by Aubry-Mather theory [46].

We begin this section by describing some of the basic phenomenology of the standard map, and reviewing the essential details of Aubry-Mather theory and the method of approximating quasiperiodic orbits by symmetric periodic orbits. For the purposes of comparison, we also summarize some previous numerical work on detecting the breakup of invariant circles. We then present connectedness data from the minimal spanning tree techniques of Chapter 2 applied to periodic orbit approximations of an invariant circle. The results obtained are not as sensitive at detecting the transition as are some previous techniques, so we investigate scaling properties of the $C(\epsilon)$ distributions in more detail. This work raises many questions about the connections between the distribution of points and the distribution of edge-lengths in their minimal spanning tree, and about the renormalization of phase space near a critical circle. This is part of a larger project that is outside the scope of this thesis.

4.3.1 Background

The standard map

$$\begin{aligned} y_{n+1} &= y_n - \frac{k}{2\pi}\sin(2\pi x_n) \\ x_{n+1} &= x_n + y_{n+1} \mod 1. \end{aligned} \tag{4.2}$$

appears in a number of physical applications. It models charged particle motion in a simple particle accelerator, the cyclotron, and is equivalent to the Frenkel-Kontorova model of a chain of atoms in a one dimensional periodic potential. The standard map is also the result of applying a first order numerical solver to the pendulum equation $\ddot{\theta} - K\sin\theta = 0$, with $x = \theta/(2\pi)$, $y = \dot{x}$, and $t = n$. The x variable is an angle in $[0, 1)$ and y is analogous to momentum. Clearly y is also periodic, so the phase space is in fact homeomorphic to the torus and we need only consider orbits in the unit square.

We label the orbits by their rotational frequency, ω, the average change in x per iteration of

the map:

$$\omega = \lim_{n\to\infty} \frac{x_n - x_0}{n}. \tag{4.3}$$

To calculate ω, we iterate $x_{n+1} = x_n + y_{n+1}$, without taking the fractional part (this corresponds to taking a lift of the map of the cylinder (4.2) and making it a map of the covering space, $\mathbb{R}^2$). The twist condition,

$$\frac{\partial x_{n+1}}{\partial y_n} \geq C > 0, \tag{4.4}$$

means that orbits with larger y have larger rotational frequency. This is compatible with the physical interpretation of y as a momentum variable. Periodic orbits satisfying $y_q = y_0$, $x_q = x_0 + p$, have rational rotation frequency p/q and are called (p, q)-periodic orbits. Quasiperiodic orbits have irrational rotation frequency. Note that the limit in (4.3) may not exist, for example, in the case the orbit is chaotic.

When the nonlinearity parameter k is zero, the map is integrable. Its orbits lie on invariant circles $y = y_0$ and evolve as rigid rotations of the angle variable, $x_{n+1} = x_n + y_0$, so that $\omega = y_0$. When $k > 0$, the geometric structure of the orbits becomes dramatically more complicated; this is illustrated in Figure 4.5. The Poincaré-Birkhoff theorem shows that every invariant circle with rational rotation frequency is destroyed, leaving a pair of periodic orbits with that frequency, one elliptic and one hyperbolic. Circles with irrational rotation frequency are guaranteed to persist for a small interval of k values by the Kolmogorov-Arnold-Moser (KAM) theorem. See Arrowsmith and Place [3] for more details about these results.

Periodic and quasiperiodic orbits

The existence of a special class of periodic and quasiperiodic orbits in any area-preserving twist map follows from Aubry-Mather theory; see [55] for an overview. These orbits minimize a quantity called the action and are monotone (the order of points is preserved under iteration). An extremely useful result from Aubry-Mather theory is that if a sequence of rationals converges to an irrational, $p_n/q_n \to \omega$, then the minimizing (p_n, q_n)-periodic orbits converge (in the Hausdorff metric) to a set of minimizing orbits with rotation number ω. Aubry-Mather theory also shows that this limit set is either an invariant circle or a Cantor set. Since it is relatively easy to find minimizing periodic orbits in the standard map, this result means we can use sequences of these orbits to approximate a quasiperiodic one. This approach, first used by Greene [31], will form the basis of approximation in our numerical investigation of invariant circles and Cantor sets in Sections 4.3.2 and 4.3.3.

It is the symmetry properties of the standard map that make minimizing (p, q)-period orbits easy to find (see [47] for more details). Each such orbit must have a point on two of the following four symmetry lines:

$$a = \{z = (x, y) \mid x = 0\} \tag{4.5}$$
$$b = \{z = (x, y) \mid x = 1/2\} \tag{4.6}$$
$$c = \{z = (x, y) \mid x = y/2\} \tag{4.7}$$
$$d = \{z = (x, y) \mid x = (y + 1)/2\}. \tag{4.8}$$

The pair of lines is determined by p and q. For action-minimizing orbits in the standard map the observed pattern is:

$$p \quad \text{odd}, \quad q \quad \text{even} \quad \Rightarrow \quad z_0 \in c \quad \text{and} \quad z_{q/2} \in d \tag{4.9}$$

$$p \quad \text{odd}, \quad q \quad \text{odd} \quad \Rightarrow \quad z_0 \in b \quad \text{and} \quad z_{(q+1)/2} \in c \tag{4.10}$$

$$p \quad \text{even}, \quad q \quad \text{odd} \quad \Rightarrow \quad z_0 \in b \quad \text{and} \quad z_{(q+1)/2} \in d. \tag{4.11}$$

These properties mean the search for an orbit is made along a single line, rather than over the whole of phase space. We use these types of orbits in Section 4.3.2.

In practice, we want the sequence of rationals p_n/q_n to converge to the irrational ω as fast as possible. This is achieved by using the continued fraction convergents of ω (see [55] for a description of how to generate continued fraction expansions). The continued fraction convergents are the "best approximants" in the sense that if p/q is a convergent of ω, then any other rational p'/q', with $q' \leq q$ is further from ω, i.e., $|\omega - p'/q'| > |\omega - p/q|$. The number with the slowest convergence of its best approximants is the golden mean,

$$\mu = \frac{1+\sqrt{5}}{2} = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}} = 1.61803\ldots . \tag{4.12}$$

The convergents, p_n/q_n, of μ are related to the Fibonacci sequence: $F_{n+1} = F_n + F_{n-1}$, with $F_0 = F_1 = 1$, by setting $p_n = F_{n+1}$, and $q_n = F_n$.

The invariant circle with golden mean rotation frequency plays an important role in the standard map, since it is the last one to be destroyed, i.e., its transition from circle to Cantor set occurs at a k-value larger than that of any other quasiperiodic orbit. This stems from the status of the golden mean as the number that is hardest to approximate by rationals, and therefore the orbit for which the KAM small divisor problems are minimal.

Renormalization near a critical circle

The phase space near an invariant circle at its critical k-value (i.e. the point of breakup) has a remarkable degree of local self-similarity [31, 75]. MacKay [47] studies this extensively and quantifies the scaling relations using renormalization techniques. We observe related scaling properties in the MST data, as discussed in Section 4.3.3.

As an illustrative example, we describe results for the standard map and an orbit with rotation frequency of $1/\mu^2 = 0.38196\ldots$. This number is closely related to the golden mean and it has continued fractions convergents, p_n/q_n with $p_n = F_n, q_n = F_{n+2}$; it also has the same critical value of $k = k_c = 0.971635406$.

Near the point where the golden circle crosses $x = 0$, the phase space is asymptotically self-similar with a scale factor of $\alpha_0 = -1.4148360$ in x and $\beta_0 = -3.0668882$ in y. The negative signs indicate a flip in orientation which is caused by the oscillation of successive convergents: $p_n/q_n < \omega < p_{n+1}/q_{n+1}$.

At $x = 1/2$, the scaling is a little more complicated. Here, the structure of the period orbits cycles with period 3. Shenker and Kadanoff [75] explain this in terms of the symmetry properties of the orbits. The Fibonacci numbers,

$$F_i = 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377\ldots$$

follow a repeating pattern of odd, odd, even. This implies that the minimizing (p_n, q_n)-periodic orbits that converge to the $1/\mu^2$ circle cycle through the three symmetry patterns of (4.9) . The renormalization must take this into account and MacKay [47] computes 3-step ratios of $\alpha_{.5} = -4.84581$ in x and $\beta_{.5} = -16.8597$. Since area is preserved, these values are related to α_0 and β_0 via:

$$4.3391 = \alpha_0\beta_0 = \sqrt[3]{(\alpha_{.5}\beta_{.5})} = (1.6922)(2.5642) = 4.3391. \tag{4.13}$$

Although we described ths scaling in terms of a particular critical circle, the scaling holds for any circle with a rotation frequency that has a continued fraction expansion that ends in ones [47].

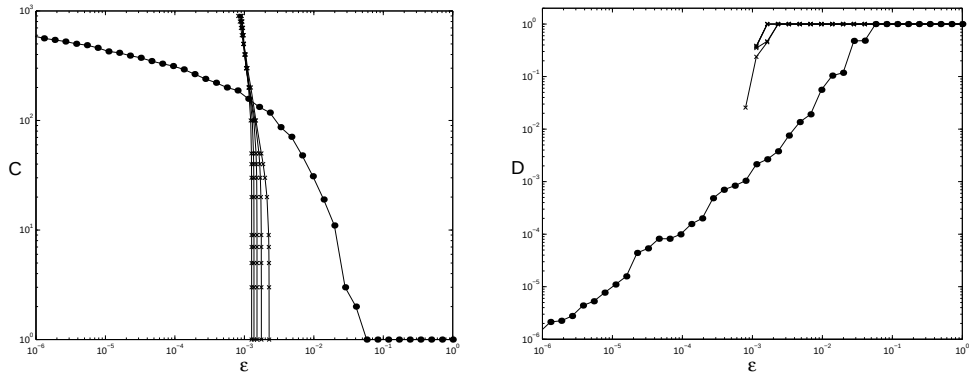
Nonexistence of invariant circles

Invariant circles that go around the cylinder (i.e. ones that are homotopic to $y = 0$) are called rotational invariant circles. There existence is important because they confine chaotic orbits to a small region of phase space and imply that the system is essentially stable. If no rotational invariant circles exist for any rotation frequency, it is possible for a single chaotic orbit to cover most of the phase space. This means that the dynamics is inherently unstable; in particular, the momentum variable, y, can increase without bound. Invariant circles are observed to exist for much larger values of k than those given by KAM theory. Many methods exist for estimating the critical value, k_c, such that for $k > k_c$, no KAM circles remain; we outline three computational approaches below. The most accurate numerical determination of k_c is due to MacKay [47, p.199], who uses Greene's criterion and renormalization techniques to estimate $k_c = 0.971635406$. The best rigorous bound on k_c for the standard map is due to MacKay and Percival [50] who show that there are no invariant circles for $k > 63/64 = 0.984375$.

A very direct method due to Chirikov [8] estimates k_c by measuring diffusion times of chaotic orbits. For different values of k, and initial conditions, (y_0, x_0) with $y_0 \approx 10^{-3}$, he records the number of iterations, N, it takes for the orbit to reach $y \approx 0.5$. There is a sharp increase in N for values of k near k_c; the longest orbits considered have $N = 10^7$. By fitting the data obtained for $N(k)$, he estimates $k_c \approx 0.989$. This is surprisingly close to the above values, given the inaccuracies of these numerical techniques.

More recently, Laskar [44] has developed a general method for detecting the destruction of invariant circles by computing the frequency of orbits with different initial conditions. Recall that the twist condition implies that along a vertical line, $x = 0$ say, orbits with larger initial y coordinate have larger frequency. Laskar therefore uses initial conditions along such a vertical line and looks for non-monotonicity in the computed frequencies. For the standard map he find k_c to five significant figures using orbits of around 10^6 points. This makes it computationally intensive, but the approach generalizes readily to higher-dimensional maps.

The most sensitive numerical technique is probably Greene's mean residue criterion [31]. This approach uses the stability of long periodic orbits that are close to a quasiperiodic one to indicate the existence of an invariant circle. The residue of a (p, q)-periodic orbit is a measure of stability related to the eigenvalues: $R = \frac{1}{4}(2 - \text{Trace } M)$, where M is the matrix linearization of the q-fold composition of the map. The mean residue is a type of geometric mean, $f = (|R|/\beta)^{1/q}$, where β is a scale factor used to improve convergence. To extend the definition of f to quasiperiodic orbits, Greene uses the continued fraction convergents p_n/q_n of an irrational ω. The value of $f(\omega)$ is found by taking the limit: $f(\omega) = \lim_{n\to\infty} f(p_n/q_n)$. For the standard map, $\beta = 1/4$, and Greene's criterion is that the orbit with rotation number ω covers a circle if and only if $f(\omega) < 1$. An attempt to generalize this criterion to higher-dimensional symplectic



(a) (b)

Figure 4.6: Connected component data for a $(377, 987)$-periodic orbit approximation to the golden $1/\mu^2$ orbit at different values of k. In both graphs, the data marked with solid circles, $\cdot$, is for $k = 1.0$. The crosses, $\times$, mark data for the other k values, $k = 0.5, 0.6, 0.7, 0.8, 0.9$. The cut-off resolution for the $k = 1.0$ data is $\rho \approx 6 \times 10^{-5}$. (a) Number of components as a function of resolution, ϵ. (b) Diameter of the largest component as a function of ϵ.

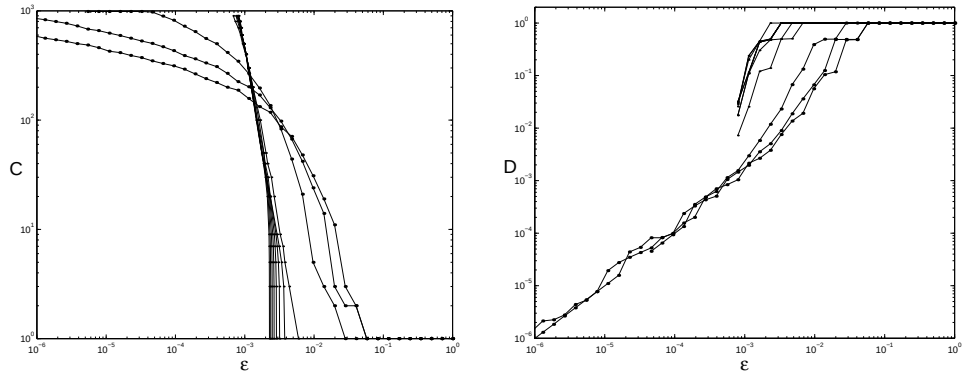
twist maps is made in [70]. A formal justification for this approach is hampered by the fact that there is no corresponding Aubry-Mather theory in higher dimensions. Greene's criterion implies that the golden invariant circle is destroyed at $k = k_c = 0.971635$. This value of k_c is determined in [31] using orbits of length around 10^4.

4.3.2 MST analysis of the transition from circle to Cantor set

In this section we investigate the transition from invariant circle to Cantor set using the computational tools based on the minimal spanning tree from Chapter 2. As one might suspect, these techniques cannot make as precise a determination of k_c as Greene's residue criterion does, for example. This is because there is a continuous metric change from invariant circle to Cantor set; the results of Chapter 3 imply that the ϵ-neighborhoods have correspondingly close topology. However, our connectedness techniques do show the transition from circle to Cantor set on a coarse k-scale.

We start by examining the structure of a single orbit in the sequence of periodic approximations to the $1/\mu^2$ orbit as k is changed. The data in Figure 4.6 is for the minimizing (377,987)-periodic orbit approximation for $k = 0.5, 0.6, \ldots, 0.9, 1.0$. There is a significant difference between the $C(\epsilon)$ and $D(\epsilon)$ graphs for $k \leq 0.9$ and $k = 1.0$, consistent with the fact that $0.9 < k_c < 1.0$. For $k \leq 0.9$, the $C(\epsilon)$ and $D(\epsilon)$ data look like that of a connected set. For $k = 1.0$, the cutoff resolution drops to 6×10^{-5} and $D(\epsilon) \to 0$, which suggests a totally disconnected set. We estimate δ as the slope of the least squares linear fit for $10^{-4} \leq \epsilon \leq 10^{-2}$, and find $\delta \approx 1.15 \pm 0.05$. The growth in the number of components, $C(\epsilon)$, resembles that of a Cantor set with $\gamma = 0$ (recall Section 2.2.3).

