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

1 Introduction **1**
1.1 Extracting qualitative information from data 1
1.2 Fractal geometry 4
1.2.1 Fractal dimensions 5
1.2.2 Other characterizations of fractal structure 6
1.2.3 Iterated function systems 10
1.3 Computational topology 11
1.4 Overview of the thesis 12
1.4.1 The basic assumptions 12
1.4.2 Organization of the thesis 13

2 Computing Connectedness **15**
2.1 Introduction 15
2.2 Foundations for computing connectedness 16
2.2.1 Concepts from point-set topology 16
2.2.2 ϵ-Resolution definitions 16
2.2.3 Disconnectedness and discreteness growth rates 18
2.3 Implementation 20
2.3.1 Minimal spanning trees 21
2.3.2 Practical issues 22
2.4 Examples 24
2.4.1 Relatives of the Sierpinski triangle 24
2.4.2 Cantor sets in the plane 32
2.5 Concluding remarks 38

3 Computing Homology **39**
3.1 Introduction 39
3.2 An overview of homology theory 40
3.2.1 Simplicial homology 41
3.2.2 The role of homotopy in homology 45
3.2.3 Inverse systems 47
3.2.4 Čech homology 48
3.2.5 Shape theory 53
3.3 Foundations for computing homology 54
3.3.1 The inverse system of ϵ-neighborhoods 54
3.3.2 Persistent Betti numbers 55
3.3.3 Growth rates for persistent Betti numbers 57

3.3.4 Finite approximations . . . 58
3.3.5 Computing persistent Betti numbers . . . 59
3.4 Implementation . . . 60
3.4.1 Alpha shapes . . . 61
3.4.2 Other algorithms for computational homology . . . 65
3.4.3 A better way? . . . 65
3.5 Examples . . . 67
3.5.1 Sierpinski triangle relatives revisited . . . 67
3.6 Concluding remarks . . . 74

4 Applications in Dynamical Systems **75**
4.1 Introduction . . . 75
4.2 The Hénon attractor . . . 76
4.3 Circle breakup in the standard map . . . 80
4.3.1 Background . . . 81
4.3.2 MST analysis of the transition from circle to Cantor set . . . 85
4.3.3 Scaling of the gaps . . . 86
4.4 Cantori from sawtooth maps . . . 94
4.5 Further applications . . . 97
4.5.1 Torus breakup in the Froeschlé map . . . 97
4.5.2 The chaotic fat fractal . . . 98
4.5.3 Pruning outlying data points . . . 99
4.6 Concluding remarks . . . 99

5 Topological Growth Rates and Fractal Dimensions **101**
5.1 Introduction . . . 101
5.2 Definitions . . . 102
5.2.1 Box-counting dimension . . . 102
5.2.2 Fat fractal exponents . . . 103
5.2.3 The Besicovitch-Taylor index . . . 103
5.2.4 Topological growth rates . . . 105
5.3 Results . . . 105
5.3.1 Subsets of the line . . . 106
5.3.2 Disconnected subsets of $\mathbb{R}^n$ . . . 109
5.3.3 Examples . . . 110
5.3.4 Other subsets of $\mathbb{R}^n$ . . . 113
5.4 Conjectures . . . 114

6 Conclusions and Future Work **117**
6.1 Summary of results . . . 117
6.2 Directions for future work . . . 118

Bibliography **121**

Code **127**

List of Notation **133**

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].

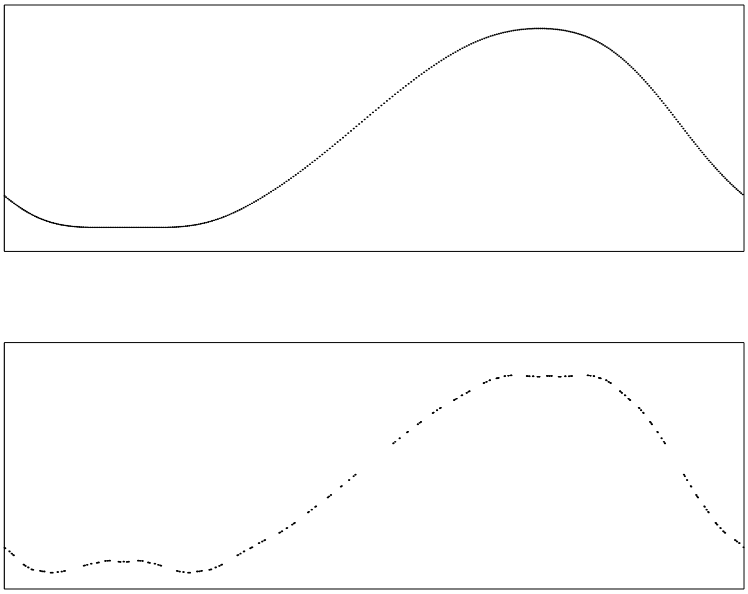

Figure 1.1: Approximations to a quasiperiodic orbit in the standard map for different values of its parameter, k. The top orbit ($k = 0.7$) approximates an invariant circle; the bottom orbit ($k = 1.0$) covers a Cantor set. The transition from circle to Cantor set is a dramatic change in the topology 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-

