DISSERTATION

IMPROVED METHODS FOR CALCULATING THE MULTIFRACTAL SPECTRUM FOR SMALL

DATA SETS

Submitted by

Leif Anderson

Department of Physics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2014

Doctoral Committee:

    Advisor: Richard Eykholt

    Raymond Steve Robinson
    Martin Gelfand
    Mingzhong Wu
    Patrick Shipman

Abstract

Improved Methods for Calculating the Multifractal Spectrum for Small Data Sets

There are multiple definitions for fractal dimension, and those definitions disagree. The multifractal spectrum provides a unifying framework for a broad family of definitions of dimension, but it requires large amounts of data to compute. We provide a description of the multifractal spectrum, one existing improvement that improves convergence for small data sets: the Extended Box Algorithm (EBA), and develop several further improvements: Local and Global Norm modifications to the EBA, the a-Norm, the Variable Box size Algorithm (VBA), and the Patchwork method.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1. MOTIVATION

One of the most interesting aspects of chaos is the universality of many of the results. One of the most famous examples is Feigenbaum's proof of the quantitative universality of his scaling factors [1] for all quadratic maps. This result proves that measurements of α and δ from the bifurcation diagrams of vastly different physical systems should produce the same universal values. It is closely related to the results of Metropolis, Stein, and Stein [2], extended by Derrida, Gervois, and Pomeau [3], which prove the existence of more self-similarity rules in the bifurcation sequences of an even broader class of systems. The beauty of these results is that despite the very different manifestations, one can show that in a way, most chaotic systems are just expressions of a more unified type of behavior, and still bear some of the quantifiable characteristics of that behavior.

In a similar vein, it can be shown that the fractal dimension of a chaotic attractor is invariant under smooth coordinate changes [4]. Thus, one can evaluate a quantity based on measurements corresponding to a particular set of initial conditions, with some very specific embedding method, and still reach a result that characterizes the underlying dynamics in a way that is completely universal. Furthermore, since chaos itself is so universal, fractal analysis can be applied to data from incredibly diverse sources, from the human heartbeat [5] to geological fracture systems [6].

One of the potential problems with this idea is that there are multiple definitions of dimension, and these definitions are known to differ from one another. This can be resolved with the introduction of the multifractal spectrum [7, 8], which provides a unifying

framework for a very broad class of fractal dimensions. However, accurate calculations of the multifractal spectrum usually require prohibitive amounts of data, especially for high-dimensional attractors. Our aim is to improve calculations of the multifractal spectrum for small data sets.

1.2. The Multifractal Spectrum

1.2.1. Introduction of D_q. There are multiple definitions of dimension, and they are known to disagree in general. In this section, we'll show how these definitions can be unified.

The simplest example is the box counting dimension D_0. Cover the space with a uniform grid of boxes[1] with side ℓ. The box counting dimension is then defined by:

$$D_0 = \lim_{\ell \to 0} \frac{\ln N(\ell)}{|\ln \ell|}$$

where $N(\ell)$ is the number of boxes that include at least some part of the attractor (not the total number of points). The box counting dimension is usually thought of as the most straightforward definition of dimension.

Another definition, which turns out to converge better than D_0 for smaller data sets, is the correlation dimension D_c [9] which can be written as

$$D_c = \lim_{\ell \to 0} \frac{\ln \left(\frac{1}{N^2} \sum_{i,j} \theta(\ell - |x_i - x_j|) \right)}{\ln \ell}$$

where x_i is the location of the ith point in the data set, and N is the total number of points in the set. The θ function is the standard Heaviside θ function, which is one if its argument is positive, and zero if its argument is negative. The θ term counts whether points i and

---

[1]Nonuniform covers will be discussed in a later section.

2

j are within a distance ℓ of each other. The numerator is basically the number of pairs of points within a given distance. There is an implied limit as $N \to \infty$ before the $\ell \to 0$ limit. The correlation dimension is generally accepted as the dimension that converges the fastest of the standard choices.

Perhaps the most philosophically useful dimension is the information dimension [10] which has been shown to characterize the rate of information loss through time, or the rate of information gain by sequential measurements. That dimension D_I is defined as:

$$D_I = \lim_{\ell \to 0} \frac{\sum_i p_i \ln p_i}{\ln \ell}$$

where again we cover the attractor with a uniform grid of boxes of size ℓ. This time p_i is the amount of probability contained in the ith box, and the sum runs only over nonempty boxes.

These definitions have differences, but clearly they must have some similarities as well, since they are all definitions of dimensions. To make the similarity more apparent, note that we can express each of them as being a mathematical statement that the amount of some quantity contained in boxes of size ℓ is varying like a power of ℓ, with that power being a dimension. The box counting dimension can be thought of as:

$$\{\text{number of boxes}\} \sim \ell^{-D_0}.$$

The correlation dimension could be described as:

$$\{\text{number of pairs within a distance } \ell\} \sim \ell^{D_c}.$$

The information dimension is basically:

$$\{\text{information}\} \sim \ell^{-D_I}.$$

The similarity in the more formal definitions becomes even more apparent if we make some modifications to the formulae.

First, note that $p^0 = 1$ when $p \neq 0$. Thus we can suggestively write the number of nonempty boxes in the box counting dimension as $\sum p^0$ instead of $N(\ell)$.

Next, let's unravel the expression for D_c. The expression inside the log in the numerator is $1/N^2$ times the number of pairs of points separated by a distance less than ℓ, with each pair counted twice and "pairs" still counted even if they are formed by counting the same point twice ($i = j$ terms). If we cover the attractor with a uniform grid of boxes of side ℓ, then the number of pairs of points separated by a distance less than ℓ will be closely approximated by the sum of the number of pairs of points in each cell of side ℓ: $\sum_i n_i^2$. This omits pairs formed between points in neighboring cells, but counts some pairs within the cell that are actually slightly farther apart than ℓ (points on diagonally opposite corners of the cell, for example). See figure 1.1 for an illustration of these cases. Moving the factor of N^{-2} inside the sum gives the numerator as $\ln \sum p_i^2$. The ratio of the original quantity to this new approximation is bounded above and below by constants, so that when we take the log and divide by $\ln \ell$, the difference vanishes and the expressions produce the same result for D_c.