Similar results are obtained with a finer k scale, as shown in Figure 4.7. Here, we present connected component data for the period-987 orbit with $k = 0.90, 0.91, \ldots, 1.0$. Again, we see a significant change in the $C(\epsilon)$ and $D(\epsilon)$ curves as k increases. This time, for $k \leq 0.97$, the orbit appears to approximate a connected set, and for $k \geq 0.98$, a Cantor set. The above results give an estimate of k_c to two significant figures using an orbit of just 987 points. To obtain higher precision we need to use longer orbits; the different topological structure of the golden



(a) (b)

Figure 4.7: Connected component data for a $(377, 987)$-periodic orbit approximation to the golden orbit for $k = 0.90, 0.91, \ldots, 0.99, 1.0$. (a) Number of components as a function of resolution, ϵ. (b) Diameter of the largest component as a function of ϵ. The change in the shape of the curves suggests that the transition from circle to cantorus occurs between $k = 0.97$ and 0.98.

orbit at $k = 0.971$ and $k = 0.972$, for example, is revealed for approximations of around 10^4 points.

The graphs in Figure 4.8 show the ambiguity that occurs when the approximation is not good enough. In each column, we fix k and plot connectedness data for a sequence of successively longer periodic orbit approximations ($q_n = 377, \ldots, 4181$) to the golden orbit. The data for $k = 0.97$ show that the orbits are probably converging to a connected set, as expected. For $k = 0.98$, it is clear from the persistence of the long MST edges, and the trend $D(\epsilon) \to 0$, that the orbits are converging to a Cantor set. At $k = 0.972$, however, the data is misleading since the orbits appear to be converging to a connected circle even though the $1/\mu^2$ orbit covers a Cantor set. The problem is that the largest gap in the Cantor set is shorter than the cutoff resolution for the data; see Figure 4.9.

The graphs of $C(\epsilon)$ for $k = 0.97$ and $k = 0.972$ in Figure 4.8 appear to keep the same form as q_n increases. The graph of $C(\epsilon)$ is really a cumulative distribution of edge lengths from the MST. It may be possible, therefore, to obtain more accurate information about the underlying topology by analyzing the relative distribution of edge lengths. We examine how distributions of MST edges change with k and the period of the orbit in the following section.

4.3.3 Scaling of the gaps

To explore the gap distributions, we again use the minimizing (p_n, q_n)-periodic orbit approximations to the $1/\mu^2$ orbit. We can simplify the numerical computation of the MST by using a result from Aubry-Mather theory which shows that a monotone invariant set (i.e., minimizing periodic and quasiperiodic orbits) must be a graph over x [55]. This lets us order the q points in an orbit $z_i = (x_i, y_i)$ by their x-coordinate, i.e., $x_i < x_j$ if $i < j$. The lengths of the MST edges are then given by the "gaps," $g_i = d(z_{i+1}, z_i)$ for $i = 0, \ldots, q-1$, and $g_q = d(z_0, \tilde{z}_q)$, where $\tilde{z}_q = (x_q - 1, y_q)$. In fact, the q gaps, g_i, give us the distance between every neighboring pair of points, whereas the MST, by its definition, omits the largest gap. We start by examining how the maximum, minimum, and mean of the gap lengths scale with the length of the periodic orbit approximations at different values of k. When $k = k_c$, our results are closely related to the local

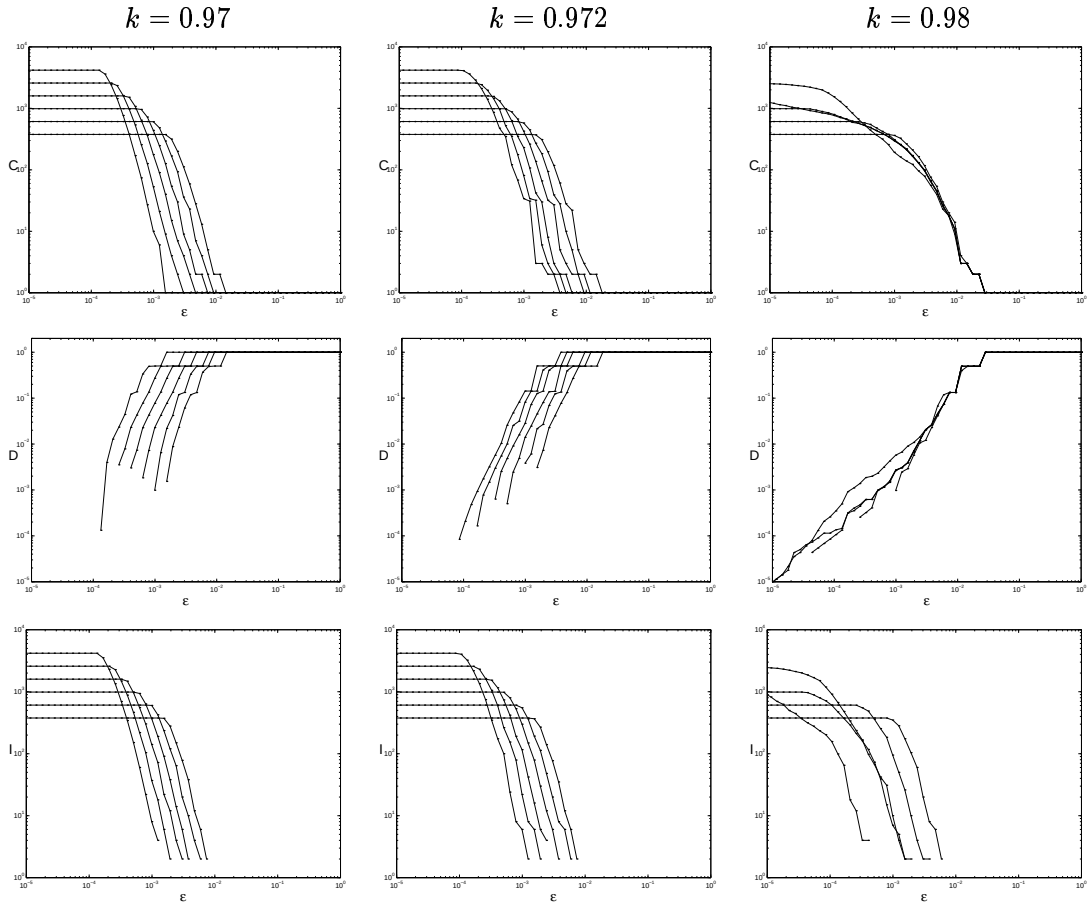
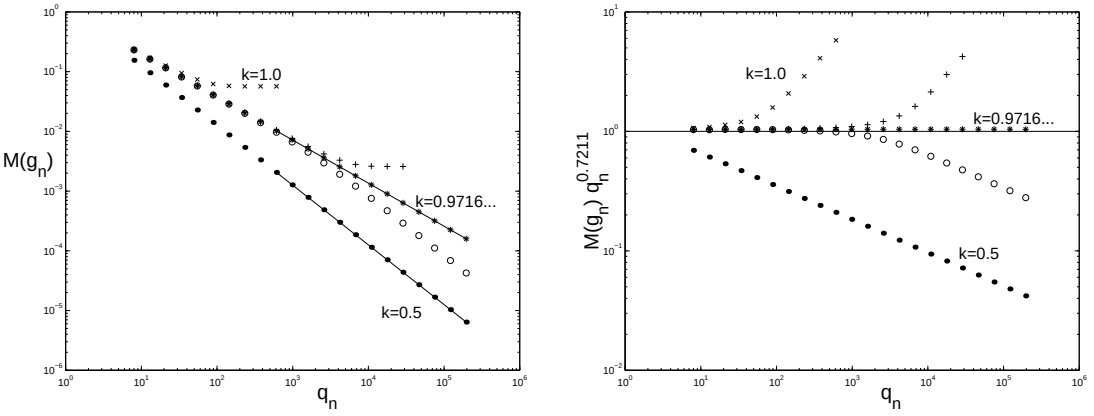


Figure 4.8: $C(\epsilon)$, $D(\epsilon)$, and $I(\epsilon)$ data for three values of k in the standard map. The left column has $k = 0.97$, the middle column, $k = 0.972$, and the right one, $k = 0.98$. In each case the data is obtained for a sequence of (p_n, q_n)-periodic orbit approximations to the $1/\mu^2$ orbit, with $q_n = 377, 610, 987, 1597, 2584, 4181$. For $k = 0.97$, the data suggests that the orbits are approximating a circle, and for $k = 0.98$, a Cantor set. When $k = 0.972$ is just larger than the critical value $k_c = 0.971635$, the golden orbit is a Cantor set, but the data given suggests the periodic orbits are converging to a circle. This is because the orbits are not long enough to allow us to resolve the gaps.



(a) (b)

Figure 4.9: (a) The largest gap, $M(g_n)$, as a function of the length of the periodic orbit approximation, q_n, to the golden orbit. The solid circles $\cdot$ are data for $k = 0.5$, the circles $\circ$ for $k = 0.971$, the stars $*$ for $k = k_c = 0.971635406$, the pluses $+$ for $k = 0.972$, and the crosses $\times$ for $k = 1.0$. The lines through the data for $k = 0.5$ and $k = k_c$ are least square fits for $610 \leq q \leq 196418$; the slopes are -1.0000 and -0.7211 respectively. (b) The same data plotted as $M(g_n)q_n^{0.7211}$ versus q_n, to exaggerate the difference between orbits converging to a circle and those converging to a Cantor set.

renormalization properties of the standard map [47]. In the final paragraphs of this section, we make a preliminary investigation of the relationship between the distribution of points and the distribution of gaps.

The largest gap

Figure 4.9(a) summarizes a study of the relationship between the largest gap, $M(g_n)$, and the period, q_n, of orbits converging to the $1/\mu^2$ orbit. There are three scaling regimes in these graphs, defined by $k < k_c$, $k = k_c$, and $k > k_c$. For $k < k_c$, $M(g_n)$ tends to zero because the orbits are converging to an invariant circle. The same is true at the critical value k_c, but the rate of convergence is different because the density of points goes to zero at every point in the orbit of $x = 0$. For $k > k_c$, $M(g_n)$ decreases and then flattens at a value that corresponds to the length of the largest gap in the Cantor set. For $k = 0.972$, this occurs when $q_n > 10^4$, and for $k = 1.0$, when $q_n > 10^2$.

We now describe the linear scaling for $k \leq k_c$ in more detail. The data for $k = 0.5$ show a clear linear relationship between $\log M(g_n)$ and $\log q_n$, which implies $M(g_n) = C/q_n^{\nu}$. From the least squares linear fit of the data for $q_n > 100$ we find that $\nu = 1.0000$ and $C = 1.256$. The slope, ν, equals one because the points are fairly evenly distributed around this circle (see Figure 4.12). There is also simple linear scaling at the critical value, $k_c = 0.971635406$. In this case the least squares fit for $q_n > 100$ gave $\nu = 0.7211$ and $C = 1.042$. The rate at which $M(g_n) \to 0$ is closely related to the renormalization scheme for a critical circle at $x = 0$ because the largest gap always occurs around $x = 0$. The scaling described in Section 4.3.1 implies that $M(g_n) \sim \alpha_0^{-n}$. Our results relate $M(g_n)$ and the period q_n as $M(g_n) \sim q_n^{-\nu}$. Since the periods are Fibonacci numbers, their growth is governed by the golden mean, $q_n \sim \mu^n$. Therefore, we must have $\alpha_0 = \mu^{\nu}$, or $\nu = \log(\alpha_0)/\log(\mu) = 0.72107$ (using the value of α_0 from [47]), which is exactly what we computed from the graph in Figure 4.9(a). When k is

just less than the critical value, e.g., $k = 0.971$, there appear to be two scaling regions. This is the result of the nonuniform distribution of points in the periodic orbit approximations to the invariant circle. For $q_n < 1000$, the orbits are close to those approximating the critical circle, so $M(g_n)$ appears to scale as $C/q_n^{0.73}$. However, for $q_n > 10^4$, the slope is very close to one; we compute $\nu = 0.998$ and $C = 8.2$.

The graph in Figure 4.9(a) suggests a new, purely geometric, criterion for determining the transition from circle to Cantor set in area-preserving maps. Given (p_n, q_n)-periodic orbits with p_n/q_n converging to an irrational, the quantity to examine is $M(g_n)q_n^\nu$ as $q_n \to \infty$. The parameter ν should be chosen to make the convergence as follows: for $k < k_c$, $M(g_n)q_n^\nu \to 0$, and for $k > k_c$, $M(g_n)q_n^\nu \to \infty$. The motivation for this is analogous to the choice of the parameter β in Greene's criterion. For the standard map, we could choose $0.7211 \le \nu < 1$. Setting $\nu = 0.7211$ gives the best results because this balances the critical scaling exactly. This data is shown in Figure 4.9(b). Based on the data given in that graph, we estimate that the value of k_c could be determined to about five or six significant figures using this criterion.

Much more numerical work is needed to test the validity of this idea. We need to look at circles with other irrational frequencies and different area-preserving maps. The advantage of this criterion is that it does not require any information about the stability of an orbit — an expensive computation involving eigenvalues of large matrices. The new criterion is unlikely to generalize to maps in higher dimensions, however, since the topological transition is very different.

The mean and minimum gap

We can summarize some information about the distribution of gap lengths by plotting the maximum, $M(g_n)$, mean, $\langle g_n \rangle$, and minimum, $m(g_n)$, values as functions of the period, q_n on the same axes. This data is given in Figure 4.10 for the same values of k as in Figure 4.9.

In all cases, we see linear scaling in the mean gap size. This follows from the fact that

$$\langle g_n \rangle = \frac{\sum_{j=1}^{q_n} g_j}{q_n} \approx \frac{L}{q_n}. \tag{4.14}$$

The quasiperiodic orbit covers an invariant circle or a Cantor set. In either case, the orbit lies on a Lipschitz graph, so its length L, is finite. We are computing the distance between points that are close to the Lipschitz graph, so the sum of the gaps converges L. This convergence is quick, so the approximation in (4.14) is a reasonable one. If we instead compute g_j as the difference in x-coordinate, then we would have $\sum_{j=1}^{q_n} g_j = 1$ for all n and k.