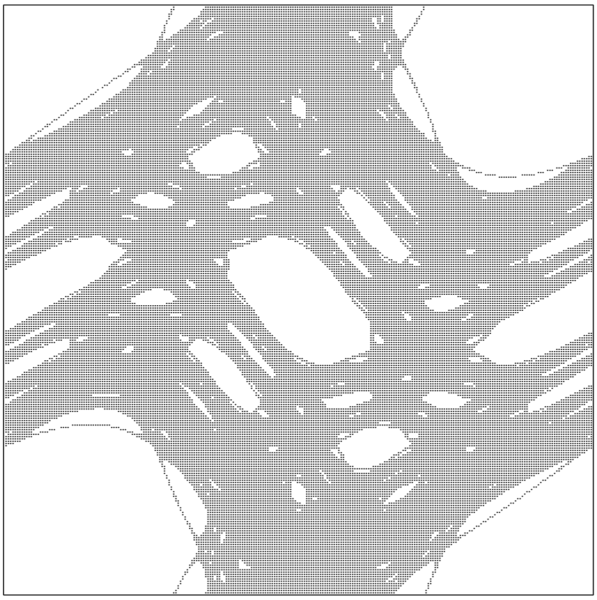

Figure 1.2: A chaotic region of the standard map with $k = 1.1$. The picture is generated by recording which boxes in a 256×256 grid are visited by a single chaotic orbit of 10^8 points. The chaotic region is connected and appears to have positive area which tells us that a significant proportion of 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.

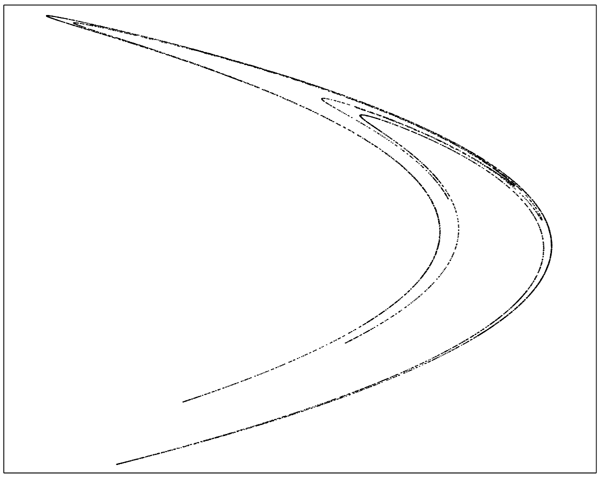

Figure 1.3: The Hénon attractor, one of the standard examples in chaotic dynamics. The set is connected and it is often described as having a Cantor set cross-section. We confirm this structure using our computational techniques in Chapter 4. The attractor has a box-counting dimension of approximately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

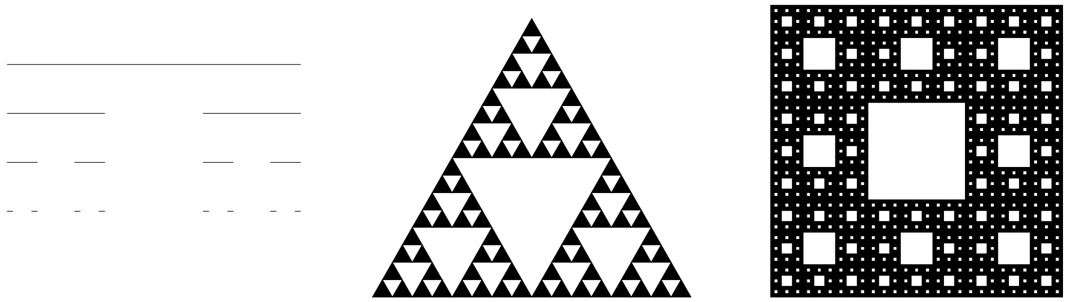

Figure 1.4: The Cantor set, Sierpinski triangle, and Sierpinski curve — standard fractal examples.