Thus we can write the three definitions of dimension that we're investigating in terms of p, which begins to make the expressions more similar to each other;

$$D_0 = \frac{\ln \sum p^0}{|\ln \ell|},$$
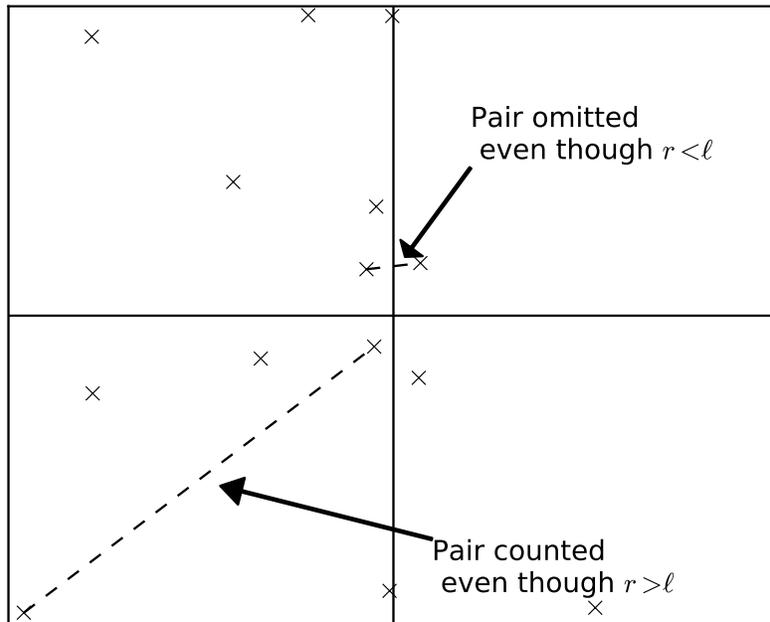
FIGURE 1.1. A schematic illustration of the errors in our approximation when replacing the true correlation function integral with $\sum p^2$. The solid lines are cell boundaries in some two-dimensional space.

$$D_c = \frac{\ln \sum p^2}{\ln \ell},$$

and

$$D_I = \frac{\sum p \ln p}{\ln \ell}.$$

The sums run only over the nonempty boxes that are covering the attractor.

Putting these definitions into similar forms should begin to motivate a common framework for a large family of definitions of dimension. Particularly, it begins to suggest that perhaps we should consider different powers of p (in other words, different moments of the probability distribution), and have the power be the parameter that we vary to obtain various special cases. We are now in a position to begin to define the multifractal spectrum.

That definition is the following. Lay a uniform grid of boxes, each of side ℓ, over the space, and let p_i be the fraction of the attractor contained in the ith box. Then

(1.1)
$$D_q = \lim_{\ell \to 0} \frac{1}{q-1} \frac{\ln \sum_i p_i^q}{\ln \ell}.$$

Note that $q = 0$ gives the box-counting dimension quite easily (and explains that troublesome absolute value in the denominator, which may have seemed out of place when compared to the other dimensions) and $q = 2$ gives the expression that we just argued must be equivalent to the correlation dimension. Examining $q = 1$ is not quite as obvious, but assuming that D_q is a continuous function of q and using L'Hospital's rule will indeed quickly reproduce the $p \log p$ expression above, so that $q = 1$ does give the information dimension. (Rényi [11] originally examined $S_q = \ln \sum p^q/(q-1)$ as a generalized entropy spectrum about 30 years before the multifractal spectrum began being investigated. By our arguments above, S_1 approaches the standard entropy, which is the negative of the information.)

The factor of $q - 1$ might seem odd, but is in fact quite natural. Consider $\sum p^q$.

$$\sum p^q = \sum p^{q-1} p = \langle p^{q-1} \rangle$$

One could argue that rather than looking at an extrinsic property such as total number of points (or pairs of points, etc.), we should examine an intrinsic property such as the average number of points (or pairs, etc.). Doing so would mean framing our quantity of interest as $\langle p^{q-1} \rangle$. To rectify our power, we should then take the $(q-1)^{th}$ root. This would give

$$\langle p^{q-1} \rangle^{1/(q-1)} \sim \ell^D$$

$$\langle p^{q-1} \rangle \sim \ell^{(q-1)D}$$

6

$$\sum p^q \sim \ell^{(q-1)D}$$

If we were to solve this for D, we would obtain the expression for D_q given in equation 1.1. With this new interpretation, it is also clear why $q = 1$ must be examined as a special case. We would be attempting to solve $\langle p^0 \rangle \sim \ell^{(0)D}$ for D.

Notice a small inconsistency in our definitions. If we shift the grid covering the attractor slightly, this could change how the probability is distributed into the boxes of the histogram, which could change the value calculated for D. This is not a truly general problem. It mostly occurs when considering sets with an artificially high degree of symmetry. For example, covering the standard Cantor set (formed by removing the center third of every interval) with a grid with a number of cells that is a power of three will match the boundaries of the set very well, if the center of the histogram matches the center of the set. However, if the grid has a number of cells that is a power of two instead of three, or if the center of the histogram is shifted off the center of the set by some small irrational fraction of the unit interval, then at every length scale, there will be fractal features that extend past the edges of histogram cells. We could tolerate a small mismatch at large scales, or even fluctuations that averaged to zero in some sense, but for a highly symmetric fractal, these slight mismatches will persist to an arbitrary scale, and will often be consistently self-similar, leading to problems. No matter how small the grid cells are, they will always be incommensurate with the actual fractal divisions.

To overcome these problems in an experimental setting, one could repeat a fractal analysis multiple times, with a slightly shifted center each time. This would allow the effects of the shifted center to average out, giving the correct dimension spectrum. There is probably even a simple rule, such as selecting the cover that minimizes some quantity. However, the mere fact that this problem could exist shows that we need a more general definition for D_q, that

deals with the choice of covers more formally. And if we wish to find a minimal or maximal cover, we should not restrict ourselves to covers composed of exactly identical boxes. As well as eliminating our incommensurate problems, using a nonuniform cover will make certain theoretical calculations much easier, and will lend itself better to the transformation between D_q and $f(\alpha)$, which we will describe shortly.

In equation 1.1, notice that if we relax our limit, and imagine investigating the ℓ dependence and extrapolating to $\ell \to 0$, we would be examining

$$\sum_i p_i^q \sim \ell^{(q-1)D_q}.$$

If we denote $(q-1)D_q = \tau$, we can suggestively express the above equation as

$$\sum_i \frac{p_i^q}{\ell^\tau} \sim const,$$

which brings us to a more or less logical reason to introduce a more general definition for D_q. Define the partition function Γ as a function of q and τ to be

(1.2)
$$\Gamma(q, \tau) = \sum_i \frac{p_i^q}{\ell_i^\tau}$$

The ℓ_i in the denominator measures how large the ith box is. It is defined as the longest length possible in the ith box. The probability p_i is the fraction of the attractor contained in that box. Now we no longer require the cover to have uniformly sized boxes. The sum over i runs over all nonempty boxes.

To find D_q from Γ, use the following procedure (from [7]). For $q > 1$ and $\tau > 0$, adjust the cover so as to maximize Γ, and for $q < 1$ and $\tau < 0$, adjust the cover to minimize Γ. Denote the largest ℓ_i in the cover by ℓ, and take the limit as $\ell \to 0$. Doing this, each p_i and

ℓ_i will be shrinking to zero, so that when raised to the q or τ, they will either explode or vanish. In general Γ will either explode or shrink to zero, depending on the values chosen for q and τ. There is a curve $\tau(q)$ along which this transition occurs. If τ is smaller than this critical value, Γ will explode. If τ is larger than the critical value, Γ will shrink to zero. These critical values $\tau(q)$ form a curve[2] which separates regions of $\Gamma \to \infty$ from regions of $\Gamma \to 0$. Find the curve $\tau(q)$ that separates these regions, and then calculate $D_q = \tau(q)/(q-1)$.