We now look at the scaling of the shortest gaps. For $k = 0.5$, the relationship between $\log m(g_n)$ and $\log q_n$ is clearly linear, implying $m(g_n) = c/q_n^\eta$. From the data for $q_n > 100$, we compute $c = 0.893$ and $\eta = 1.0000$. Again, the slope is one because the points are fairly evenly distributed around the circle. For $k = k_c$ there is also linear scaling. Initially, we estimate $\eta = 1.0928$ from the data for $q_n > 100$. Since the density of points is highest at $x = 0.5$, we expect the smallest gaps to occur there; this is corroborated by our observations. It follows that we should see the period-3 oscillation described in Section 4.3.1, caused by the cycling symmetry patterns of the approximating orbits. Because the oscillations are small, it is not possible to see this in the graph of Figure 4.10. We evaluate the difference between the least-squares linear fit and the data; the graph of this in Figure 4.11 shows a very definite period-3 oscillation. We therefore compute the slope of $\log m(g_n)$ versus $\log q_n$ using every third data point and find an average value of $\eta = 1.0931$. This number is related to the renormalization scaling at $x = 0.5$ via a similar argument to that for maximum gap. Using the values from

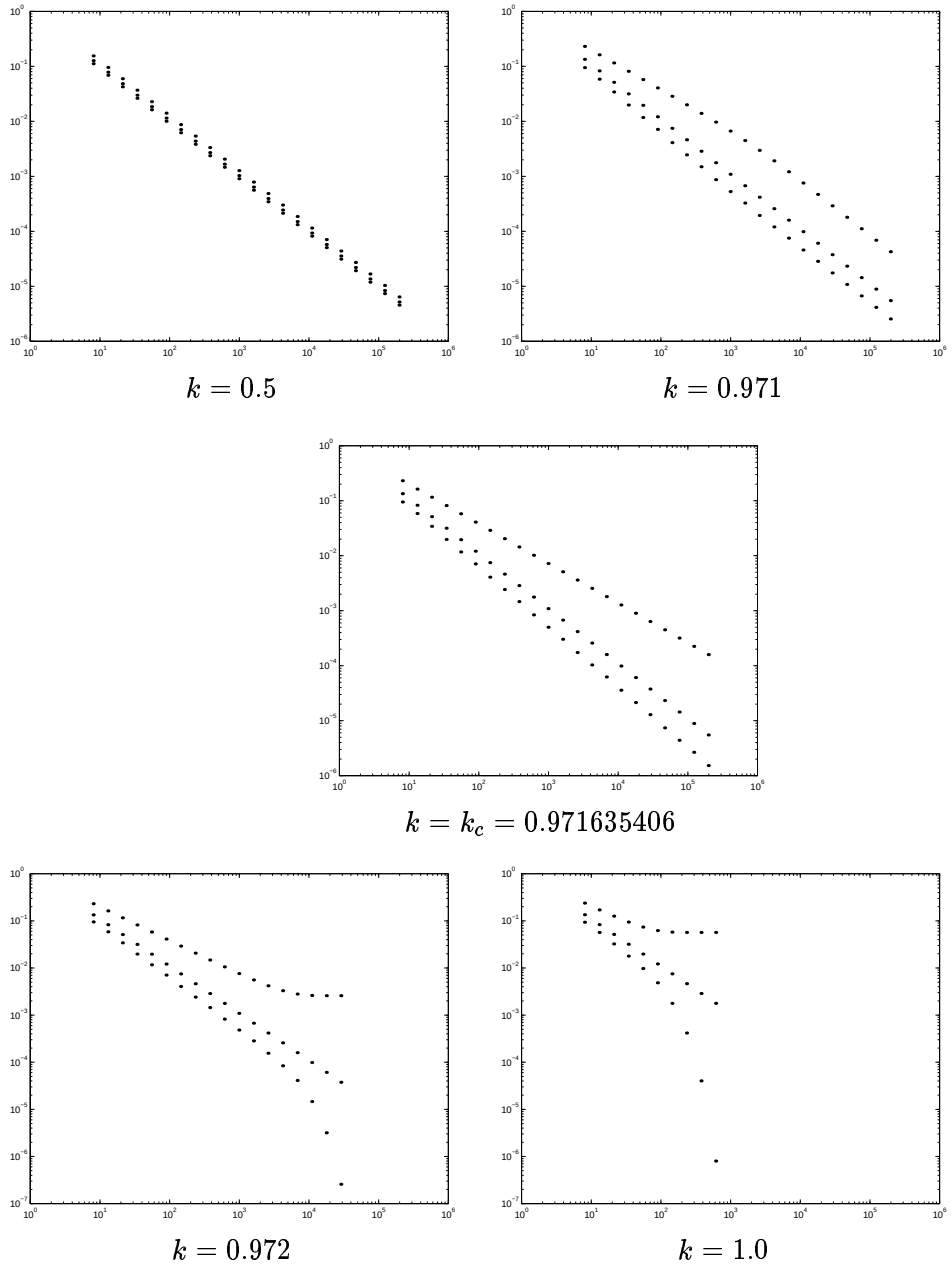


Figure 4.10: Maximum, minimum, and mean of the gap lengths for periodic orbit approximations to the golden orbit at different values of k.

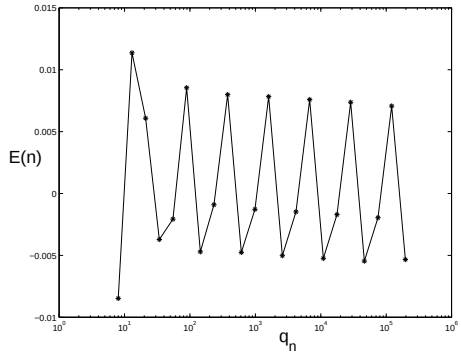


Figure 4.11: Error in the least squares linear fit to the minimum gap data for $k = k_c$ (see Figure 4.10). Specifically, we use $100 < q_n < 2 \times 10^5$ to fit the line and then compute the error as $e(n) = \log_{10}(m(g_n)) - (-1.0928 \log_{10}(q_n) - 0.0268)$. The period-3 oscillation is due to the periodic orbit symmetry properties.

(4.13), we therefore expect $\eta = \log(\sqrt[3]{\alpha_{.5}})/\log(\mu) = 1.09315$, which is exactly what we found. At $k = 0.971$ we see the same type of transitional behavior as before, with the data close to that of the critical circle for small q_n and slopes that converge to 1.00 as q_n gets larger.

For approximations to the Cantor sets, we see $m(g_n)$ drop off exponentially. A possible explanation for this starts with a model that gaps in the Cantor set decrease as $g_j = c\lambda^j$. A finite approximation to the Cantor set should then have smallest gap $m(g_n) \approx c\lambda^{q_n}$. In the log-log plot we therefore expect to see

$$\log m(g_n) = \log c + e^{\log q_n} \log \lambda.$$

For $k = 0.972$ we estimate $\lambda = 0.9997$ and $c = 0.0002$ and for $k = 1.0$, $\lambda = 0.983$ and $c = 0.02$. These numbers are calculated from least-square fits for data from the longest four approximations at each k, since these are orbits for which the largest gaps are resolved properly. Since such few data points are used to fit the line, these values of c and λ are not likely to be very accurate.

The linear scaling of the mean suggests it is the appropriate number to rescale by to get a limiting density distribution of gaps in the golden orbit. It is not obvious that this should exist or be well defined, but it might help give an expected form for the $C(\epsilon)$ curve which could then help distinguish between connected and disconnected sets. In the following section we pose some questions about the relationship between the density of points on an invariant circle and the distribution of gap sizes. This leads us into the problem of how to relate the density-measure for set of points to the statistics of the minimal spanning tree — a much larger project that is outside the scope of this thesis.

Renormalized gap distributions

We start by describing the distribution of points around an invariant circle; see Figure 4.12 for some example densities. We generate these graphs as histograms of x-coordinate values from very long minimizing periodic orbit approximations to the $1/\mu^2$ orbit. Sussman *et al.* [79] generate similar graphs using the linearization of the standard map and a highly accurate estimate of the tangent direction at a single point on the circle. For $k = 0$, the points of each periodic orbit are perfectly evenly spaced around the circle. This implies that the limiting

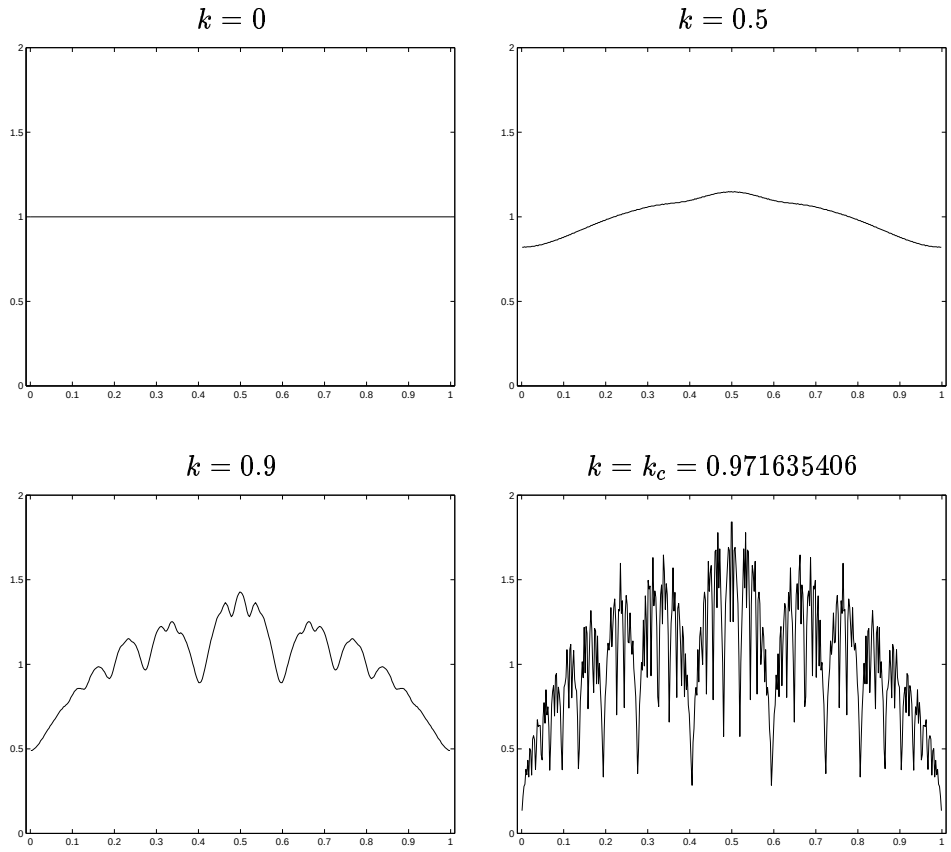


Figure 4.12: Relative density distribution of x-coordinate values for long periodic orbits ($q \sim 10^5$) approximating the golden circle at different choices of $k \leq k_c$. The number of bins in the histogram is the square-root of the number of points.

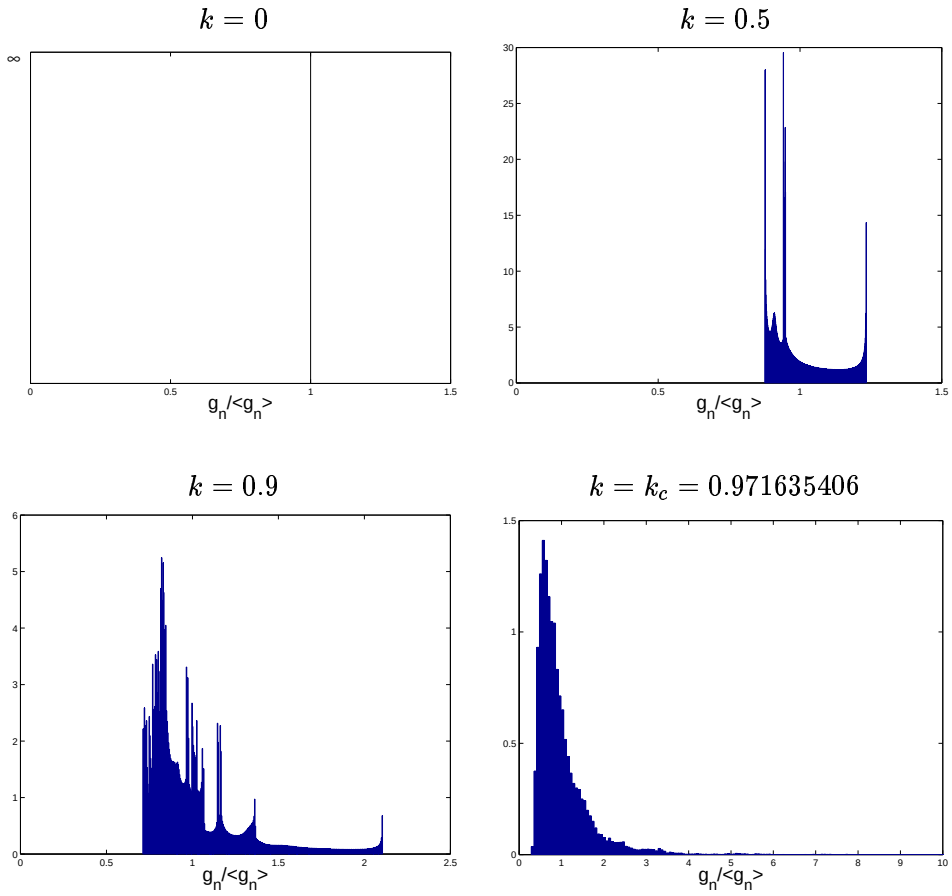


Figure 4.13: Distribution of gap lengths normalized by their mean for the same orbits and k values used in Figure 4.12. For $k = 0$, the distribution is a delta-function since the points in the orbit are evenly spaced around the circle.

distribution of points for an invariant circle is uniform. As k increases, this distribution becomes slightly less uniform with a minimum at $x = 0$ and maximum at $x = 0.5$. For k close to the critical value, the density becomes highly nonuniform, though still nonzero everywhere. At k_c the density drops to zero at $x = 0$, and at every point in the orbit of $x = 0$. The histogram approximation does not show this because of discretization effects. When $k > k_c$, the orbit approximates a Cantor set, so there are intervals of x-values where the density is zero. Since the cantori have Lebesgue measure zero, the support of the density has measure zero; it is therefore very difficult to generate a graph of this function using histograms.

The gap distributions graphed in Figure 4.13 are computed from the same orbits and k-values as Figure 4.12. Here, we have rescaled the gap lengths to make the mean equal to one in each case. Since the mean, $\langle g_n \rangle \sim L/q_n$, and $L \approx 1$, this rescaling amounts to multiplying the gaps by the number of points in the orbit. For $k = 0$, the gaps are exactly the same length and the distribution is therefore a delta function. When $k < k_c$, the gap distributions have a sharp maximum and minimum which correspond to the constants C and c computed above. A possible explanation for the spikes in the distribution is given in the next paragraph. For $k = k_c$, we know that $M(g_n) \sim C/q_n^{0.7211}$ and $m(g_n) \sim c/q_n^{1.093}$. We therefore expect the rescaled distributions to have a minimum that goes to zero and a maximum that goes to infinity as the number of points is increased. The rescaled gap distributions for the Cantor sets (i.e. $k > k_c$) should also have a minimum that goes to zero and a maximum that goes to infinity. The difference between a Cantor set and the critical circle is that the relative number of gaps less than the mean should diverge for the Cantor set.

It is not obvious how to analytically relate the density of points to the distribution of gap lengths. It seems intuitively reasonable that the longest gap will occur at the minimum in density and the shortest gap at the maximum and this is what we observe. In fact, the local rescaled gap size, $g(x)$, should be inversely proportional to the density of the points, i.e. $g(x) \sim 1/f(x)$. The relative number of gaps of a particular size, y, should then be related to the number of points where $1/f(x) = y$. This explains the spikes in the gap distributions for $k < k_c$ as the places where the density, $f(x)$, has a turning point. A formal derivation of these relationships requires some work but should be fairly straightforward. The results should be related to order statistics [10].

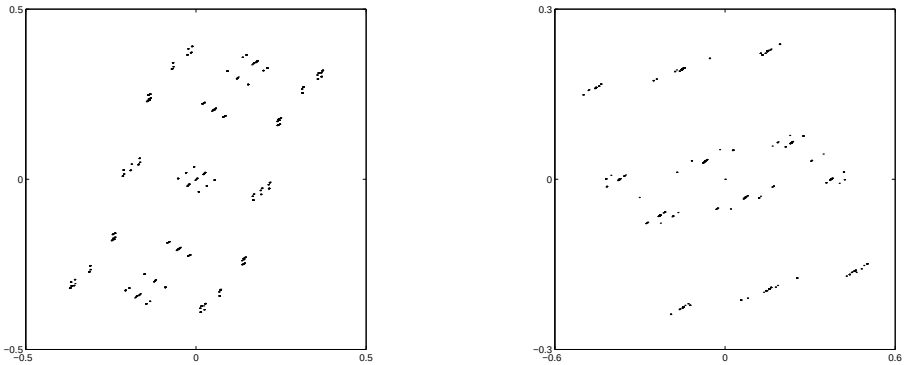
4.4 Cantori from sawtooth maps

The next set of examples we examine are orbits from 4d symplectic sawtooth maps. A cantorus is an invariant Cantor set of a symplectic mapping on which the dynamics is semi-conjugate to an incommensurate rotation. It is known that close enough to an "anti-integrable" limit, symplectic maps have cantori for all such rotations [48]. As we saw in the previous section, cantori in area-preserving twist maps arise from the destruction of invariant circles. For higher-dimensional maps, it is not known if the breakup of an invariant torus leads immediately to a cantorus, or if there are some intermediate stages; this question is one we intend to explore further (see Section 4.5.1). Below, we look at cantori from the anti-integrable limit.

A simple model for which cantori are analytically computable is the following piecewise linear map, called a sawtooth map:

$$\begin{aligned} p_{n+1} &= p_n + \nabla V(x_n) \\ x_{n+1} &= x_n + p_{n+1} \mod 1 \,. \end{aligned}$$

Here $x \in \mathbb{T}^2$ (the 2-torus) and $p \in \mathbb{R}^2$ and $V = \frac{1}{2} x^t A x$ is a quadratic potential. By choosing



(a) (b)

Figure 4.14: Two examples of cantori generated by symplectic sawtooth maps. Each orbit has 10^4 points.

