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

$$\mu(X) = b - a - \sum_{1}^{\infty} u_i.$$

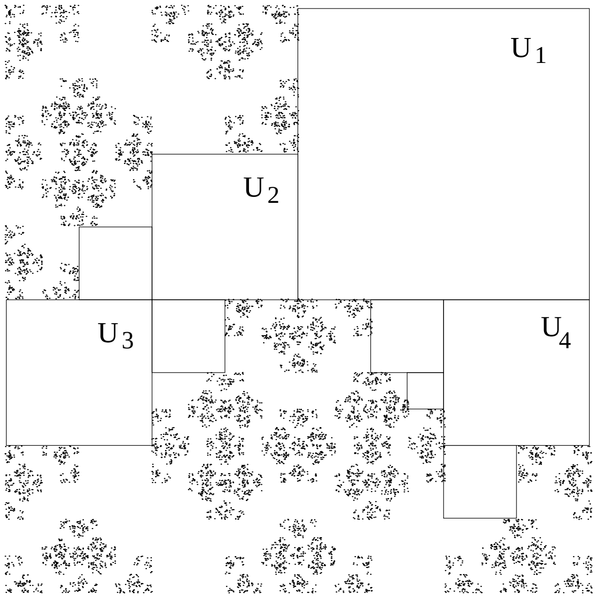

Figure 5.1: Packing the complement of a fractal.

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 $\gamma^{\text{inf}}, \gamma^{\text{sup}}, \delta^{\text{inf}}$ or δ^{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 $n \geq N$, $n u_n^{\beta} < 1$, 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 $g_k < 2\epsilon \le g_{k-1}$. 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

$$(\tfrac{3}{10})(\tfrac{1}{2})^k \le \mu(K_\epsilon) - \mu(K) \le (\tfrac{3}{10})(\tfrac{1}{2})^{k-1}.$$

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

$$X = \{0, 1, \tfrac{1}{2}, \tfrac{1}{3}, \ldots\}.$$

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 $g_{n+1} \le \epsilon_n < g_n$. Then the points $1, \ldots, 1/n$ 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 $\delta \ne 1$. 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, $r_1 \geq r_2 \geq r_3 \geq \dots$ 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.