To show why the switch between maximizing and minimizing happens, consider a cover for the set, while calculating Γ for $\tau > 0$. If we consider a cell on the edge of the set and expand it, we increase ℓ for that particular cell without changing p. Since τ is positive and $\Gamma = \sum p^q/\ell^\tau$, this has the effect of lowering Γ. In fact, if we cover the set with a single cell, we can increase the size of that cell to make Γ arbitrarily small. Similarly, when $\tau < 0$, we can repeat those steps to make Γ arbitrarily large. This can be accomplished regardless of the actual properties of the set we are covering. On the other hand, returning to $\tau > 0$, to maximize Γ with minimal changes, one could examine every nonempty cell and bring its edges in to eliminate empty parts of the cell near the edge, until that cell was as small as it could be while still containing the same amount of probability. This would make each individual ℓ smaller, which would make Γ larger. Similarly, for $\tau < 0$, making each ℓ smaller would make Γ smaller. In this way, the most snug cover, containing a minimum amount of empty space, which naturally matches the boundaries of the set, is closer to being the optimal cover than an incommensurate cover would be. There are problems with our test cases that could potentially be solved with a $\lim_{\ell \to 0}$, or overcome by starting with different cell configurations, but hopefully this simple sketch helps motivate the maximization and minimization. A schematic showing the different regions of the $q\tau$ plane is shown in figure 1.2.

---

[2]Along this $\tau(q)$ curve, Γ is often finite, but this is not necessarily the case.
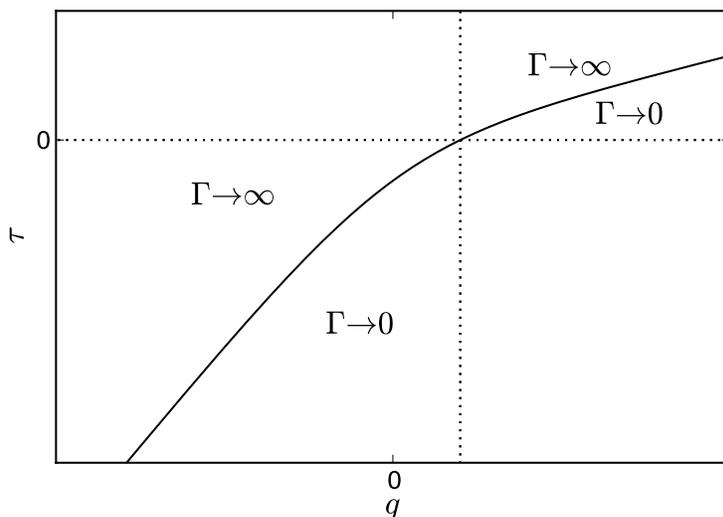
FIGURE 1.2. A plot of $\tau(q)$ for a typical multifractal. Different sets will have different asymptotic limits, and slightly different curvatures around $q = 0$, but are highly similar in their basic shape. As we will show in appendix A, it is highly general for the $\tau(q)$ curve to approach lines of constant slopes as $q \to \pm\infty$. The vertical line is at $q = 1$, the horizontal line at $\tau = 0$. In the upper right quadrant, Γ in maximized. In the lower left quadrant, Γ is minimized. The correct $\tau(q)$ curve is the curve that separates the $\Gamma \to 0$ region from the $\Gamma \to \infty$ region.

Clearly, if we happen to choose a cover of uniform size, this definition will reduce to the same one as in equation 1.1. Along the "correct" $\tau(q)$ curve, $\Gamma \sim 1$, meaning that if the cover is uniform,

$$1 \sim \sum_i \frac{p_i^q}{\ell^\tau}$$

$$\ell^\tau \sim \sum_i p_i^q$$

Remembering $\tau = (q - 1)D_q$ and taking the $\ell \to 0$ limit exactly reproduces equation 1.1.

Although we have not yet described exactly how to numerically generate a $\tau(q)$ curve or a D_q curve, be it from a theoretical knowledge of the set or from an actual data stream, the spectra typically look like the examples shown in figure 1.3 and figure 1.2.
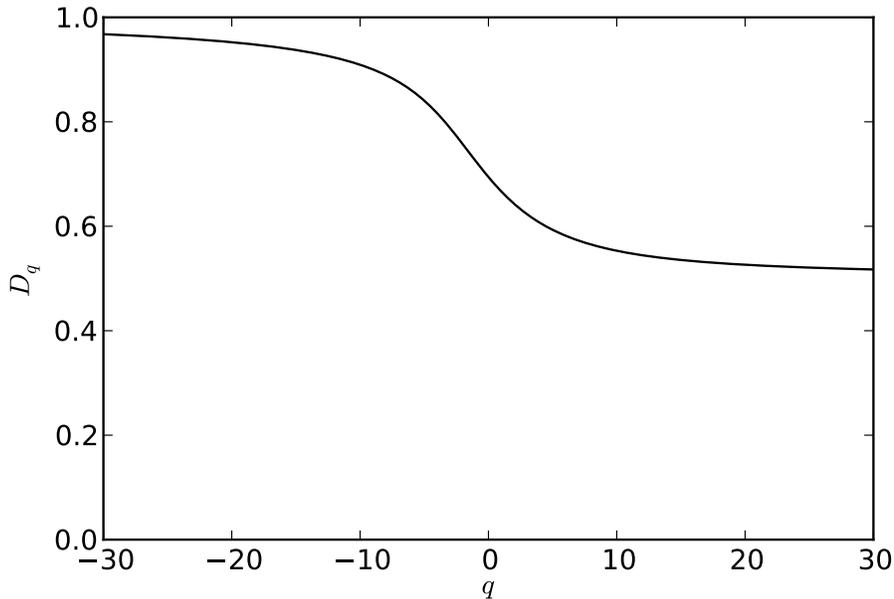
FIGURE 1.3. A plot of D_q for a typical multifractal. Different sets will have different asymptotic limits, and slightly different curvatures around $q = 0$, but are highly similar in their basic shape. (This particular spectrum is for a two-scale Cantor set that places half the probability in the left quarter of the interval, and half the probability in the right half of the interval. Two-scale Cantor sets are described in section 1.2.2.)

1.2.2. THE MOTHER-DAUGHTER FORMULA AND EXACT RESULTS. When calculating dimensions with the standard definitions (the three introduced in the previous subsection, for example), there is a set of standard tricks that can be used to generate analytic results for a limited group of fractals, such as fractals where the self-similarity rules are known. In this section, we will introduce the mother-daughter formula, which is a technique allowing us to analytically calculate the multifractal spectrum in some cases.

Imagine a fractal distribution generated by a known self-similarity rule. Say that it is a Cantor-like probability distribution. Begin with a uniform distribution on the unit interval. At each step, take any uniform sections and divide them into smaller intervals whose lengths are fractions f_i of the original length, with each containing a portion p_i of the probability in the original interval. Figure 1.4 shows the first few stages of construction of a simple

11

case of one of these sets, with only two subintervals. Sets with only two intervals, but with different p and f for the two intervals, are known as two-scale asymmetric Cantor sets. We use them frequently, since they are some of the simplest fractals that can be built that have a non-constant multifractal spectrum.