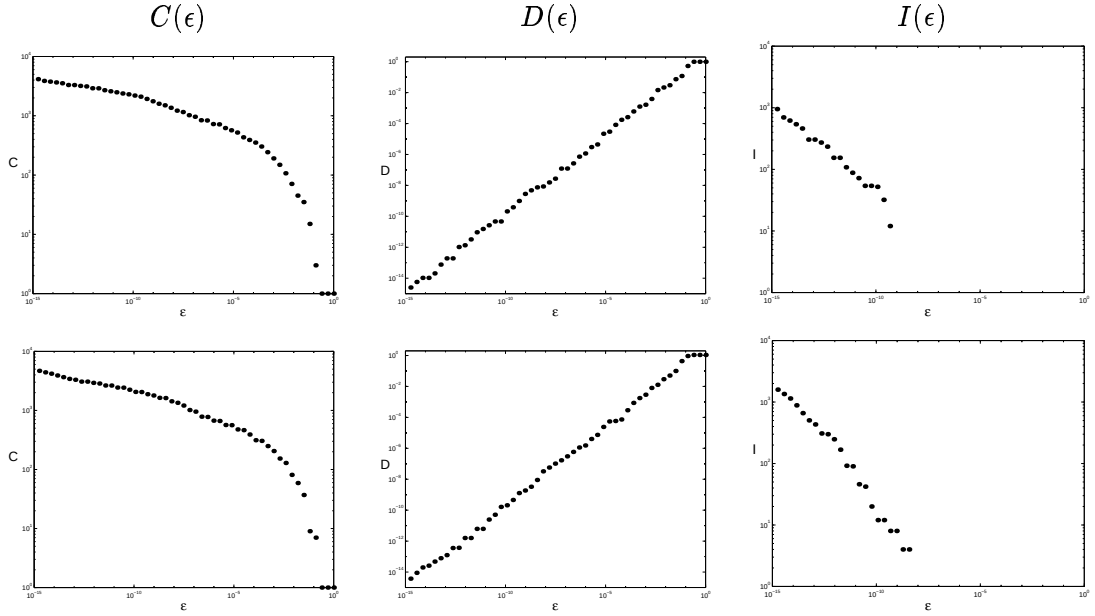
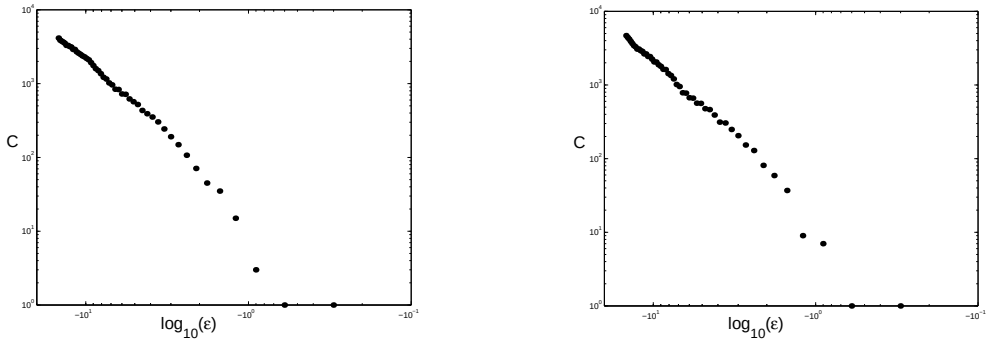


Figure 4.15: $C(\epsilon)$, $D(\epsilon)$ and $I(\epsilon)$ data for the two cantori. Top row: data for the cantorus in Figure 4.14(a). Bottom row: data for the cantorus in Figure 4.14(b). All axes are logarithmic. The horizontal axis range is $10^{-15} < \epsilon < 1$.



(a) (b)

Figure 4.16: $C(\epsilon)$ versus $\log(\epsilon)$ for the two cantori of Figure 4.14(a) and Figure 4.14 (b) respectively. All axes are logarithmic. The horizontal range is $-20 < \log(\epsilon) < -0.1$

an incommensurate pair of irrational rotation numbers, for example, the golden mean $\frac{\sqrt{5}-1}{2}$ and $\frac{1}{\sqrt{2}}$, we can find an orbit analytically by looking for the semi-conjugacy, $x_t = X(\theta + \omega t)$. In Figure 4.14, we show the projection of the cantorus onto the configuration plane (x coordinates) for two choices of ω: for Figure 4.14(a) $\omega = (\frac{1}{\sqrt{2}}, \frac{\sqrt{5}-1}{2})$, and for Figure 4.14(b) ω is the the spiral mean, (τ^{-1}, τ^{-2}) where τ is the real root of $\tau^3 - \tau - 1$. These orbits are generated by the same dynamical system since in both cases the matrix for the quadratic form is:

$$A \approx \begin{pmatrix} 1.9152 & -2.0358 \\ 0.5214 & 0.0847 \end{pmatrix}.$$

The graphs of $D(\epsilon)$ in Figure 4.15 are similar to those for previous Cantor sets. They tell us that the cantori are totally disconnected, because $D(\epsilon) \to 0$. We estimate δ to be very close to one: for the cantorus of Figure 4.14(a) $\delta = 0.99 \pm 0.05$; for that of Figure 4.14(b) $\delta = 1.03 \pm 0.04$. The graphs of $C(\epsilon)$ exhibit scaling behavior similar to that of the Cantor sets in the standard map (see Figure 4.6). There is no linear scaling region from which to determine γ, and this suggests that our assumption that $C(\epsilon) \sim \epsilon^{-\gamma}$ does not hold here. Since the underlying sets are Cantor sets, we must still see $C(\epsilon) \to \infty$ as $\epsilon \to 0$, but the growth rate is possibly logarithmic, rather than polynomial, in ϵ. To test this hypothesis, we plot $C(\epsilon)$ versus $\log(\epsilon)$ on logarithmic axes in Figure 4.16. The observed linear relationship tells us that $C(\epsilon) \sim (\log \epsilon)^{-\nu}$ near $\epsilon = 0$. We estimate the value of ν using a least-squares fit of the slope of the line, and find that $\nu = 2.14 \pm 0.05$ for both cases. This also implies that $\gamma = 0$. A result due to A. Fathi [25] implies that these cantori have Hausdorff dimension equal to zero, so again we see that γ and the dimension are equal for a zero-measure Cantor set.

The graphs of $C(\epsilon)$ for the two cantori are almost identical, as are their scaling rates, ν. Experimental observations show that the $C(\epsilon)$ data vary most strongly with the choice of eigenvalues for the quadratic form, rather than the eigenvectors or the frequencies. This suggests that there may be some relationship between the eigenvalues and the scaling rate, ν. A relationship between these numbers is not surprising, since the eigenvalues govern the contraction along an orbit, and contraction ratios are often associated with fractal dimensions.

4.5 Further applications

In this section we describe some possible further applications of our computational topology techniques. The list is far from exhaustive!

4.5.1 Torus breakup in the Froeschlé map

Hamiltonian flows with $d \geq 3$ degrees of freedom are modelled by symplectic maps of $\mathbb{R}^{2d}$ — higher-dimensional analogues of area-preserving maps. One example that symmetrically couples two standard maps, is called Froeschlé map:

$$\begin{aligned} y_1' &= y_1 + \frac{a}{2\pi}\sin(2\pi x_1) + \frac{c}{2\pi}\sin(2\pi(x_1+x_2)) \\ y_2' &= y_2 + \frac{b}{2\pi}\sin(2\pi x_2) + \frac{c}{2\pi}\sin(2\pi(x_1+x_2)) \\ x_1' &= x_1 + y_1' \mod 1 \\ x_2' &= x_2 + y_2' \mod 1. \end{aligned} \tag{4.15}$$

This map was originally introduced as a model of the time evolution of elliptic galaxies [27]. When $a = b = c = 0$, orbits live on two-tori embedded in four-dimensional space. These tori are labelled by the rotation frequencies of the orbit. The KAM-theorem implies that tori with incommensurate frequencies persist for small values of a, b, and c. We saw in Section 4.4 that far from integrability, the remnants of these tori are Cantor sets, but little is known about the structure of torus remnants for moderate perturbations. It is possible that these sets have the structure of a Cantor set of lines or a Sierpinski carpet [41]. We believe that the computational topology tools from Chapters 2 and 3 could be used to help investigate this problem.

The invariant tori do not have quite the same importance dynamically as the invariant circles of area-preserving maps since a two-dimensional surface cannot enclose a four-dimensional region. This observation led Arnold to reason that a chaotic orbit of any Hamiltonian system with three or more degrees of freedom can diffuse through a dense subset of the phase space. This Arnold diffusion is typically extremely slow, and one expects that transport will be more rapid when there are few invariant tori. It is therefore still desirable to have a better understanding of the structure of torus remnants.

Recall that in Section 4.3 we approximated quasiperiodic orbits by long periodic ones. We would like to use similar techniques to approximate orbits on invariant tori in the Froeschlé map. Unfortunately, many properties of area-preserving maps do not necessarily hold in higher-dimensional symplectic maps. There is no nice generalization of continued fractions, so the first problem is to generate rational approximants to pairs of incommensurate irrational numbers. One reasonably effective approach is a generalization of the Farey tree, due to Kim and Ostlund [38]. Second, although it is easy to generalize the twist condition, there has been no generalization of Aubry-Mather theory to higher-dimensional symplectic maps. However, numerical simulations suggest that symmetric periodic orbits do converge to quasiperiodic ones. It may also be possible to use frequency analysis techniques such as those of Laskar [43] or Sussman *et al.* [79] to approximate quasiperiodic orbits.

Preliminary work with periodic orbit approximations to the Froeschlé map shows that it is easy to find long orbits close to invariant tori when the perturbation parameters, a, b, and c, are small. As these parameters increase, the orbits become unstable; it becomes impossible to find orbits with more than a few hundred points. This suggests that orbit stability is a good predictor of the existence of invariant tori [70]. It also means that we cannot get enough data

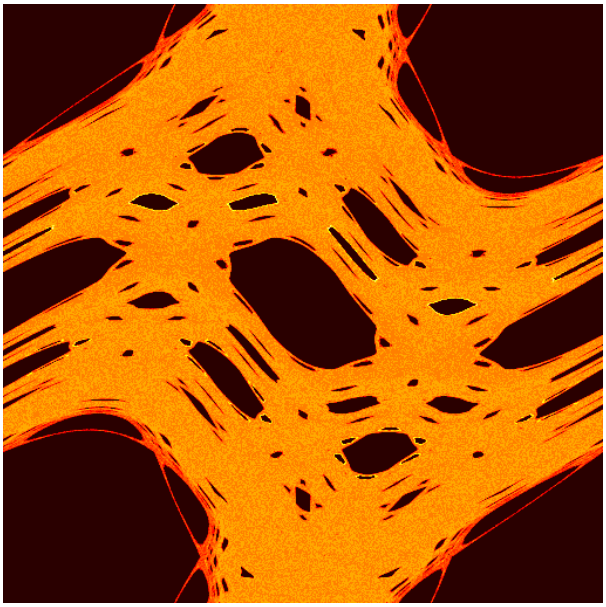


Figure 4.17: A two-dimensional histogram for a single chaotic orbit in the standard map with $k = 1.1$. We iterated the map 10^7 times, and binned the points in a 512×512 grid of boxes in the unit square $x, y \in [-0.5, 0.5]$. The color indicates the relative number of points in each box; yellow is high density, red is low density, black is zero.

to make a confident analysis of the topology — the cutoff resolution is too high. Until we can generate good approximations to quasiperiodic orbits in the Froeschlé map we will be unable to say much about their topological structure.

4.5.2 The chaotic fat fractal

As we described in Section 4.3, when $k > k_c$ all the rotational invariant circles of the standard map are destroyed and it is possible for a single chaotic orbit to cover most of phase space. The regions that remain inaccessible are the elliptic island resonance zones. This means the chaotic region contains holes at many different resolutions. Umberger and Farmer [84] show that the chaotic region is a fat fractal, and claim that the scaling they observe implies the existence of global scaling laws for the dynamics. For us, the chaotic region provides a nice example where we can test the relationship between the growth rate in the number of holes, γ_1, and the fat fractal exponent (defined in Chapter 5).

There are a few technical issues to address here. The natural way to generate a finite set of points that cover the chaotic region is to use one or several orbits with initial conditions near an unstable point. The problem with this is that an orbit takes a very long time ($\sim 10^{10}$ iterations) to cover the chaotic region uniformly [56]. Using so many data points is impractical, but if $k \gtrsim 1$, then orbits with around 10^4 points typically cover only a fraction of the entire chaotic region. We can reduce the amount of data we store by binning — i.e., we draw an $N \times N$ grid on the phase space and record only those squares the orbit hits. This essentially digitizes the data, and means the cutoff resolution is at least $1/N$. The alpha shape approach to computational homology described in Chapter 3 is not the most efficient way find holes in gridded data. The cubical complexes proposed in Section 3.4.3 would be a more natural setting.

4.5.3 Pruning outlying data points

In [90], Zahn proposed using the minimal spanning tree to identify outlying points from experimental measurements of particle tracks in bubble chambers. We think these ideas could be extended and used to remove noisy points from embedded chaotic time series data. Regular noise filtering techniques work by deleting a band of frequencies from the signal. This is inappropriate for chaotic signals since they typically have a wide spectrum of frequencies. It is possible to compute information about the dynamics, e.g. the Lyapunov exponents, from a chaotic attractor that is reconstructed by embedding the time series data. Reducing the amount of noise in the signal is desirable, as this will improve the accuracy of such information.

Our idea is to use the minimal spanning tree and nearest neighborhood graph to identify points as "outliers at resolution ϵ." Essentially, these will be the points isolated at resolution ϵ, so the NNG is of more interest than the MST. These ideas are very preliminary but they suggest the versatility of our techniques.

4.6 Concluding remarks

We have demonstrated the effectiveness of our computational topology algorithms by applying our techniques to a number of well-understood discrete dynamical systems. Our analysis of some sections of the Hénon attractor in Section 4.2 provides strong evidence confirming the intuition that they have Cantor-set structure. The cantori of Section 4.4 have sub-polynomial growth in the number of components —i.e., $\gamma = 0$— even though they are Cantor sets. We resolved this problem by showing that the number of components grew logarithmically. The most interesting results of the chapter came from studying the transition from invariant circle to Cantor set in the standard map. In Section 4.3 we showed that the graphs of the number and the size of ϵ-components distinguished between approximations to invariant circles and Cantor sets at a fairly crude level. We developed a more sensitive criterion for detecting the transition by examining the largest gaps in a sequence of periodic orbit approximations. An interesting avenue for further work is to investigate the relationship between the MST edge-lengths and the underlying point distributions from orbits on invariant circles or Cantor sets. We presented numerical evidence that as the number of points increases, the gap distributions scale by their means. Since the invariant circles and Cantor sets are essentially one-dimensional objects, this problem should have parallels with the study of order statistics [10]. A better understanding of the MST edge-lengths may have implications for the renormalization of the standard map phase space.

It should be noted that our techniques use purely geometric information, no knowledge of the underlying dynamics is needed. This is an advantage in some applications where nothing is known about the dynamical system. Typically, however, there is some information that can be exploited. Thus, computational approaches, such as Mischaikow's [59], that make explicit use of dynamical systems theory, have greater success at diagnosing the dynamics. We deliberately kept our theory and algorithms as general as possible, so that they can be adapted to different circumstances.

Chapter 5

Topological Growth Rates and Fractal Dimensions

5.1 Introduction

Throughout this thesis, we observe close correlations between values of the topological growth rates and various other fractal indices. These observations are based on both analytic derivations and numerical computations of the relevant exponents. In this chapter we derive inequalities that relate our topological growth rates to existing scaling indices such as the box-counting dimension and the Besicovitch-Taylor exponent. Such relationships lead to a better understanding of the topological growth rates.