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

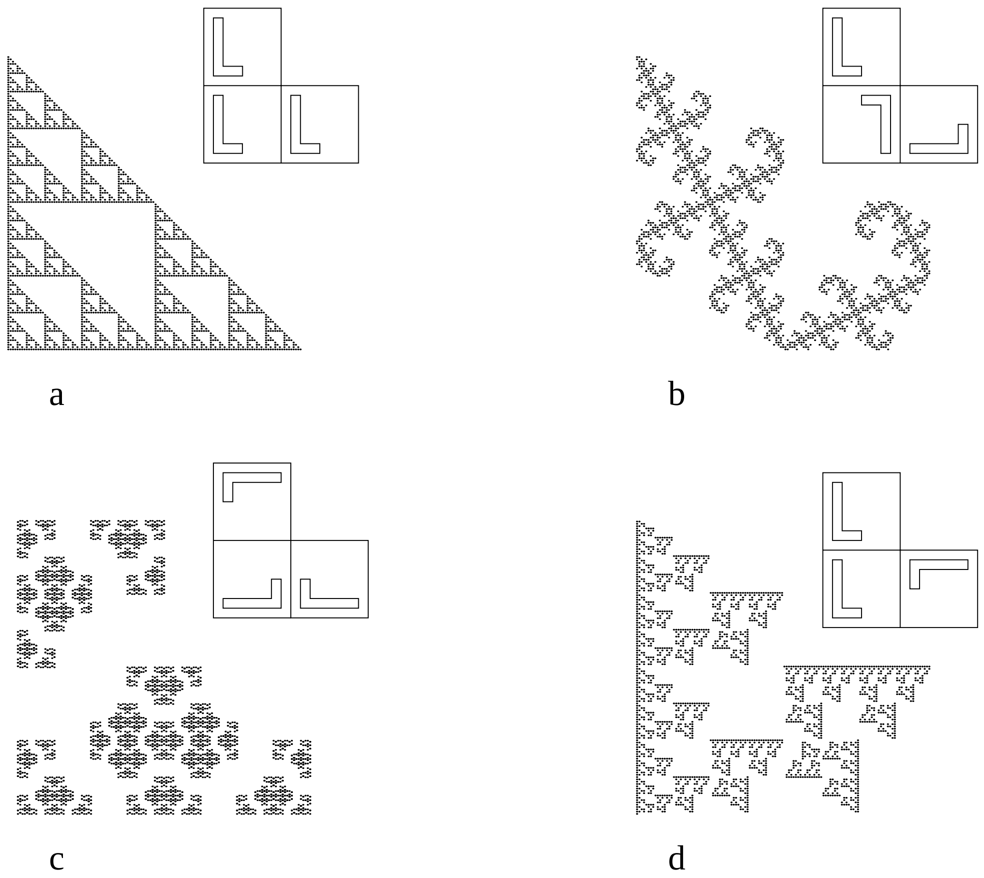

Figure 1.5: Four relatives of the Sierpinski triangle. Each of these fractals is generated as the attractor for an iterated function system (Section 1.2.3) that maps the unit square into three squares of one half the size. The “L” notation designates the rotation or reflection used in each case. These 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.

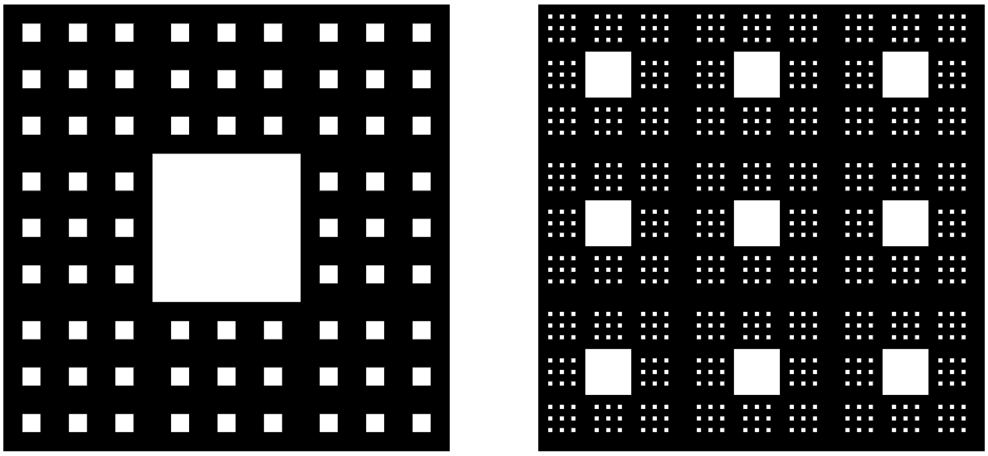

Figure 1.6: Two fractals with the same dimension but different lacunarity (after Mandelbrot [51]). Each fractal is built from $9^2 - 9$ copies of the unit square with an edge of $\frac{1}{9}$. They differ in the positioning of the 9 deleted squares, and this gives them different coarse-scale structure. 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.

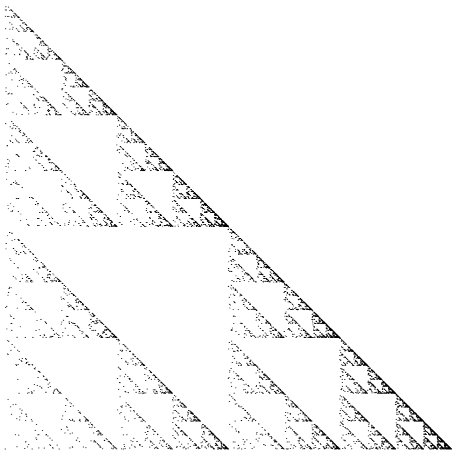

Figure 1.7: A multifractal measure on the Sierpinski triangle. This picture is generated by choosing each of the three functions from the IFS with uneven probabilities. Such a nonuniform distribution of points leads to inaccurate estimations of the box-counting dimension. In our work, the effect is an 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 $r \to 0$:

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

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 areapreserving 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.