To begin approximating Γ, follow the same steps as in the construction. Partition the space into intervals that match the first divisions. Any further subdivisions will have redistributed the probability within each interval, but will not have changed the total amount in any partition. Thus, our first approximation for Γ will be:

$$\Gamma_1 = \sum_i \frac{p_i^q}{f_i^\tau} = \tilde{\Gamma}.$$

We will also call this function $\tilde{\Gamma}$, since it is the prototypical simplest approximation we could build for Γ. It encapsulates all the information we need from the construction rule. To make our description more concrete, if at each stage of construction, we took each uniform interval and divided it into a left and right portion, and moved one third of the probability into one quarter of the length on the left, and two thirds of the probability into half the length on the right, we would have

$$\Gamma_1 = \frac{(1/3)^q}{(1/4)^\tau} + \frac{(2/3)^q}{(1/2)^\tau} = \tilde{\Gamma}.$$

Now, for our next approximation, we divide our partition to follow the construction of the fractal. Following our example, we should write out the next version of Γ with four terms.

$$\Gamma_2 = \frac{(1/3 \cdot 1/3)^q}{(1/4 \cdot 1/4)^\tau} + \frac{(1/3 \cdot 2/3)^q}{(1/4 \cdot 1/2)^\tau} + \frac{(2/3 \cdot 1/3)^q}{(1/2 \cdot 1/4)^\tau} + \frac{(2/3 \cdot 2/3)^q}{(1/2 \cdot 1/2)^\tau}$$

$$= \frac{(1/3)^q}{(1/4)^\tau} \left(\frac{(1/3)^q}{(1/4)^\tau} + \frac{(2/3)^q}{(1/2)^\tau} \right) + \frac{(2/3)^q}{(1/2)^\tau} \left(\frac{(1/3)^q}{(1/4)^\tau} + \frac{(2/3)^q}{(1/2)^\tau} \right)$$
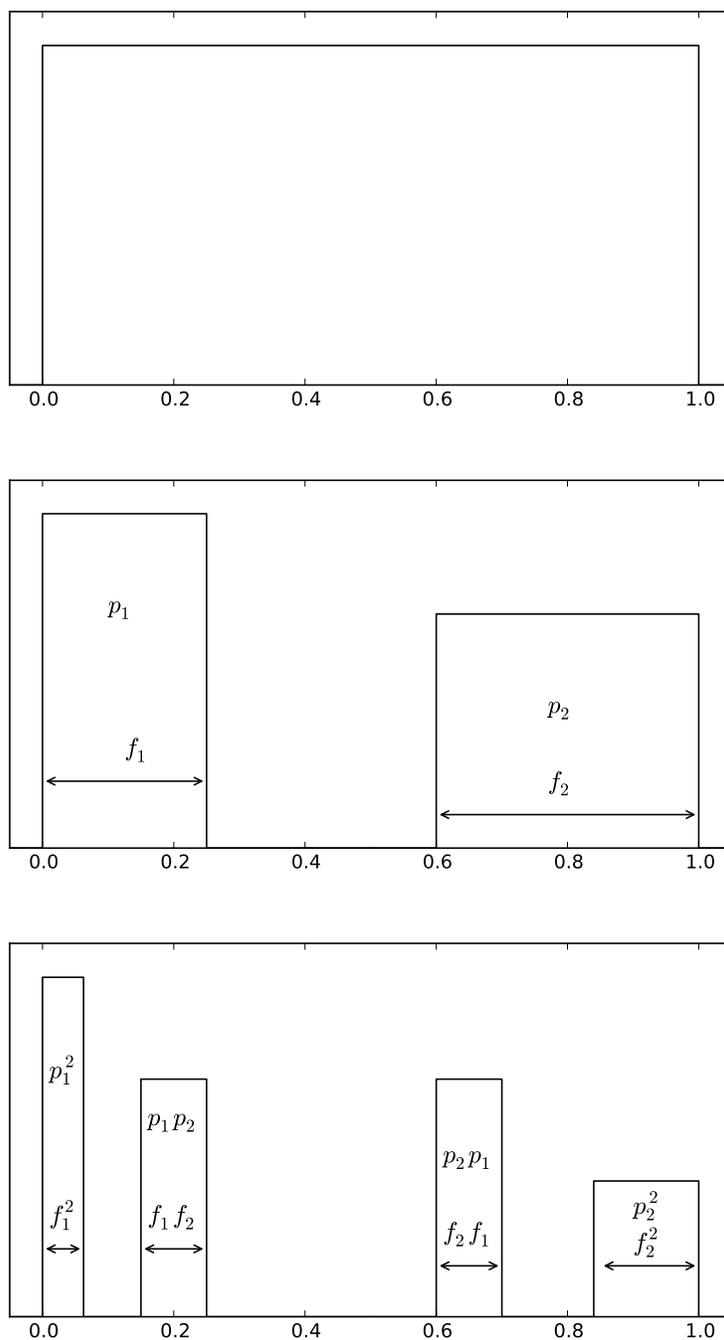
FIGURE 1.4. Illustration of the first few steps of the construction of a two-scale asymmetric Cantor set. Not to scale. In general, it is possible to construct multiscale Cantor sets where at each stage, uniform intervals are split into more than just two sub-intervals.

$$= \left(\frac{(1/3)^q}{(1/4)^\tau} + \frac{(2/3)^q}{(1/2)^\tau} \right)^2$$

In other words, since each interval is subdivided in the exact same manner as every other, we can append a factor of $\tilde{\Gamma}$ to every term in Γ_1, which, for more general sets, factors as:

$$\Gamma_2 = \sum_{i,j} \frac{(p_i p_j)^q}{(f_i f_j)^\tau} = \left(\sum_i \frac{p_i^q}{f_i^\tau} \right) \left(\sum_j \frac{p_j^q}{f_j^\tau} \right) = (\tilde{\Gamma})^2.$$

Moving to the next stage of construction is equivalent to introducing another factor of $\tilde{\Gamma}$. At each step, the partition function will continue to factor in this same way. Each mother segment will have the same daughters. Thus, if we subdivide N times, we will have

$$\Gamma_N = \tilde{\Gamma}^N.$$

As N increases, there are three possibilities: $\Gamma_N \to \infty$, $\Gamma_N \to 0$, or $\Gamma_N = 1$ if $\tilde{\Gamma} = 1$. This is compatible with our assertion that the $q\tau$ plane is divided into regions where Γ either vanished or grew to infinity, with the correct $\tau(q)$ curve dividing the two. Apparently, along that line, for this class of sets, if Γ is finite, $\Gamma = 1$. However, far more importantly, if we know the generating rule, we can solve $\tilde{\Gamma} = 1$ instead of $\Gamma = 1$, using the coarse partition function from the very first step, which only has a handful of terms, rather than the infinite number of terms that one might expect. To make this difference more apparent, we have carefully used f to specify the fraction of the interval in the construction rule prototype $\tilde{\Gamma}$, instead of ℓ, which would be the infinitesimal length of one of the boxes in the fully formed partition function Γ. The equation $\tilde{\Gamma} = 1$ is known as the mother-daughter formula.