The chapter has three sections. We start by giving definitions of box-counting dimension, fat fractal exponents and Besicovitch-Taylor index. These measures of fractal scaling have close connections with one another, and with the topological growth rates. Sections 5.3.1 to 5.3.3 examine the disconnectedness and discreteness indices, γ and δ for subsets of $\mathbb{R}$ and $\mathbb{R}^n$. The most detailed results are for compact totally disconnected subsets of the line; these are given in Section 5.3.1. Such sets are defined in terms of countably many complementary open intervals. It is well known that the fractal dimension is related to the scaling of the lengths of these deleted intervals. We adapt this result to show that γ and δ are also related to this scaling. In Section 5.3.2 we study subsets of higher-dimensional spaces, and obtain simple inequalities involving γ, δ, and the box-counting dimension $\dim_B$. We give examples in Section 5.3.3 to illustrate some of the cases for the inequalities of Sections 5.3.1 and 5.3.2. A consequence of the results in this chapter is that for zero measure Cantor subsets of $\mathbb{R}$

$$\gamma = \dim_B \quad \text{and} \quad (1-\gamma) \leq \delta \leq 1,$$

providing the appropriate limits exist. Although the disconnectedness index γ takes the same value as the box-counting dimension under these conditions, we emphasize that this does not imply that γ is a "fractal dimension." Any definition of fractal dimension should extend the classical notion, and therefore an m-dimensional manifold must have dimension m, for example. The disconnectedness index, however, is zero for any compact, connected manifold. In Section 5.3.4, we take a first step towards relating the growth rate of "k-dimensional holes" γ_k to the box-counting dimension.

We discuss some of the many open questions in the concluding section of this chapter. The most interesting unproven conjecture concerns strictly self-similar fractals. We have observed, for the examples in this thesis, that when a self-similar fractal has a non-zero γ_k, then it takes

the same value as the similarity dimension. This is not surprising — self-similarity is a strong condition and we expect it to dominate any scaling properties.

5.2 Definitions

We recall the necessary definitions of box-counting dimension, fat fractal exponents, and the Besicovitch-Taylor index — a scaling index that provides a link between our topological growth rates and fractal dimensions.

5.2.1 Box-counting dimension

We discussed the box-counting dimension and its relationship to the Hausdorff dimension briefly in Chapter 1. Here, we restate the definition and give an equivalent formulation in terms of ϵ-neighborhoods of a set.

Box-counting dimension

Recall from Chapter 1 that the box-counting dimension is defined in terms of covers of the fractal by sets of size ϵ. If $N(\epsilon)$ is the smallest number of sets with diameters at most ϵ needed to cover X, then

$$\dim_B = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{-\log \epsilon} \tag{5.1}$$

Of course, this limit may not exist, in which case the $\limsup$ and $\liminf$ are used. The corresponding limits are the upper and lower box-counting dimensions, $\overline{\dim}_B$ and $\underline{\dim}_B$. The number $N(\epsilon)$ can be defined in many ways, all of which yield an equivalent value of $\dim_B$ (see Falconer [23] for details). The definitions of $N(\epsilon)$ that we use in Section 5.3 are

1. the smallest number of closed balls of radius ϵ that cover X; and
2. the largest number of disjoint balls of radius ϵ with centers in X.

Minkowski dimension

The Minkowski dimension is the scaling rate of the Lebesgue measure of the ϵ-neighborhoods of X. We write $\mu(X)$ for the Lebesgue measure in $\mathbb{R}^n$, and X_ϵ for an ϵ-neighborhood of X. The definition of Minkowski dimension is as follows:

$$\dim_M = \lim_{\epsilon \to 0} \left[n - \frac{\log \mu(X_\epsilon)}{\log \epsilon} \right]. \tag{5.2}$$

If the limit does not exist we use the $\limsup$ and the $\liminf$. This definition of dimension is equivalent to box-counting. To see this, let $N(\epsilon)$ be the largest number of disjoint balls from definition 2 above. If we write c_n for the volume of the unit ball in $\mathbb{R}^n$ — i.e., $c_1 = 1,\ c_2 = \pi,\ c_3 = 4\pi/3$, etc. — then

$$\mu(X_\epsilon) \geq c_n \epsilon^n N(\epsilon).$$

If we triple the radius of the balls, we have that

$$\mu(X_\epsilon) \leq c_n (3\epsilon)^n N(\epsilon).$$

These bounds imply that $\dim_B = \dim_M$; see [23] for a more detailed proof. The two different formulations of box-counting dimension mean we can derive different types of relationships with the topological growth rates.

5.2.2 Fat fractal exponents

We have given a few examples of fat fractals in this thesis. Recall that these sets have positive Lebesgue measure and therefore integer Hausdorff and box-counting dimensions. It is possible to characterize the irregular structure of a fat fractal by modifying the definition of Minkowski dimension. The measure of the ϵ-neighborhoods X_ϵ converges to the measure of X; the fat fractal exponent characterizes the convergence rate as follows:

$$d_F = \limsup_{\epsilon \to 0} \left[n - \frac{\log(\mu(X_\epsilon) - \mu(X))}{\log \epsilon} \right] \tag{5.3}$$

The exponent d_F is not a dimension because it gives inconsistent values for the dimension of the unit n-cube, I^n, depending on the ambient space. If $I^n \subset \mathbb{R}^n$, then $d_F = n - 1$; but if $I^n \subset \mathbb{R}^m$ with $m > n$, then $d_F = n$.

Umberger *et al.* [20, 22, 84] give a finer characterization of the scaling of $\mu(X_\epsilon)$ by separating out contributions from "fattening" and "filling in holes" of the fractal. This involves considering the fattening of X to its ϵ-neighborhood X_ϵ and then the unfattening of X_ϵ to a set U_ϵ. The unfattening operation is achieved by fattening the complement of X_ϵ, i.e.,

$$U_\epsilon = \mathbb{R}^n - (\mathbb{R}^n - X_\epsilon)_\epsilon.$$

The set U_ϵ is larger than X — any structures of size less than ϵ are filled in or smoothed out. Now define $F(\epsilon) = \mu(U_\epsilon) - \mu(X)$, and $G(\epsilon) = \mu(X_\epsilon) - \mu(U_\epsilon)$. $F(\epsilon)$ is the measure of filled in holes — i.e., the small-scale structure — and $G(\epsilon)$ is the measure of fattening caused by large-scale structure. Umberger *et al.* define scaling rates for both $F(\epsilon)$ and $G(\epsilon)$ as $\epsilon \to 0$, and use these rates as a characterization of fat fractal structure. In the notation of [22]

$$\beta = \lim_{\epsilon \to 0} \frac{\log F(\epsilon)}{\log \epsilon} \quad \text{and} \quad \overline{\beta} = \lim_{\epsilon \to 0} \frac{\log G(\epsilon)}{\log \epsilon}.$$

Note that $F(\epsilon) + G(\epsilon) = \mu(X_\epsilon) - \mu(X)$ so $d_F = n - \min\{\beta, \overline{\beta}\}$.

5.2.3 The Besicovitch-Taylor index

This index is derived from a process of packing the complement of a set with regular cells. We start by describing the case for compact totally disconnected subsets of $\mathbb{R}$. Let $X \subset [a, b]$ be such a set. The complement $[a, b] - X$ is the union of a countable number of open intervals, U_i, for $i = 1, 2, \ldots$ [81]. That is,

$$X = [a, b] - \bigcup_1^\infty U_i. \tag{5.4}$$

We let $u_i = \mu(U_i)$, the Lebesgue measure of U_i, and assume that the sets are ordered by decreasing length, i.e., $u_1 \geq u_2 \geq \ldots$. Since $X \subset [a, b]$ and the U_i are disjoint, we have that

$$\mu(X) = b - a - \sum_1^\infty u_i.$$

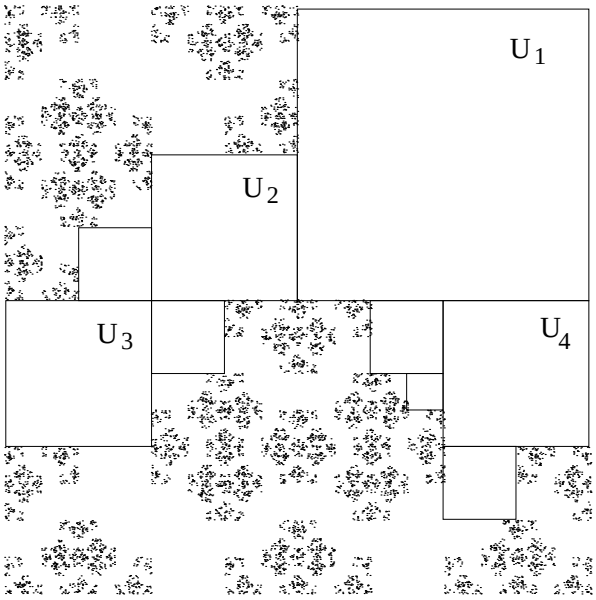


Figure 5.1: Packing the complement of a fractal.

The convergence of the series $\sum u_i$ can be characterized in a number of ways. The original formulation of Besicovitch and Taylor is the index:

$$d_{BT} = \inf\{\alpha \mid \sum_{1}^{\infty} u_i^{\alpha} < \infty\}. \tag{5.5}$$

Properties of convergent monotone series can be used to show that the following rates are equivalent to the Besicovitch-Taylor index [81].

$$d_{BT} = \limsup_{k\to\infty} \frac{\log k}{-\log u_k} \tag{5.6}$$

$$= \limsup_{k\to\infty} \left[1 - \frac{\log \sum_k^{\infty} u_i}{\log u_k}\right] . \tag{5.7}$$

We show in Section 5.3.1 that (5.6) has a close connection with our disconnectedness index.

Tricot [80] extends the definition of the Besicovitch-Taylor index to subsets $X \subset \mathbb{R}^n$ by packing a bounded complementary region with regular cells. For example, we can use n-cubes with faces that are parallel to the coordinate axes. Let U_0 be the smallest closed n-cube that contains X, and let u_0 be the side length of U_0. Now let U_1 be the largest cube in $U_0 - X$, U_2 be the largest cube in $U_0 - X - U_1$, and so on; see Figure 5.1. If we set u_i equal to the side length of U_i, then we can define the Besicovitch-Taylor index as in (5.5). The Lebesgue measure $\mu(U_i) = u_i^n$, so the series $\sum u_i^n$ converges, and therefore $d_{BT} \leq n$. The equivalent formula in (5.6) remains the same, but (5.7) becomes

$$d_{BT} = \limsup_{k\to\infty} \left[n - \frac{\log \sum_k^{\infty} u_i^n}{\log u_k}\right] . \tag{5.8}$$

We also note that the sets used in the packing can be more general than n-cubes; see [80] for details. The "cut-out sets" described in Falconer [24] involve similar ideas.

5.2.4 Topological growth rates

For ease of reference, we recall definitions from Chapters 2 and 3 for scaling rates in the number of components, size of components, and persistent Betti numbers.

Disconnectedness and discreteness

For disconnected sets, the rate of growth in the number of ϵ-connected components, $C(\epsilon)$, is measured by the disconnectedness index, γ. That is, $C(\epsilon) \sim \epsilon^{-\gamma}$, and

$$\gamma = \lim_{\epsilon \to 0} \frac{\log C(\epsilon)}{-\log \epsilon}. \tag{5.9}$$

The size of the ϵ-components is measured by the largest component diameter, $D(\epsilon)$. If a set is totally disconnected then $D(\epsilon) \sim \epsilon^{\delta}$ and

$$\delta = \lim_{\epsilon \to 0} \frac{\log D(\epsilon)}{\log \epsilon} \tag{5.10}$$

is the discreteness index. If the limits do not exist, we use the $\liminf$ or $\limsup$ and write $\gamma^{\inf}, \gamma^{\sup}, \delta^{\inf}$ or $\delta^{\sup}$ for the corresponding indices. We also note that the resolution parameter ϵ is related to distances between points in the set.

Growth rates of Betti numbers

In Chapter 3 we introduced the notion of persistent Betti number, $\beta_k^0(X_\epsilon)$, to count the number of k-dimensional holes in a space as a function of resolution. Here, the parameter ϵ relates to the ϵ-neighborhood, so it is a radius measurement. If $\beta_k^0(X_\epsilon) \to \infty$ as $\epsilon \to 0$, then we characterize the rate of growth by the following index

$$\gamma_k = \lim_{\epsilon \to 0} \frac{\log \beta_k^0(X_\epsilon)}{-\log \epsilon}. \tag{5.11}$$

As always, if the limit does not exist, we use the $\limsup$ or $\liminf$. Recall that for $k = 0$, the Betti number is just the number of connected components, so the definition of γ_0 agrees with that for the disconnectedness index, γ. In the definition of γ_i, we compare the number of holes to their size. The Besicovitch-Taylor index also compares a number with a size parameter; this is the reason we expect a link between the two.

5.3 Results

In this section, we derive a number of inequalities that relate our topological growth rates to different fractal scaling indices. Sections 5.3.1 and 5.3.2 examine the disconnectedness and discreteness indices. The first results are for totally disconnected subsets of an interval. The inequalities are straightforward consequences of existing results that relate the Besicovitch-Taylor index and the Minkowski dimension. We then consider, in Section 5.3.2, totally disconnected subsets of higher-dimensional spaces. The examples of Section 5.3.3 are mainly Cantor subsets of $[0, 1]$, chosen so as to illustrate various cases of equality and inequality for the results of the two preceding sections. Finally, in Section 5.3.4, we take a first step towards relating the growth rates of Betti numbers to fractal dimensions.

5.3.1 Subsets of the line

We start with compact totally disconnected subsets of the real line and show that the disconnectedness index, γ, is closely related to the Besicovitch-Taylor index, which is in turn related to the Minkowski dimension and the fat fractal exponent. We then derive general bounds for the discreteness index, δ.

Disconnectedness

Suppose $X \subset \mathbb{R}$ is compact and totally disconnected, and let X_ϵ denote an ϵ-neighborhood of X. As in (5.4) the complement of X is a union of open intervals, U_i. The number of ϵ-connected components $C(\epsilon)$ of X is just one more than the number of complementary intervals with length $u_i \geq \epsilon$. This gives us a way to relate the disconnectedness index, γ, and d_{BT}. Given ϵ, choose n so that $u_n < \epsilon \leq u_{n-1}$. Then $C(\epsilon) = n$ and

$$\frac{\log n}{-\log u_n} \leq \frac{\log C(\epsilon)}{-\log \epsilon} \leq \frac{\log n}{-\log u_{n-1}} = \frac{\log u_n}{\log u_{n-1}} \, \frac{\log n}{-\log u_n} \, . \tag{5.12}$$

Following [21] we define

$$L = \lim_{n\to\infty} \frac{\log u_n}{\log u_{n-1}}.$$

This quantity satisfies $1 \leq L \leq \infty$. Taking the limit of each each quantity in (5.12), we have that

$$\boxed{d_{BT} \leq \gamma \leq L \, d_{BT} \, .} \tag{5.13}$$

It is argued in [21] that for physical examples, $L = 1$, and then $d_{BT} = \gamma$. In general, however, L can be arbitrarily large — e.g., if $a, b > 1$, set $u_i = a^{-(b^i)}$, then $L = b$. If the limits in (5.12) do not exist, we can obtain similar results to (5.13) by using the $\limsup$ or $\liminf$.

Both Falconer [24] and Tricot [81] derive inequalities involving d_{BT} and the Minkowski dimension $\dim_M$ when X has zero Lebesgue measure. These results therefore extend to γ. In summary, the theorem of Section 3.4 in [81] shows that

$$d_{BT}^{\sup} = \overline{\dim}_M . \tag{5.14}$$

Slightly different results in Falconer [24] imply the above, and also that

$$\frac{\underline{\dim}_M (1 - \overline{\dim}_M)}{(1 - \underline{\dim}_M)} \leq d_{BT}^{\inf} \leq \underline{\dim}_M . \tag{5.15}$$

The above inequalities tell us that for totally disconnected subsets of $\mathbb{R}$ with zero measure, the limit $\dim_M$ exists if and only if the limit d_{BT} exists, in which case they are equal. Translating this into an expression for γ we have, providing the limits exist,

$$\boxed{\dim_M \leq \gamma \leq L \, \dim_M \, .} \tag{5.16}$$

A proof

To illustrate the techniques involved in proving the above inequalities, we give a proof of (5.14) following that in Tricot [81]. We start by observing that since the lengths u_i are decreasing, for sufficiently small $\epsilon > 0$ we can find an integer n such that

$$u_n \leq 2\epsilon < u_{n-1}.$$

Now consider the measure of the ϵ-neighborhood of X — this can be broken down as follows:

$$\mu(X_\epsilon) = \mu(X) + 2\epsilon n + \sum_{i=n}^{\infty} u_i. \tag{5.17}$$

The second term represents the overlap of the ϵ-neighborhood into gaps of length greater than 2ϵ and the third term is the length of the gaps that are filled in completely. For zero-measure sets, the first term disappears.

The following proof uses critical exponent definitions of d_{BT} and $\dim_M$, rather than the limit formulations given in (5.2) and (5.5). Specifically, the Minkowski dimension is

$$\dim_M = \inf\{\alpha \mid \epsilon^{\alpha-1}\mu(X_\epsilon) \to 0\}. \tag{5.18}$$

Version (5.6) of the Besicovitch-Taylor index is equivalent to

$$d_{BT} = \inf\{\alpha \mid n u_n^\alpha \to 0\}. \tag{5.19}$$

See [81] for a proof that these definitions are equivalent to the earlier ones. We now compare critical exponents for the left and right sides of (5.17) proceeding in two stages. The first step shows that $d_{BT} \leq \dim_M$, the second that $\dim_M \leq d_{BT}$.

Step 1. $d_{BT} \leq \dim_M$. Multiplying both sides of (5.17) by $\epsilon^{\alpha-1}$, we have

$$\epsilon^{\alpha-1}\mu(X_\epsilon) = 2\epsilon^\alpha n + \epsilon^{\alpha-1}\sum_{i=n}^{\infty} u_i.$$

If $\alpha > \dim_M$, then by definition, $\epsilon^{\alpha-1}\mu(X_\epsilon) \to 0$, which implies that the right side also tends to zero. Thus, $2\epsilon^\alpha n \to 0$ and since $\epsilon \geq u_n/2$, we have that $2^{1-\alpha}u_n^\alpha n \to 0$ and this implies $\alpha \geq d_{BT}$. Therefore, $d_{BT} \leq \dim_M$.

Step 2. $\dim_M \leq d_{BT}$. Conversely, without loss of generality we can assume that $d_{BT} < 1$. (This is because if $d_{BT} = 1$, then step 1 shows that $\dim_M \geq 1$, but $\dim_M \leq 1$ from its definition, so we are done.) Now choose α such that $d_{BT} < \alpha < 1$. Again we have that

$$\epsilon^{\alpha-1}\mu(X_\epsilon) = 2\epsilon^\alpha n + \epsilon^{\alpha-1}\sum_{i=n}^{\infty} u_i.$$

Since $u_n \leq 2\epsilon < u_{n-1}$, and $\alpha - 1 < 0$ we have that $\epsilon^\alpha < (u_{n-1}/2)^\alpha$ and $\epsilon^{\alpha-1} \leq (u_n/2)^{\alpha-1}$. Therefore

$$\epsilon^{\alpha-1}\mu(X_\epsilon) \leq 2^{1-\alpha}u_{n-1}^\alpha n + 2^{1-\alpha}u_n^{\alpha-1}\sum_{i=n}^{\infty} u_i.$$

We want to show that the right side goes to zero. The first term does because $\alpha > d_{BT}$ which means $u_{n-1}^{\alpha} n \to 0$. We can rewrite the second term (dropping the $2^{1-\alpha}$) as

$$u_n^{\alpha-1} \sum_{i=n}^{\infty} u_i = u_n^{\alpha} + \sum_{i=n+1}^{\infty} \frac{u_i}{u_n^{1-\alpha}}.$$

We next show that $\sum_{i=n}^{\infty} u_i^{\alpha} \to 0$ as $n \to \infty$. Choose β such that $d_{BT} < \beta < \alpha$. Since $\beta > d_{BT}$ there is an integer N so that for $n \geq N$, $nu_n^{\beta} < 1$, i.e. $u_n < n^{-1/\beta}$. Thus

$$\sum_{i=n}^{\infty} u_i^{\alpha} \leq \sum_{i=n}^{\infty} 1/i^{\alpha/\beta},$$

and $\alpha/\beta > 1$ so the right side tends to zero. Putting all the pieces back together, we have that $\epsilon^{\alpha-1}\mu(X_\epsilon) \to 0$ as $\epsilon \to 0$, implying that $\alpha \geq \dim_M$ and therefore that $\dim_M \leq d_{BT}$.

Remark. If X is a fat fractal, we can subtract $\mu(X)$ from each side of (5.17) and obtain results identical to (5.14) and (5.15) for the fat fractal exponent, d_F, instead of $\dim_M$.

Discreteness

For a totally disconnected subset $X \subset [a, b]$, the disconnectedness index, γ, is independent of the arrangement of the complementary intervals, U_i, within $[a, b]$. This is not true of the discreteness index, δ. In this section, we derive bounds on δ that are independent of the arrangement of complementary intervals. The argument is the same as one we used in Chapter 2 for a Cantor set with $\gamma = 0$.

Let $u_1 \geq u_2 \geq u_3 \geq \ldots$ be the lengths of the U_i. If $u_{n+1} < \epsilon \leq u_n$, then the largest ϵ-component must be longer than the next interval to be removed, so

$$D(\epsilon) \geq u_{n+1}.$$

If n is large enough that $u_n < 1$ and $D(\epsilon) < 1$, then

$$\frac{\log D(\epsilon)}{\log \epsilon} \leq \frac{\log u_{n+1}}{\log u_n}.$$

Taking the limit on both sides we have that $\delta \leq L$.

On the other hand, the diameter cannot exceed the total length of what remains of the interval $[a, b]$, so

$$D(\epsilon) \leq b - a - \sum_{1}^{n} u_i = \mu(X) + \sum_{n+1}^{\infty} u_i.$$

We assume again that $u_{n+1} < 1$ and $D(\epsilon) < 1$ so that

$$\frac{\log D(\epsilon)}{\log \epsilon} \geq \frac{\log[\mu(X) + \sum_{n+1}^{\infty} u_i]}{\log u_{n+1}}.$$

If $\mu(X) = 0$, then the quantity on the right is related to the Besicovitch-Taylor index via (5.7). Taking the limit on both sides, we find that $\delta \geq (1 - d_{BT})$. If $\mu(X) > 0$, then all we have is that $\delta \geq 0$. To summarize, if $\mu(X) = 0$ and the appropriate limits exist, then

$$\boxed{(1 - d_{BT}) \leq \delta \leq L.} \tag{5.20}$$

We give examples in Section 5.3.3 to illustrate the results obtained here.

5.3.2 Disconnected subsets of $\mathbb{R}^n$

In this section we explore connections between the box-counting dimension $\dim_B$ and the disconnectedness and discreteness indices γ and δ, when X is a compact totally disconnected subset of $\mathbb{R}^n$. We start by showing that for any set X for which the limits exist,

$$\boxed{\gamma \leq \dim_B .} \tag{5.21}$$

This follows from comparing the number of ϵ-connected components, $C(\epsilon)$ with the largest number of disjoint $\epsilon/2$-balls with centers in X, $N(\epsilon/2)$ (i.e., definition 2 on page 102). Since any two ϵ-components are separated by a distance of at least ϵ, any two balls of radius $\epsilon/2$ with centers in different ϵ-components must be disjoint. It follows that

$$C(\epsilon) \leq N(\epsilon/2).$$

If $\epsilon < 1$ we have that

$$\frac{\log C(\epsilon)}{-\log \epsilon} \leq \frac{\log N(\epsilon/2)}{-\log \epsilon} = \frac{\log N(\epsilon/2)}{-\log(\epsilon/2) - \log 2}.$$

By taking the limit as $\epsilon \to 0$ on each side, it follows that $\gamma \leq \dim_B$. If the limits do not exist, we still have that

$$\gamma^{\inf} \leq \underline{\dim}_B \quad \text{and} \quad \gamma^{\sup} \leq \overline{\dim}_B.$$

Any connected fractal (e.g., the Sierpinski triangle) has $\gamma < \dim_B$, since a connected set with more than one point has $\gamma = 0$ and $\dim_B \geq 1$. More interesting examples — for which the inequality is strict — are fat Cantor sets in $\mathbb{R}^1$ for which $\dim_B = 1$, but $\gamma < 1$ (see the example in Section 5.3.3). We have also seen examples of self-similar Cantor sets where equality holds in (5.21).

Next, we show that if X is totally disconnected and the appropriate limits exist, then

$$\boxed{\dim_B \leq \frac{\gamma}{\delta}.} \tag{5.22}$$

We again start by considering the ϵ-connected components of X. The number of ϵ-components is $C(\epsilon)$ and the largest ϵ-component diameter is $D(\epsilon)$. We set $r = D(\epsilon)/2$, and let $N(r)$ be the smallest number of r-balls needed to cover X (i.e., definition 1 on page 102). Clearly $C(\epsilon)$ balls with radius r will cover X, so that

$$N(r) \leq C(\epsilon).$$

From this inequality it follows that when $\epsilon < 1$,

$$\frac{\log N(r)}{-\log \epsilon} \leq \frac{\log C(\epsilon)}{-\log \epsilon}.$$

If we multiply the left side by $\log r / \log r$ and rearrange we have

$$\frac{\log r}{\log \epsilon} \, \frac{\log N(r)}{-\log r} \leq \frac{\log C(\epsilon)}{-\log \epsilon}.$$

But $r = D(\epsilon)/2$ so

$$\frac{(\log D(\epsilon) - \log 2)}{\log \epsilon} \frac{\log N(r)}{-\log r} \leq \frac{\log C(\epsilon)}{-\log \epsilon}. \tag{5.23}$$

Since X is totally disconnected we know that $D(\epsilon) \to 0$ (Lemma 3 in Chapter 2). If we assume that the limit defining δ exists and is nonzero, then the limit as $\epsilon \to 0$ and the limit as $r \to 0$ are equivalent. We can therefore take the limits on both sides of the inequality and find

$$\delta \dim_B \leq \gamma.$$

If the limits do not exist then we can use the limsup or liminf instead. We must be a little more careful when deriving the inequalities since for positive functions, $f, g > 0$

$$\liminf[f(x)g(x)] \geq [\liminf f(x)][\liminf g(x)]$$

and

$$\limsup[f(x)g(x)] \leq [\limsup f(x)][\limsup g(x)].$$

Taking the $\liminf$ in (5.23) we have

$$\liminf \frac{\log D(\epsilon)}{\log \epsilon} \liminf \frac{\log N(r)}{-\log r} \leq \liminf \frac{\log C(\epsilon)}{-\log \epsilon}.$$

And for the $\limsup$

$$\limsup \frac{\log N(r)}{-\log r} \leq \limsup \frac{\log C(\epsilon)}{-\log \epsilon} \limsup \frac{\log \epsilon}{\log D(\epsilon)}.$$

Since $\limsup(1/x) = 1/(\liminf x)$, it follows that:

$$\underline{\dim}_B \leq \frac{\gamma^{\inf}}{\delta^{\inf}} \quad \text{and} \quad \overline{\dim}_B \leq \frac{\gamma^{\sup}}{\delta^{\inf}} \ .$$

Finally, putting (5.21) and (5.22) together, tells us that when the limits exist and $\gamma \neq 0$ then

$$\boxed{\delta \leq 1.} \tag{5.24}$$

All of the above inequalities are consistent with the results obtained in the previous section for Cantor subsets of the line. In fact, for totally disconnected subsets of $\mathbb{R}$ with zero Lebesgue measure, we have from (5.16) and (5.21) that $\gamma = \dim_B$, and if $\gamma \neq 0$, then $(1 - \gamma) \leq \delta \leq 1$.

5.3.3 Examples

We now discuss some examples that illustrate various cases of the relationships between dimensions and the discreteness and disconnectedness indices.

Middle-third Cantor set

This Cantor set is constructed by successively removing the middle third of each remaining interval. There are 2^{k-1} complementary intervals with lengths $g_k = (\frac{1}{3})^k$. From the formulas for middle-α Cantor sets in Chapter 2 we have that

$$\gamma = \frac{\log 2}{\log 3} \quad \text{and} \quad \delta = 1.$$

Since the set is self-similar, we know that

$$\dim_B = \dim_M = \frac{\log 2}{\log 3}.$$

The convergence rate of the gap lengths is the limit

$$L = \lim \frac{\log g_{k+1}}{\log g_k} = \lim \frac{(k+1)\log \frac{1}{3}}{k \log \frac{1}{3}} = 1.$$

To compute d_{BT} we need the total number of gaps with lengths $\geq g_k$; this is just

$$n_k = \sum_{i=1}^{k-1} 2^i = 2^k.$$

Therefore,

$$d_{BT} = \lim \frac{\log n_k}{-\log g_k} = \frac{\log 2}{\log 3}.$$

It follows that equality holds in all the appropriate relationships derived above — i.e. (5.13), (5.16), (5.22), and (5.24).

A fat Cantor set

We examine the same fat Cantor set as in Chapter 2. Recall that, $K \subset [0,1]$ and there are 2^{k-1} gaps of size $g_k = (\frac{1}{2})^{2k-1}(\frac{1}{10})$ for $k = 0, 1, 2, \ldots$. The set has positive Lebesgue measure so $\dim_B = \dim_M = 1$. We showed that when the gaps are removed from the centers of intervals,

$$\gamma = \frac{1}{2} \quad \text{and} \quad \delta = \frac{1}{2}.$$

The convergence rate of the gap lengths is again

$$L = \lim \frac{\log g_{k+1}}{\log g_k} = \lim \frac{(2k+1)\log \frac{1}{2} + \log \frac{1}{10}}{(2k-1)\log \frac{1}{2} + \log \frac{1}{10}} = 1.$$

For the Besicovitch-Taylor index we have that the total number of gaps with lengths $\geq g_k$ is 2^k, so

$$d_{BT} = \lim \frac{\log n_k}{-\log g_k} = \lim \frac{k \log 2}{(2k-1)\log 2 + \log 10} = \frac{1}{2}.$$

We see that $\gamma = d_{BT}$, and $\dim_B = \gamma/\delta$.

Finally, we show that the fat fractal exponent for this set is also $\frac{1}{2}$, using the formula (5.17) for the measure of the ϵ-neighborhood of K. Given ϵ, choose k so that $g_k < 2\epsilon \leq g_{k-1}$. There are a total of 2^{k-1} gaps longer than 2ϵ and the length of these gaps is the sum:

$$\sum_{n=1}^{k-1} (\tfrac{1}{10}) 2^{n-1} (\tfrac{1}{2})^{2n-1} = \tfrac{1}{10}(1 - (\tfrac{1}{2})^{k-1}).$$

Since $\sum_1^\infty 2^{n-1} g_n = \frac{1}{10}$, it follows that the total length of all gaps less than 2ϵ is $(\frac{1}{10})(\frac{1}{2})^{k-1}$. From (5.17) we therefore have that

$$\mu(K_\epsilon) - \mu(K) = 2\epsilon(2^{k-1}) + (\tfrac{1}{10})(\tfrac{1}{2})^{k-1}.$$

By our choice of ϵ, we have

$$(\tfrac{3}{10})(\tfrac{1}{2})^k \leq \mu(K_\epsilon) - \mu(K) \leq (\tfrac{3}{10})(\tfrac{1}{2})^{k-1}.$$

Using this in the definition of fat fractal exponent (5.3), we find that $d_F = \frac{1}{2}$. Thus, we see that $d_{BT} = d_F = \gamma$.