Once this simpler partition function has been generated, solving it is straightforward, although the equation may be transcendental, requiring numerical approximation. If numerical results are used, they are still vastly more accurate than results gained from actually covering the space with boxes. Throughout this dissertation, these numerical results will be referred to as "exact" or "analytic" results, despite any small differences between the numerical solutions and the true spectrum. The difference between the spectra obtained through actual analysis of histograms and the exact result is many orders of magnitude larger than the difference between the numerical solution to the transcendental equation and the true exact result.

As a final note, notice that if the set happens to have two subintervals, and if the length of one interval is the square of the length of the other, then $\tilde{\Gamma} = 1$ reduces to a simple quadratic equation. That is to say, if the left interval is a fraction f of the original and the right interval is a fraction f^2 of the original, then the equation becomes

$$p_l^q(f^{-\tau}) + p_r^q(f^{-\tau})^2 = 1,$$

which is a quadratic equation for $f^{-\tau}$, easily solved for $f^{-\tau}$ and then $\tau(q)$. This may seem like an elaborate and unlikely set of conditions, but recall Feigenbaum's proof in [1] that during the period-doubling cascade of all quadratic maps, the orbits double into new orbits that are rescaled spatially by factors of α and α^2. This applies very directly to data generated by the logistic map [12], among others. To find the multifractal spectrum of the probability distribution formed by a system that obeys one of these maps, we could begin by covering the distribution with a cover that subdivided into a self-similar Cantor-like distribution, with the spatial subdivisions being fractions $1/\alpha$ and $1/\alpha^2$ of the whole, and probabilities

being evenly distributed between the two halves. This allows us to analytically calculate the multifractal spectrum for the distribution to which all quadratic maps must converge[3].

1.2.3. CONNECTION TO $f(\alpha)$. Although we have shown how the definition in equation 1.1 reduces to several pre-existing definitions of dimension, and shown how to compute a few exact results with the mother-daughter formula, the basic idea for the multifractal spectrum is still just an arbitrary definition. In this section, we will define the more intuitive quantity $f(\alpha)$, and then sketch out the basics of the proof that $\tau(q) = (q-1)D_q$ and $f(\alpha)$ are Legendre transforms of each other.

Consider some fractal attractor. Different areas of the attractor are qualitatively and quantitatively different from each other. Define a local dimension $\alpha(x)$ through:

$$(1.3) \qquad\qquad n_\ell(x) \sim \ell^{\alpha(x)},$$

with $n_\ell(x)$ being the number of points contained in a box of size ℓ centered on position x. There is an implied limit as $\ell \to 0$. The exponent α is similar to the quantity that one might have expected to define as the information dimension (since the information dimension corresponded to $q = 1$ in the p^q spectrum). For a single isolated point, n_ℓ will be constant and independent of ℓ, so the local dimension $\alpha(x)$ would be zero, as expected. Points distributed along a line will give $n_\ell = \lambda\ell \sim \ell^1$ if ℓ is small enough, giving a local dimension of 1. A well-behaved surface will have $n_\ell = \sigma\ell^2$ as $\ell \to 0$, giving a local dimension of 2. In general, for a smoothly changing probability distribution that fills d dimensions, there will be algebraic prefactors, but the dominant behavior will be $n_\ell \sim \ell^d$, matching our expectations for a

---

[3]It turns out that in reality there are some interesting subtleties. The data generated by the logistic map at the period-∞ accumulation point, for example, does not exactly match the spectrum calculated under these assumptions. See [13] for a more detailed investigation. However, this still serves to illustrate the power of the mother-daughter formula.

smooth manifold. If the point in question is part of a fractal, the local dimension may turn out to be something other than an integer.

Imagine a map or flow with many spikes in its histogram. For example, consider an unstable fixed point. If the system landed near the fixed point, it would evolve away, but it would do so slowly. The closer to the fixed point, the slower it would leave. In fact, the speed of the departure is often approximately proportional to the distance from the fixed point, leading to a distance which increases exponentially with time (one of the hallmarks of chaos is the exponential separation of nearby trajectories). The speed of the motion away from the fixed point would also grow exponentially, so that the time spent at any location near the fixed point would diverge as the distance from the fixed point vanishes. That would lead to a spike in the histogram of exactly the type described earlier. Now consider an unstable periodic orbit. Each of the points in the orbit is an iterate of all the others. The same phenomenological behavior will occur: points closer to the fixed orbit linger near the orbit longer. The iterates could be rescaled, but could all have the same local dimension α. For a flow, there could be a single unstable periodic orbit that winds through the space. Points surrounding that trajectory could reasonably be expected to all have the same dimension as each other, even if they are locally squeezed or stretched a little at different parts of the orbit.

Choosing a particular value for $\alpha = \alpha_0$ selects a collection of points. Perhaps we will select out a dust of points that are the discrete iterates of a single fixed point. Perhaps we will pick the entire boundary of some region, or the curve describing some unstable periodic orbit. Maybe there are no points at all that have local dimension α_0. For each value of α, find the dimension of the set of points selected in this manner and call the dimension of that

17

subset $f(\alpha)$. If there are no points with $\alpha = \alpha_0$ for some particular α_0, then $f(\alpha_0)$ does not exist; the $f(\alpha)$ curve is not necessarily continuous.

In general, selecting different values of α will pick out different interpenetrating fractals from the attractor. Each one of these sub-fractals has its own dimension, giving rise to an entire $f(\alpha)$ spectrum. The spectrum might not be continuous, but for most physical systems f is continuous over a single connected region. This curve $f(\alpha)$ is another description of the multifractal spectrum. This is not the same quantity that we defined in equation 1.1, but after exploring some of the consequences of this definition of $f(\alpha)$, we will show that this quantity is the Legendre transform of $\tau = (q-1)D_q$.