A countable totally disconnected set

Finally, we consider the set

$$X = \{0, 1, \tfrac{1}{2}, \tfrac{1}{3}, \dots\}.$$

This set is totally disconnected but not perfect. Falconer shows [23] that the Hausdorff and box-counting dimensions differ for this set — the set is countable, so $\dim_H = 0$, but $\dim_B = \frac{1}{2}$. The distance between neighboring points in the set is

$$g_n = \frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)},$$

so $L = 1$, and $d_{BT} = \frac{1}{2}$. To compute the disconnectedness and discreteness growth rates, let ϵ_n be any number such that $g_{n+1} \leq \epsilon_n < g_n$. Then the points $1, \dots, 1/n$ are ϵ_n-isolated and the rest belong to a single ϵ_n-component so that $C(\epsilon_n) = n + 1$. The largest ϵ_n-component is always the tail of the sequence: $[0, 1/(n+1)]$, which means $D(\epsilon_n) = 1/(n+1)$. Thus,

$$\delta = \lim_{n\to\infty} \frac{-\log(n+1)}{-\log n(n+1)} = \frac{1}{2},$$

and

$$\gamma = \lim_{n\to\infty} \frac{\log(n+1)}{\log n(n+1)} = \frac{1}{2}.$$

This example shows that it is possible to have $\gamma = \dim_B$ but $\dim_B < \gamma/\delta$.

We observed in Chapter 2 that the Cantor set examples with zero Lebesgue measure had $\delta = 1$. We conjecture that this is the case for all zero-measure Cantor sets. The example of the countable sequence of points described above is a totally disconnected set with zero measure, but $\delta \neq 1$. It follows that if our conjecture is true, then the proof will have to make explicit use of the fact that Cantor sets are perfect, i.e., that they have no isolated points.

5.3.4 Other subsets of $\mathbb{R}^n$

We now examine fractal subsets of $\mathbb{R}^n$ that have unbounded growth in the number of $(n-1)$-dimensional non-bounding cycles. Suppose that $X \subset \mathbb{R}^n$ and that X is a compact, connected fractal with persistent Betti number $\beta^0_{n-1}(\epsilon) \sim \epsilon^{-\gamma_{n-1}}$ as $\epsilon \to 0$. Under these conditions we can show that the growth rate γ_{n-1} is bounded above by the Minkowski dimension $\dim_M$ if $\mu(X) = 0$ More generally, if $\mu(X) \geq 0$, γ_{n-1} is bounded by the fat fractal exponent d_F:

$$\boxed{\gamma_{n-1} \leq d_F.} \tag{5.25}$$

We start by defining a type of Besicovitch-Taylor index for the sequence of persistent hole sizes. From the definition of persistent Betti number, we know that if $\beta^0_{n-1}(\epsilon) = N > 0$, then there are N distinct $(n-1)$-cycles in the ϵ-neighborhood of X. The presence of an $(n-1)$-cycle in X_ϵ implies the existence of an n-ball with radius ϵ in the bounded complement of X. Therefore, if U_0 is the smallest n-ball containing X, and $\beta^0_{n-1}(\epsilon) = N > 0$, then there are N disjoint balls $B_i(\epsilon) \subset U_0 - X$. Now consider the values of ϵ where there is a jump in the value of $\beta^0_{n-1}(\epsilon)$. These ϵ-values characterize the size of a newly-created persistent hole since they define the largest possible radius of a ball that fits inside the corresponding hole in X. Let ϵ_i be the sequence of values where $\beta^0_{n-1}(\epsilon)$ is discontinuous, and let N_i be the difference between the left and right limits of $\beta^0_{n-1}(\epsilon)$ at ϵ_i, i.e., the number of holes with size ϵ_i. In order to define a Besicovitch-Taylor index, we list the radii of the persistent holes in decreasing order, with their multiplicity, and obtain a sequence, $r_1 \geq r_2 \geq r_3 \geq \ldots$ with $r_k \to 0$. The index is then just

$$d_r = \limsup_{k\to\infty} \frac{\log k}{-\log r_k}\ . \tag{5.26}$$

This index has identical equivalent formulations as for the Besicovitch-Taylor index in Section 5.2.3. Despite this similarity, d_r is not the same index as that obtained by packing the complement with cubes; the latter will detect fractal boundaries as well as the growth rate of holes.

The index d_r is closely related to γ_{n-1}. Given a sufficiently small $\epsilon > 0$, we can choose k so that $r_{k+1} < \epsilon \leq r_k$. It follows that $\beta^0_{n-1}(\epsilon) = k$ and therefore that

$$\frac{\log k}{-\log r_{k+1}} < \frac{\log \beta^0_{n-1}(\epsilon)}{-\log \epsilon} \leq \frac{\log k}{-\log r_k}. \tag{5.27}$$

The limit of the quantity on the left is not quite d_r; the convergence of the sequence r_k plays a role. As in Section 5.3.1, we introduce the factor

$$L = \lim_{k\to\infty} \frac{\log r_{k+1}}{\log r_k}.$$

Taking limits of each quantity in (5.27) we find that

$$\tfrac{1}{L} d_r \leq \gamma_{n-1} \leq d_r. \tag{5.28}$$

If $L = 1$ (a common case) then $d_r = \gamma_{n-1}$.

We now show that

$$d_r \leq d_F.$$

The proof is similar to one in [80], where inequalities involving the Besicovitch-Taylor index and fat fractal exponent are derived. The idea is to relate the size of sets that fill in the complement of X to the measure of X_ϵ. As we remarked earlier, X is compact and connected, so there are balls of each radius r_k in the bounded complement of X, i.e. $B(r_k) \subset U_0 - X$. These balls are disjoint, and if $r_k < \epsilon$, then $B(r_k) \subset X_\epsilon - X$. It follows that

$$\mu(X_\epsilon - X) \geq \sum_{i=k}^{\infty} \mu(B(r_i)) = \sum_{i=k}^{\infty} c_n r_i^n.$$

The integer k is the smallest such that $r_k < \epsilon$ and the constant c_n is the measure of the unit n-ball in $\mathbb{R}^n$ (as on page 102).

From this inequality it follows that

$$\log \mu(X_\epsilon - X) \geq \log \sum_{i=k}^{\infty} c_n r_i^n.$$

Assuming $r_k \leq \epsilon < 1$, we have that $0 < -\log \epsilon \leq -\log r_k$ so

$$\frac{\log \mu(X_\epsilon - X)}{-\log \epsilon} \geq \frac{\log \sum_{i=k}^{\infty} c_n r_i^n}{-\log r_k}.$$

From the definition of d_F (5.3)

$$d_F \geq \limsup_{k \to \infty} \left[n - \frac{\log \sum_{i=k}^{\infty} r_i^n + \log c_n}{\log r_k} \right].$$

The quantity on the RHS is equivalent to d_r by (5.8), so $d_r \leq d_F$.

It follows from this result that $\gamma_{n-1} \leq d_F$. If X has zero Lebesgue measure, then $d_F = \dim_M$, the Minkowski dimension. Thus, we have that $\gamma_{n-1} \leq \dim_M$ when $\mu(X) = 0$. As an example where equality holds, we saw in Chapter 3 that $\gamma_1 = \dim_M = \log 3 / \log 2$ for the Sierpinski triangle.

The above proof does not apply to γ_k with $k < n - 1$ because the assumption that the n-balls in the complement are disjoint is not valid.

5.4 Conjectures

In this section, we briefly discuss some relationships that we conjecture to hold, based on the examples in this thesis. The first problem concerns the discreteness index of zero-measure Cantor sets. The second conjecture is that self-similar fractals should have topological growth rates equivalent to their similarity dimension. We finish with some questions about additional inequalities involving the fat fractal exponent and the γ_i.

Conjecture 1. *Cantor sets with zero Lebesgue measure have $\delta = 1$.*

This holds for all the zero measure Cantor set examples that we have studied in this thesis and we believe it to hold in generally. In Section 5.3.2, we showed that for any totally disconnected set with $\gamma \neq 0$, $\delta \leq 1$. Therefore, all that remains is to show $\delta \geq 1$ under suitable assumptions on the set X. Since we have seen examples of a fat Cantor set and a totally disconnected non-perfect set with $\delta < 1$, the assumptions on X must include that is has zero measure and is perfect. It may also be the case that $\delta = 1$ only holds for a more restricted class of

sets — for example, self-similar Cantor sets. We have attempted to prove the conjecture under this condition, but have so far been unsuccessful. The index δ is defined in terms of the largest ϵ-component diameter. It is possible that a different measure of component size is needed — perhaps the smallest ϵ-component diameter, since this is related to the property of perfectness.

Conjecture 2. *If X is a self-similar fractal and $\gamma_i \neq 0$, then $\gamma_i = \dim_S$.*

This has been the case for the examples of Chapters 2 and 3. It is a reasonable conjecture because self-similarity is such a strong property that we expect it to dominate any scaling law. A proof of this conjecture might use related constructions to those used in proving that the Hausdorff and similarity dimensions are equivalent for self-similar sets that satisfy the open set property; see [23], for example. It seems that the easiest place to start is with self-similar Cantor sets that satisfy a "closed set condition." That is, $X = \bigcup f_i(X)$ with this union disjoint. Not all self-similar Cantor sets have this property — for example, some Cantor set relatives of the Sierpinski triangle do not.

As mentioned in Section 5.3.4, it may be possible to derive further inequalities involving the topological growth rates and the Minkowski dimension or fat fractal exponent. For example, if $X \subset \mathbb{R}^n$, can we show that $\gamma_i \leq d_F$ for $0 \leq i \leq n-2$? The results of this chapter have used Tricot's formulation of fat fractal scaling [80]. We may be able to obtain different results by comparing our indices with the fat fractal exponents of Umberger *et al.* [22].

This is only a partial list; there are many promising avenues to explore.

Chapter 6

Conclusions and Future Work

This thesis has considered the problem of extracting topological information about a set from a finite approximation to it. The essence of our approach is to coarse-grain the data at a sequence of resolutions and extrapolate the limiting trend. Our theoretical work and numerical investigations show that this multiresolution approach can successfully recover information about the underlying topology when the data approximate a compact subset of a metric space. The extrapolation is always constrained by the finite nature of the finite-precision data; we identify a cutoff resolution to measure this. Although the examples studied in this thesis are fairly simple, the theory applies in very general contexts. With faster numerical implementations, we believe that our approach to computational topology could be a useful tool for analyzing data from both physical and numerical experiments.

In the following sections we summarize the main results of this thesis then outline directions for further research.

6.1 Summary of results

The main contribution of this thesis is the multiresolution approach to computational topology developed in Chapters 2 and 3. This approach has a number of advantages over existing single resolution techniques. First, it is applicable to both smooth and fractal sets, the only condition is that they be compact subsets of a metric space. Second, by examining data at a sequence of resolutions we obtain more accurate knowledge of the underlying topology by identifying persistent features. Finally, it leads to a practical method for estimating the cutoff resolution — a measure of confidence in the results. At present, the major drawback to computing topological information — especially homology — at many resolutions is the high time-cost of the computations.

In Chapter 2, we considered the problem of distinguishing between connected and disconnected sets. The key step was introducing the functions $C(\epsilon)$, $D(\epsilon)$, and $I(\epsilon)$ to count the number of ϵ-connected components, the largest ϵ-component diameter, and the number of ϵ-isolated points respectively. Results from Section 2.2 show that the behavior of these functions as the resolution parameter ϵ tends to zero tells us whether or not a compact space is connected, totally disconnected, and/or perfect. For arbitrary point-set data, $C(\epsilon)$, $D(\epsilon)$, and $I(\epsilon)$ are easily computed from the minimal spanning tree. Another consequence of these ideas is a technique for estimating the inherent accuracy of the data. When the data approximate a perfect set, we estimate the cutoff resolution as the smallest ϵ-value for which there are no ϵ-isolated points. Our characterization of connected components as a function of resolution has many potential

applications, some of which are discussed in the following section.

The topic of Chapter 3 was computational homology - in particular using the Betti numbers to count the number and type of non-bounding cycles in a space. Since the zeroth-order Betti number is the number of path-connected components of a space, this forms a natural extension of the work in Chapter 2. The central lesson from this chapter is that it is not enough to examine the Betti numbers as a function of resolution. This is because coarse-graining a set can introduce spurious holes that are cause by the geometry, rather than the topology, of the space. Instead, an inverse system of ϵ-neighborhoods is necessary. The inclusion maps from the inverse system identify holes that persist in the limit as ϵ tends to zero. We quantify this by the persistent Betti number $\beta_k^\lambda(\epsilon)$ which counts the number of holes in the ϵ-neighborhood that have a preimage in a smaller λ-neighborhood. This enables us to detect those holes that are due to the coarse-graining rather than the underlying topological structure. The system of ϵ-neighborhoods also allows us to formalize the relationship between the data and the underlying space. In particular, we derive inequalities involving the persistent Betti numbers of the data and the underlying space. We anticipate that both the persistent and regular Betti numbers of an ϵ-neighborhood will be useful in characterizing the structure of data. The persistent Betti numbers reflect the underlying topological structure while the regular Betti numbers of ϵ-neighborhoods give additional information about how the space is embedded. As we discussed in Chapter 3, more efficient numerical implementations are needed before these techniques can be fully applied to real data.

In Chapter 4, we applied the techniques from Chapter 2 to study some simple examples from dynamical systems. These examples each have well understood structure, so they provide a test of our techniques and illustrate the versatility of our approach. In the first example, we confirmed the Cantor-set structure of cross-sections from the Hénon attractor. We then studied the breakup of invariant circles in an area-preserving twist map. The transition from circle to Cantor set is continuous in a metric sense, so the functions $C(\epsilon)$ and $D(\epsilon)$ are not very sensitive to this transition. However, by adapting our techniques to examine the scaling of the "largest gap," we develop a new criterion for finding the critical parameter value that compares well with previous results.

Many of the examples in this thesis are fractals. By definition, a fractal has structure on arbitrarily fine scales, so it is possible for $C(\epsilon)$ or $\beta_k^0(\epsilon)$ to go to infinity as ϵ goes to zero. In Chapter 5, we derive inequalities that relate the topological growth rates to various existing measures of fractal scaling. We find that the growth rates of the number of components or holes are closely related to the Minkowski dimension and fat fractal exponents via the Besicovitch-Taylor index. Our exponents, however, distinguish between fractals that have the same dimension but different topological structure. They are therefore a useful addition to the collection of tools for characterizing fractal structure.

6.2 Directions for future work

A number of open problems were discussed in the body of the thesis as they arose. These ranged from easy extensions of the work presented in this thesis, to potential applications, to general questions about whether we can use similar techniques to compute other topological properties, such as branching structure or local connectedness, from finite data. The most interesting problems from each chapter are revisited below.

We start by describing extensions of our work in Chapter 2 on connected components and minimal spanning trees. The first two items are simple generalizations that may be of interest in

applications. The last question concerns the distribution of edge lengths in the minimal spanning tree.

1. We could use different measures of the size of an ϵ-component. Examples include the relative number of points in an ϵ-component or the n-dimensional volume of space occupied by a component. Such measures are often used in applications of percolation theory. Recall that we only examined scaling in the largest ϵ-component diameter, since this was our test for total disconnectedness. It is likely that the entire distribution of ϵ-component sizes will give interesting information in applications. This would require only a slight modification of our algorithms.

2. We observed in Chapter 2 that the cutoff resolution for nonuniformly distributed data is larger than that for a uniform covering of the underlying set. A large cutoff resolution leads to low confidence in the extrapolated underlying topology. It may be possible to reduce the cutoff resolution for nonuniform data by weighting the MST edges by the nearest neighbor distance for each point. This idea is appealing heuristically but needs some formal justification.

3. The function $C(\epsilon)$ is essentially the cumulative distribution of edge lengths in the minimal spanning tree. For finite data this distribution has two parts. When $\epsilon > \rho$ the distribution carries information about the topology of the underlying set — the focus of this thesis. We conjecture that for $\epsilon < \rho$ the distribution of MST edge-lengths is related to the distribution of the data points, i.e., a measure associated with the underlying set. It is possible that formal results about this already exist in statistics. In [78] there is a result that relates the total length of a MST to the underlying point distribution. For subsets of $\mathbb{R}$, the relationship between distributions of points and corresponding MST edge lengths should reduce to a problem in order statistics [10].

Our work on computational homology in Chapter 3 focussed on the mathematical foundations rather than the implementations, and there is a significant amount of work to be done on the latter.

1. The alpha shape algorithm we described in Section 3.4.1 is a subcomplex approach to generating simplicial complexes at multiple resolutions. We argued in Section 3.4.3 that a more efficient approach is to use subdivisions of cubical complexes. This requires a slight adjustment of the theory and a substantial amount of work on the implementations.

2. We derived a formula for computing the persistent Betti numbers in Section 3.3.5. This is certainly not the only way to compute them. Algorithms for computing regular Betti numbers have exploited many different results from algebraic topology. It may be possible to adapt some of these to our problem. Efficient implementations will also be highly dependent on the type of cell complexes used.

3. In terms of theory, we need a more complete understanding of the continuity of the persistent Betti numbers, $\beta_k^\lambda(\epsilon)$, as λ and ϵ tend to zero. This is related to continuity and tautness results for Čech homology.

We gave some detailed descriptions of potential applications in dynamical systems at the end of Chapter 4. The most challenging of these is the break-up of invariant tori in four-dimensional symplectic twist maps. In general, we anticipate that our computational techniques will be particularly useful in such higher-dimensional settings where visualization is difficult.

More theoretical questions that are related to the study of dynamical systems include the following.

1. Just as the dimension can vary at different points of a multifractal, the scaling of components or holes with resolution may differ for subsets of a fractal. Is it possible to localize our theory to quantify this?

2. Newhouse defined the *thickness* of Cantor subsets of $\mathbb{R}$ to analyze the existence of homoclinic tangencies of stable and unstable manifolds [62]. The definition is given in terms of ratios of diameters and deleted intervals. It may be possible to generalize this notion to Cantor subsets of $\mathbb{R}^n$ using techniques from Chapter 2.

As we emphasized in Chapter 5 there is ample room for many more results relating our topological growth rates to fractal dimensions. See Section 5.4 for details.

From the number of open problems in this short list, it should be clear that computational topology is a rich, interesting, and rapidly evolving discipline. The work in this thesis suggests that further research in this field is likely to be fruitful.

Bibliography

[1] Software available from NCSA via anonymous ftp from ftp.ncsa.uiuc.edu/Visualization/Alpha-shape/.

[2] C. Allain and M. Cloitre. Characterizing the lacunarity of random and deterministic fractal sets. *Physical Review A*, 44:3552–3558, 1991.

[3] D.K. Arrowsmith and C.M. Place. *An introduction to dynamical systems.* Cambridge University Press, 1990.

[4] M.F. Barnsley. *Fractals Everywhere.* Academic Press, Boston, second edition, 1993.

[5] R. Blumenfeld and B.B. Mandelbrot. Levy dests, Mittag-Leffler statistics, mass fractal lacunarity, and perceived dimension. *Physical Review E*, 56:112–118, 1997.

[6] M. Brady, J.M. Hollerbach, T.L. Johnson, T. Lozano-Perez, and M. T. Mason, editors. *Robot Motion: Planning and Control*. MIT Press, 1984.

[7] B. Burch and J.C. Hart. Linear fractal shape interpolation. *Proceedings / Graphics Interface '97*, pages 155–162, 1997.

[8] B.V. Chirikov. A universal instability of many-dimensional oscillator systems. *Physics Reports*, 52(5):263–379, 1979.

[9] T.H. Cormen, C.E. Leiserson, and R.L. Rivest. *Introduction to Algorithms.* MIT Press, Cambridge, MA, 1990.

[10] H.A. David. *Order Statistics.* John Wiley & Sons, 1970.

[11] C.J.A. Delfinado and H. Edelsbrunner. An incremental algorithm for Betti numbers of simplicial complexes on the 3-sphere. *Computer Aided Geometric Design*, 12:771–784, 1995.

[12] M. Dellnitz and O. Junge. A subdivision algorithm for the computation of unstable manifolds and global attractors. *Numerische Mathematik*, 75:293–317, 1997.

[13] T.K. Dey, H. Edelsbrunner, and S. Guha. Computational topology. In B. Chazelle, J.E. Goodman, and R. Pollack, editors, *Advances in Discrete and Computational Geometry*, volume 223 of *Contemporary Mathematics.* American Mathematical Society, 1999.

[14] T.K. Dey and S. Guha. Computing homology groups of simplicial complexes in $\mathbf{R}^3$. *Journal of the ACM*, 45:266–287, 1998.

[15] R.W. Easton. *Geometric Methods for Discrete Dynamical Systems*. Oxford University Press, 1988.

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

[17] H. Edelsbrunner, D.G. Kirkpatrick, and R. Seidel. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, 29(4):551–559, 1983.

[18] H. Edelsbrunner and E.P. Mücke. Three-dimensional alpha shapes. *ACM Transactions on Graphics*, 13:43–72, 1994.

[19] G.A. Edgar. *Measure, Topology, and Fractal Geometry*. Springer, 1990.

[20] R. Eykholt and D.K. Umberger. Characterization of fat fractals in nonlinear dynamical systems. *Physical Review Letters*, 57:2333–2336, 1986.

[21] R. Eykholt and D.K. Umberger. Relating the various scaling exponents used to characterize fat fractals in nonlinear dynamical systems. *Physica D*, 30:43–60, 1988.

[22] R. Eykholt and D.K. Umberger. Extension of the fat-fractal exponent β to arbitrary sets in d dimensions. *Physics Letters A*, 163:409–414, 1992.

[23] K.J. Falconer. *Fractal Geometry: Mathematical Foundations and Applications.* Wiley, New York, 1990.

[24] K.J. Falconer. *Techniques in Fractal Geometry*. John Wiley & Sons, New York, 1997.

[25] A. Fathi. Expansiveness, hyperbolicity and Hausdorff dimension. *Communications in Mathematical Physics*, 126:249–262, 1989.

[26] J. Friedman. Computing Betti numbers via combinatorial Laplacians. *Algorithmica*, 21:331–346, 1998.

[27] C. Froeschlé. Numerical study of a four-dimensional mapping. *Astronomy and Astrophysics*, 16:172–189, 1972.

[28] R.W. Ghrist, P.J. Holmes, and M.C. Sullivan. *Knots and Links in Three-Dimensional Flows*. Number 1654 in Lecture Notes in Mathematics. Springer, 1997.

[29] G.H. Golub and C.F. Van Loan. *Matrix Computations.* Johns Hopkins, third edition, 1996.

[30] C. Grebogi, S.W. McDonald, E. Ott, and J.A. Yorke. Exterior dimension of fat fractals. *Physics Letters A*, 110:1–4, 1985.

[31] J.M. Greene. A method for determining a stochastic transition. *Journal of Mathematical Physics*, 20:1183–1201, 1979.

[32] A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2000. Also available at http://www.math.cornell.edu/ hatcher/.

[33] G.T. Herman and E. Zhao. Jordan surfaces in simply connected digital spaces. *Journal of Mathematical Imaging and Vision*, 6:121–138, 1996.

[34] I.N. Herstein. *Topics in Algebra*. Blaisdell Publishing Company, 1964.

[35] J.G. Hocking and G.S. Young. *Topology*. Addison-Wesley, 1961.

[36] J.E. Hutchinson. Fractals and self similarity. *Indiana University Mathematics Journal*, 30:713–747, 1981.

[37] W.D. Kalies, K. Mischaikow, and G. Watson. Cubical approximation and computation of homology. *Banach Center Publications*, 47:115–131, 1999.

[38] S. Kim and S. Ostlund. Simultaneous rational approximations in the study of dynamical systems. *Physical Review A*, 34:3426–3434, 1986.

[39] T. Kong, A.W. Roscoe, and A. Rosenfeld. Concepts of digital topology. *Topology and its Applications*, 46:219–262, 1992.

[40] T. Y. Kong and A. Rosenfeld. Digital topology: Introduction and survey. *Computer Vision, Graphics and Image Processing*, 48:357–393, 1989.

[41] H. Kook and J.D. Meiss. Period orbits for reversible, symplectic mappings. *Physica D*, 35:65–86, 1989.

[42] K. Kuratowski. *Topology*, volume II. Academic Press, 1968.

[43] J. Laskar. Frequency analysis for multi-dimensional systems. global dynamics and diffusion. *Physica D*, 67:257–281, 1993.

[44] J. Laskar, C. Froeschlé, and A. Celletti. The measure of chaos by the numerical analysis of the fundamental frequencies. application to the standard mapping. *Physica D*, 56:253–269, 1992.

[45] C.-N. Lee, T. Poston, and A. Rosenfeld. Winding and Euler numbers for 2D and 3D digital images. *CVGIP: Graphical Models and Image Processing*, 53:522–537, 1991.

[46] R.S. MacKay. Greene's residue criterion. *Nonlinearity*, 5:161–187, 1992.

[47] R.S. MacKay. *Renormalisation in Area-Preserving Maps*, volume 6 of *Advanced Series in Nonlinear Dynamics*. World Scientific, 1993.

[48] R.S. MacKay and J.D. Meiss. Cantori for symplectic maps near the anti-integrable limit. *Nonlinearity*, 5:149–160, 1992.

[49] R.S. MacKay, J.D. Meiss, and J. Stark. Converse KAM theory for symplectic twist maps. *Nonlinearity*, 2:555–570, 1989.

[50] R.S. MacKay and I.C. Percival. Converse KAM: Theory and practice. *Communications in Mathematical Physics*, 98:469–512, 1985.

[51] B.B. Mandelbrot. Measures of fractal lacunarity: Minkowski content and alternatives. In C. Bandt, S. Graf, and M. Zähle, editors, *Fractal geometry and stochastics*, volume 37 of *Progress in Probability*, pages 15–42. Birkhäuser Verlag, 1995.

[52] S. Mardešić and J. Segal. *Shape Theory*. North-Holland, 1982.

[53] J. Mather. A criterion for the non-existence of invariant circles. *Publications Mathématiques, I.H.E.S.*, 63:153–204, 1986.

[54] A.I. Mees. Tesselations and dynamical systems. In M. Casdagli and S. Eubank, editors, *Nonlinear Modeling and Forecasting*, volume 12 of *Proceedings volumes in the SFI Studies in the Sciences of Complexity*, pages 3–24. Addison-Wesley, 1992.

[55] J.D. Meiss. Symplectic maps, variational principles, and transport. *Reviews of Modern Physics*, 64:795–847, 1992.

[56] J.D. Meiss. Transient measures in the standard map. *Physica D*, 74:254–267, 1994.

[57] G. Mindlin and R. Gilmore. Topological analysis and synthesis of chaotic time series. *Physica D*, 58:229–242, 1992.

[58] K. Mischaikow, M. Mrozek, J. Reiss, and A. Szymczak. Construction of symbolic dynamics from experimental time series. *Physical Review Letters*, 82(6):1144–1147, 1999.

[59] K. Mischaikow, M. Mrozek, A. Szymczak, and J. Reiss. From time series to symbolic dynamics: An algebraic topological approach. Preprint, http://www.math.gatech.edu/ mischaik/papers/paperlist.html, December 1997.

[60] M.R. Muldoon, R.S. MacKay, J.P. Huke, and D.S. Broomhead. Topology from a time series. *Physica D*, 65:1–16, 1993.

[61] J.R. Munkres. *Elements of Algebraic Topology*. Benjamin Cummings, 1984.

[62] S. Newhouse. Diffeomorphisms with infinitely many sinks. *Topology*, 13:9–18, 1974.

[63] H.E. Nusse and J.A. Yorke. *Dynamics: Numerical Explorations.* Springer, 2nd edition, 1998.

[64] E. Ott, T. Sauer, and J.A. Yorke. *Coping with Chaos.* Wiley, 1994.

[65] L.M. Pecora, T.L. Carroll, and J.F. Heagy. Statistics for mathematical properties of maps between time series embeddings. *Physical Review E*, 52:3420–3433, 1995.

[66] H.-O. Peitgen, S. Jürgens, and D. Saupe. *Chaos and Fractals: New Frontiers of Science.* Springer-Verlag, 1992.

[67] F.P. Preparata and M.I. Shamos. *Computational Geometry: An Introduction.* Springer-Verlag, 1985.

[68] R.S. Prim. Shortest connection networks and some generalizations. *The Bell System Technical Journal*, pages 1389–1401, November 1957.

[69] V. Robins. Towards computing homology from finite approximations. Submitted to *Topology Proceedings*, October 1999.

[70] V. Robins. Periodic orbits and invariant tori in symplectic maps. Honours thesis, Department of Mathematics, Australian National University, 1994.

[71] V. Robins, J.D. Meiss, and E. Bradley. Computing connectedness: An exercise in computational topology. *Nonlinearity*, 11:913–922, 1998.

[72] V. Robins, J.D. Meiss, and E. Bradley. Computing connectedness: Disconnectedness and discreteness. *Physica D*, 139:276–300, 2000.

[73] C. Robinson. *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos.* CRC Press, Boca Raton, 1995.

[74] A. Rosenfeld. Connectivity in digital pictures. *Journal of the ACM*, 17:146–160, 1970.

[75] S.J. Shenker and L.P. Kadanoff. Critical behavior of a KAM surface: I. Empirical results. *Journal of Statistical Physics*, 27:631–656, 1982.

[76] F.J. Solis and L. Tao. Lacunarity of random fractals. *Physics Letters A*, 228:351–356, 1997.

[77] E.H. Spanier. *Algebraic Topology*. McGraw-Hill, New York, 1966.

[78] J.M. Steele. Growth rates of Euclidean minimal spanning trees with power weighted edges. *The Annals of Probability*, 16:1767–1787, 1988.

[79] G.J. Sussman and M. Wisdom, J. with Mayer. Structure and interpretation of classical mechanics. Technical Report 1667, MIT Artifical Intelligence Laboratory, 1999.

[80] C. Tricot. The geometry of the complement of a fractal set. *Physics Letters A*, 114:430–434, 1986.

[81] C. Tricot. *Curves and Fractal Dimension.* Springer-Verlag, 1995.

[82] N.B. Tufillaro, P. Wyckoff, R. Brown, T. Schreiber, and T. Molteno. Topological time-series analysis of a string experiment and its synchronized model. *Physical Review E*, 51(1):164–174, 1995.

[83] J.K. Udupa. Surface connectedness in digital spaces. In A. Rosenfeld and T.Y. Kong, editors, *Topological Algorithms for Digital Image Processing.* Elsevier, Amsterdam, 1993.

[84] D.K. Umberger and J.D. Farmer. Fat fractals on the energy surface. *Physical Review Letters*, 55:661–664, 1985.

[85] E.W. Weisstein. *CRC Concise Encyclopedia of Mathematics.* CRC Press, 1998.

[86] R. van de Weygaert, B.J.T. Jones, and V.J. Martinez. The minimal spanning tree as an estimator for generalized dimensions. *Physics Letters*, 169A:145–150, 1992.

[87] G.T. Whyburn. Topological characterization of the Sierpinksi curve. *Fundamenta Mathematicae*, 45:320, 1958.

[88] J. Williams, editor. *Geographic Information from Space: processing and applications of geocoded satellite images*. Wiley, 1995.

[89] K. Yip. *KAM : A system for intelligently guiding numerical experimentation by computer*. MIT Press, 1991.

[90] C.T. Zahn. Graph-theoretical methods for detecting and describing Gestalt clusters. *IEEE Transactions on Computers*, C-20:68–86, 1971.

List of Notation

X	Underlying compact space
X_ϵ	ϵ-neighborhood of X
S	Data set
d_H	Hausdorff distance/metric
$\mu(X)$	Lebesgue measure in $\mathbb{R}^n$
ϵ	Resolution parameter
ρ	Cutoff resolution
$C(\epsilon)$	Number of ϵ-connected components
$D(\epsilon)$	Largest ϵ-component diameter
$I(\epsilon)$	Number of isolated points
γ	Disconnectedness index, growth rate of $C(\epsilon)$
δ	Discreteness index, growth rate of $D(\epsilon)$
σ^k	k-simplex
$\lvert\mathcal{C}\rvert$	Underlying space of a simplicial complex
∂	Boundary operator
C_k	Chain group
B_k	Boundary group
Z_k	Cycle group
H_k	Homology group
β_k	Betti number
$\beta_k^\lambda(\epsilon)$	Persistent Betti number
γ_k	Growth rate of $\beta_k^0(\epsilon)$
$\varprojlim$	Inverse limit
ω	Rotation number
μ	Golden mean
k	Standard map parameter
k_c	Critical value of k for the golden mean circle
$N(\epsilon)$	Number of boxes in the definition of $\dim_B$
$\dim_B$	Box-counting dimension
$\dim_H$	Hausdorff dimension
$\dim_M$	Minkowski dimension
$\dim_S$	Similarity dimension
d_{BT}	Besicovitch-Taylor index
d_F	Fat fractal exponent