An example can help clarify the $f(\alpha)$ spectrum. Consider the interval $0 < x \leq 1$ with the probability distribution $p(x) = \beta\, x^{\beta-1}$, for $0 < \beta < 1$. This is a probability distribution with a single singularity at $x = 0$. Elsewhere, the distribution is smoothly changing. Forming a small box of size ℓ will enclose $n(\ell) = \int_x^{x+\ell} p(x')dx'$. If we are not near $x = 0$, we can choose ℓ small enough that the probability distribution is approximately constant across the small window over which we are integrating. This will give $n(\ell) \approx p(x)\ell \sim \ell^1$. However, at $x = 0$, there is a spike in the histogram, so that the probability distribution is rapidly changing no matter how small ℓ is, and we cannot ignore the integration. For that one point we will have $n = \int_0^\ell \beta\, x^{\beta-1}dx = \ell^\beta$, which clearly scales like ℓ^β. Thus, the local dimension α is β for the single point $x = 0$, and 1 for all the other points in the interval. This means that there is a one-dimensional set of points with $\alpha = 1$, and a zero-dimensional set with $\alpha = \beta$. So $f(\alpha)$
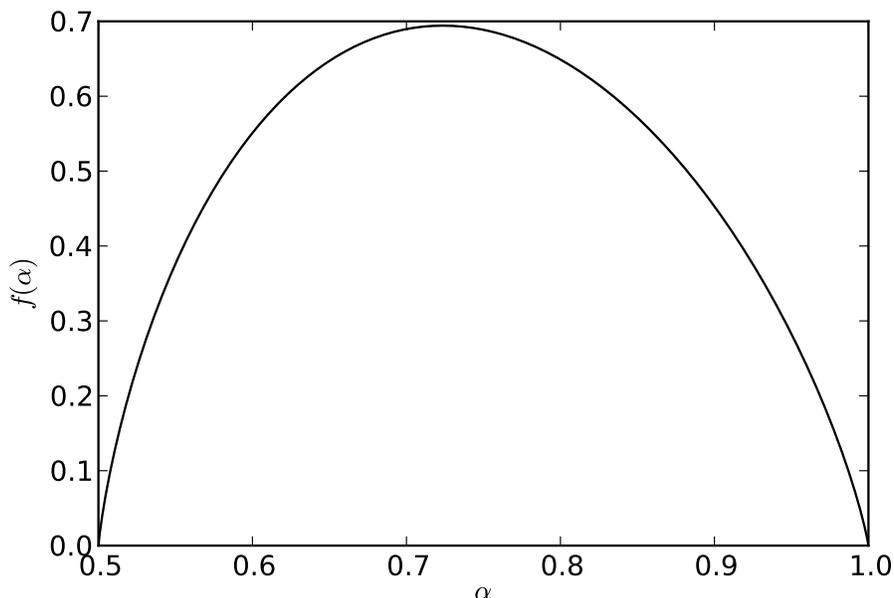
FIGURE 1.5. The $f(\alpha)$ curve for a typical set. The $f(\alpha)$ spectra found in nature are typically very similar to this figure: continuously defined and concave down on an interval between two extremal values of α that represent the highest and lowest local dimensions that exist within the set. (This particular curve characterizes a two-scale Cantor set whic places half the probability in a quarter of the interval, half the probability in half the interval.)

is undefined for most values of α, except for the two points $f(\beta) = 0$ and $f(1) = 1$.

$$
(1.4) \qquad f(\alpha) = \begin{cases} 0 & \alpha = \beta \\ 1 & \alpha = 1 \\ undefined & \text{elsewhere} \end{cases}
$$

This "spectrum" is highly discontinuous, and is not common. For most physical systems, $f(\alpha)$ turns out to be a smooth function defined continuously between a maximum and minimum α value. A more typical $f(\alpha)$ spectrum is shown in figure 1.5.

Now let's discuss the actual connection to D_q. Observant readers will have noticed that although our definition of α is relatively exact, we have been quite vague about $f(\alpha)$. We defined f as the dimension of the set of points with local dimension α, but we did not

specify which definition of dimension we should use. All these definitions of dimension can be thought of as scaling exponents of some quantity. For f, in order to make the Legendre transform to D_q work, we choose the scaling exponent defined by:

$$p(\alpha) \sim \ell^{-f(\alpha)}$$

where $p(\alpha)d\alpha$ is the frequency with which the dimension α takes a value between α and $\alpha + d\alpha$. In other words, we define the dimension f to be the scaling exponent associated with the distribution of values of α. It is analogous to the definition of α in equation 1.3.

Now, examining the definition for D_q in equation 1.1, we approximate $\sum p^q$ with an integral. When integrating, we could choose to integrate over position, but it is more useful to integrate over the measure defined by α. The probability contained in the ith box should be $p_i = n_i/N \sim \ell^\alpha$. Our integral becomes:

$$\sum p^q \sim \int p(\alpha)d\alpha \, (\ell^\alpha)^q \sim \int \ell^{\alpha q - f(\alpha)} d\alpha.$$

This integral is dominated by the minimum of $q\alpha - f$. Calling that value $\alpha' = \alpha'(q)$, we obtain

$$\sum p^q \sim \ell^{\alpha q - f(\alpha)}.$$

There is some prefactor here, but the prefactor is at most logarithmic in size. To compute D_q, we take $D_q = \frac{1}{q-1}\left(\frac{\ln \sum p^q}{\ln \ell}\right)$, so the ℓ dependence is all that really matters. This process will give

$$D_q = \frac{1}{q-1}\left(\frac{((\alpha'q - f(\alpha'))\ln \ell}{\ln \ell}\right) = \frac{\alpha'q - f(\alpha')}{q-1}$$

where α' is defined by the condition that it minimizes $\alpha q - f$. Recalling that we also defined $D_q = \tau/(q-1)$, we can identify

(1.6)
$$\tau = \min_{\alpha}\left(\alpha q - f(\alpha)\right).$$

This minimization condition is exactly the definition of a generalized Legendre transform[4]. Thus we have just shown that τ and f are Legendre transforms of each other (since the Legendre transform is its own inverse). If the minimum is a local minimum instead of an endpoint minimum or some sort of singular point, then this minimization can be performed by taking a derivative. The minimization condition becomes

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}[q\alpha - f(\alpha)]\Big|_{\alpha=\alpha'(q)} = 0$$

or

$$f'(\alpha') = q.$$

To ensure that this is a minimum, not a maximum, we take another derivative of $\alpha q - f$ and require it to be positive, giving $f''(\alpha') < 0$.

Since the Legendre transform is invertible, this means that we can calculate either D_q or $f(\alpha)$. Either curve characterizes the same underlying information. The $f(\alpha)$ picture is more philosophically motivated, and has an easier interpretation, namely that f is the dimension of the set of points with local dimension α. The D_q picture is much easier to calculate, since it is easily expressed in terms of boxes in histograms and has a much more direct series of steps leading from data to results.

---

[4]We should in fact use $\tau = \inf_{\alpha}(\alpha q - f(\alpha))$. A more rigorous derivation can be found in [7].

There are a few common results which can help develop an understanding of the relation between $f(\alpha)$ and D_q. Consider $q = 0$ for a set with a smooth D_q and $f(\alpha)$. In that case, $\tau(0) = (0 - 1)D_0$ can be found through:

$$-D_0 = \min_\alpha \left(q\alpha - f(\alpha)\right) = \min_\alpha(-f(\alpha)),$$

which gives the fairly general result:

$$D_0 = \max_\alpha f(\alpha).$$

As $q \to \pm\infty$, we know that $D_q \to D_{\pm\infty}$. Our Legendre transform gives:

$$(q - 1)D_q = \min_\alpha(q\alpha - f(\alpha)).$$

As $q \to \pm\infty$, f remains finite, so that we have, for $|q|$ large,

$$qD_q \approx \min_\alpha(q\alpha).$$

For $q \to +\infty$, this gives $D_\infty = \min_\alpha \alpha$, and for $q \to -\infty$, this gives $D_{-\infty} = \max_\alpha \alpha$. So it turns out that for many fractals, and the majority of real physical systems, the $f(\alpha)$ curve is a concave down curve defined on $\alpha \in [D_\infty, D_{-\infty}]$, with a maximum value of $f_{max} = D_0$. Since the extremal values of α correspond to points with extremal local dimensions, and those points are often rare, is not uncommon for the values of f corresponding to these limiting α values to be zero. In other words, there is often a single point that realizes the maximum local dimension, and a single point that realizes the minimum local dimension,

which would give $f(\alpha_{max}) = f(\alpha_{min}) = 0$. However, this particular result is far less general than our relations concerning D_0 and $D_{\pm\infty}$. A typical $f(\alpha)$ curve is shown in figure 1.5.

1.2.4. INTERPRETATION OF D_q AND $f(\alpha)$. Recall that D_q is basically found by solving $\Gamma = \sum p^q/\ell^\tau = 1$. When q is large and positive, this sum will be dominated by the areas of the attractor that have the largest probabilities associated with them, and when q is large and negative, the sum will be dominated by the sparse areas of the attractor. Thus, regardless of the particulars of the process of actually solving that equation, in the $q \to \pm\infty$ extremes the dimension D_q can only depend on characteristics of the dense or sparse parts of the attractor. Extending this idea, D_q for any particular value of q can be loosely interpreted as the dimension of a constant-density subset of the attractor. The value of q selects the density in question. This is a very rough interpretation, and is not going to be backed up by any more rigorous arguments. However, it can provide an intuitive handle for the multifractal spectrum. If D_∞ is very low, this can be roughly interpreted as meaning that the dense areas of the attractor are a low-dimensional (i.e., concentrated) subset. It has been proven [8] that D_q is a monotonically decreasing function of q under a relatively general set of assumptions. This can be interpreted as meaning that the sparse areas of the attractor are always more widespread than the dense areas of the attractor, since lower q values will always have a higher D_q than higher q values do.

The same general trends apply to $f(\alpha)$, although this time the relations are far more rigorous. The $f(\alpha)$ spectrum is the dimension of the subset of points where the local dimension is α. The collection of points that share the minimum value for α are the points that have the lowest local dimension. These are the areas of the attractor that are most point-like. They are not guaranteed to be the sparsest areas of the attractor, but they play an intuitively similar role. Similarly, the areas of the attractor that have the highest values

23

of α are the highest-dimensional parts of the attractor, which are the most bulk-like parts of the attractor. These are the regions that are closest to being part of a filled volume, or the appropriate higher-dimensional analog. If $f(\alpha)$ is large for some particular value of α, that intuitively means that the attractor contains 'a lot' of that type of point. We are still being a little vague in our interpretation, but to become more precise would be to return to the actual definitions.

There are some oddities in our interpretation. In appendix A we show that if we consider an asymmetric two-scale Cantor set, D_q asymptotically approaches $\log p / \log f$ as $q \to -\infty$ (p and f are the probability and fraction associated with the least dense branch of the attractor in the construction rule). It is entirely possible for this value for $D_{-\infty}$ to be larger than the dimension of the space containing the attractor. For example, in a one-dimensional space, we could have a two-scale cantor set that placed $\frac{3}{4}$ of the probability in the left $\frac{1}{4}$ of the interval and $\frac{1}{4}$ of the probability in the right $\frac{1}{2}$ of the attractor. This would give $D_{-\infty} = \log \frac{1}{4} / \log \frac{1}{2} = 2$, which is quite surprising for an object that is contained in a one-dimensional space. This is not as strange as it seems, though.

Interpreting $D_{-\infty}$ as the dimension of the sparsest area of the attractor, to truly calculate that dimension, we would form a symmetric Cantor set with the same properties as that sparse area. That would be impossible globally. In our simple example with $\frac{3}{4}$ probability in $\frac{1}{4}$ of the interval and $\frac{1}{4}$ probability in $\frac{1}{2}$ of the interval, in order to produce a set that had a global dimension that matched the dimension of the sparse part of the set, we would have to place $\frac{1}{4}$ of the probability into the left $\frac{1}{2}$ of the attractor, and $\frac{1}{4}$ of the probability into the right half of the attractor. In other words, at each construction step, we would have to remove half of the probability, and distribute the remainder evenly over the unit interval. It is not completely clear what this would even mean. One interpretation would be that in

order to follow this construction rule, we would need additional directions to expand into, so that there was somewhere to put the 'missing' probability at each step. This is a little hard to picture, but at this point we are beginning to assign too specific of a meaning to what was intended to be a loose and qualitative interpretation.

Alternatively, considering this same fractal with $D_{-\infty} > 1$ in the $f(\alpha)$ framework, the local dimension of the rightmost point in the attractor (the point reached by taking the less dense branch of the attractor every time during construction) is the exponent characterizing how the amount of probability varies with box size for a box centered on that point. As we increase the size of a box centered on that point, the amount of probability grows more rapidly than ℓ^1, since the box encounters other branches of the Cantor set. This means that the exponent can be larger than 1 even though the actual dimension of the space is 1. Just like in the D_q picture, one way of understanding this is that if we had a set that followed this local structure on a global scale, that fractal that was thus generated would have to be a higher dimensional shape.

So for either framework, dimensions larger than the dimension of the space are acceptable, even though they disrupt some of the simpler ways of describing the interpretation of D_q and $f(\alpha)$. The limiting value $D_{-\infty}$ is a certain type of dimension of the sparsest parts of the attractor, but it's subtly different from being just the box counting or even topological dimension of the support of those points. A more careful exploration of these issues would eventually lead back to the definitions. Let us leave our description at this level.

Strange values aside, the basic intuitive idea is that α plays a role somewhat similar to q, and f plays a role somewhat similar to D. This should be expected, since the previous section describes how those two pairs of variables are related by a Legendre transform. The $f(\alpha)$ picture could be interpreted as focusing on dimension, while the D_q picture focuses on

density, but the basic relations are similar: lower q or α generally refers to sparser areas of the attractor and larger q or α generally refers to denser areas of the attractor. Higher D_q or $f(\alpha)$ generally means more of the attractor is of the type described by q or α, and lower D or f generally means less.

1.3. Implementation of the Naive Method

Let us now present a description of how the basic multifractal spectrum is actually calculated. How do we apply the theoretical framework from section 1.2 to an actual series of measurements?

As described in section 1.2, we seek to find the curve $\tau(q)$ along which $\Gamma = \sum \frac{p^q}{\ell^\tau}$ is finite, from which we can compute $D_q = \tau/(q-1)$. The problem with finding $\tau(q)$ is that for a finite number of points in a histogram, we can never really realize accurate results as $\ell \to 0$. As ℓ shrinks, we will eventually see a set of discrete points, which should always have dimension zero. To avoid this, we must let ℓ become small, but not smaller than some critical size. Doing this, Γ cannot reach exactly zero or truly diverge to infinity. So how large is large enough for Γ to be considered infinite, and how small does Γ need to be to be approximately zero?

For some sets, we know that along the $\tau(q)$ curve separating the $\Gamma \to 0$ region from the $\Gamma \to \infty$ region, we must have Γ approach a finite value, which may even be exactly 1. However, even if the set under examination is one of the sets where $\Gamma \to 1$, this convergence is often quite slow, and another approach is needed.

Instead of forming cutoffs for the size of Γ and computing Γ at every point on the $q\tau$ plane, or trying to solve $\Gamma = 1$, we instead examine the scaling of Γ with ℓ and extrapolate to $\ell = 0$. Specifically, we begin by using a uniform grid of boxes all with side ℓ. Then our

condition that Γ is finite becomes the scaling relation

$$\Gamma = \sum \frac{p^q}{\ell^\tau} \sim 1$$

$$\sum p^q \sim \ell^\tau$$

$$\sum p^q = A\,\ell^\tau$$

$$\ln \sum p^q = \tau \ln \ell + const.$$

(We could tolerate an A that was not truly constant with respect to ℓ, so long as its ℓ dependence is no stronger than logarithmic.) Finding the ℓ dependence of Γ apparently reduces to finding the slope of $\ln \sum p^q$ versus $\ln \ell$. If that slope is known, we can assume that the same scaling would continue to $\ell = 0$. Also note at this stage that we could write

$$\ln \sum (n/N)^q = \tau \ln \ell + const.$$

$$\ln \sum n^q - q \ln N = \tau \ln \ell + const.$$

$$\ln \sum n^q = \tau \ln \ell + const.$$

Thus it is safe to examine the number of points in a cell instead of the probability in a cell. The normalization affects the offset of the $\ln \sum n^q$ versus $\ln \ell$ plot, but not its slope, which will still be τ. Converting between n and p is trivial, but using n makes the discretization problems all the more apparent, and does save us that one small calculation. We will often switch between p and n within this dissertation.

In order to find that slope, we need $\sum p^q$ for the same attractor, when covered by grids of several different sizes. To implement this, we begin with the finest grid that we anticipate needing. The data is deposited into a histogram with $\ell = \ell_{min}$: the smallest cell size to be
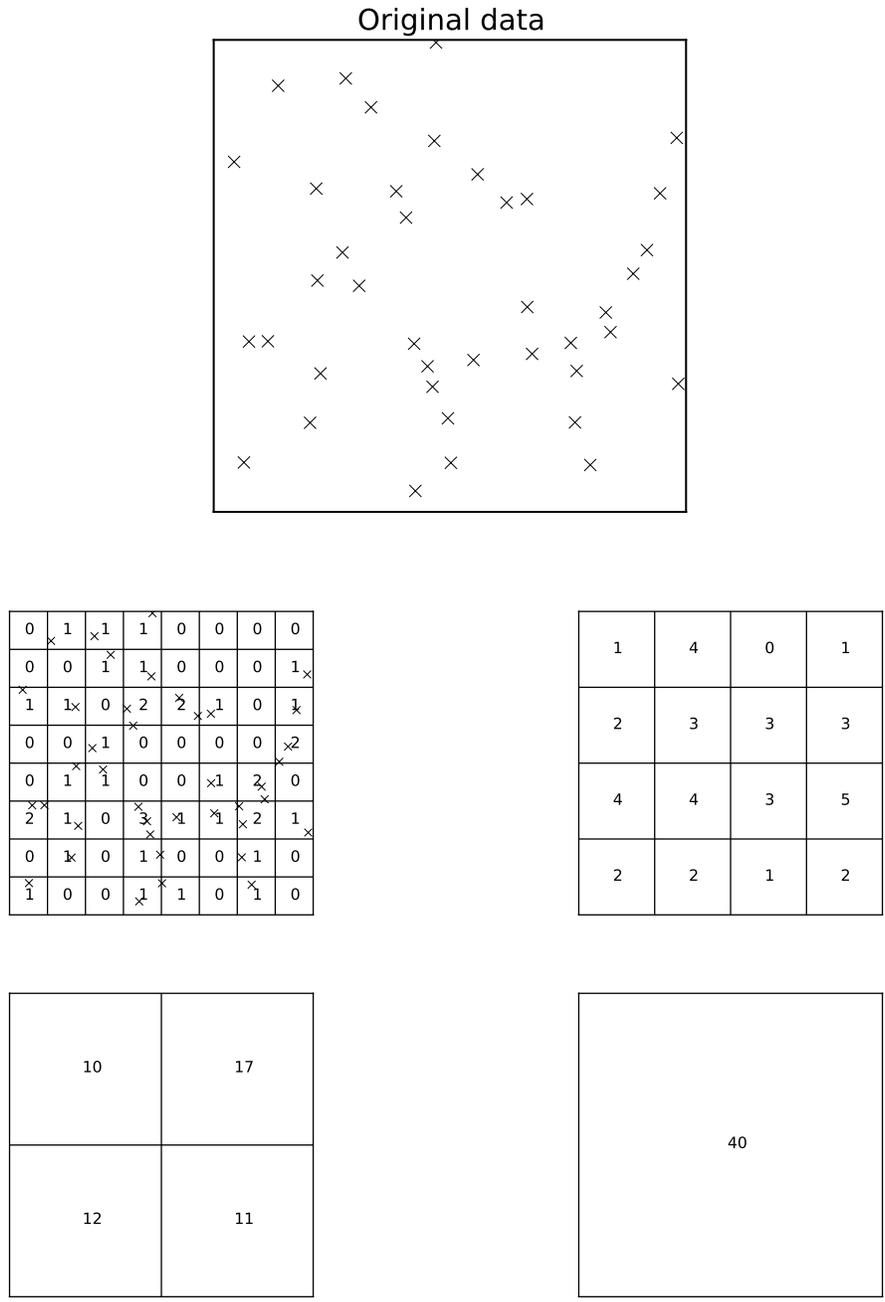
FIGURE 1.6. The rebinning process. The original data is placed in bins, then the values from those bins are combined together to make larger bins, giving a list of p values at different length scales. For clarity, the numbers in these figures are actually n values, not p. The total number of points in the histogram is 40, so each p_i is $n_i/40$.

used[5]. Then neighboring cells in this histogram are combined to produce a new histogram characterizing the same attractor, but with a larger ℓ (larger bins). This new histogram is rebinned into yet another histogram with larger ℓ, and so on until we have generated histograms at many ℓ values. At each rebinning step, the length of a side of a cell is exactly doubled, to ensure that the rebinning does not move the boundaries of the cells in the histograms, it only removes some of the divisions between cells.

Next, we pick a value for q and compute $\sum p^q$ for each of these histograms, in order to produce a plot of $\ln \sum p^q$ versus $\ln \ell$. Each histogram is a copy of the original at a different length scale, so that each histogram gives a single point for our plot of $\ln \sum p^q$ versus $\ln \ell$. A least-squares fit gives the slope of this plot, which is τ for that particular value of q. Finding the slope is an easy process to automate. Then a new value for q is chosen, and the sums are re-computed and the slope re-fitted. Point by point, this procedure generates the $\tau(q)$ curve, which then gives D_q, since $D_q = \tau/(q-1)$.

An illustration of the entire process for a small two-dimensional histogram is show in figure 1.6, table 1.1, and figure 1.7. First, in figure 1.6, we place grids of different sizes over the space containing the data. Then in table 1.1, we choose a q value ($q = 3$ in this case), and compute $\sum p^q$ at each length scale. Those $\sum p^q$ values are plotted in a log-log scale in figure 1.7, and the slope of $\ln \sum p^q$ versus $\ln \ell$ is τ for that particular value of q. To get $\tau(q)$ for other values of q, we would return to table 1.1 and select another value of q, then generate

---

[5]In order to make our approach more general, we used a three dimensional histogram, under the theory that if our method could work on an attractor placed in a space with more than one dimension, it could be easily extended to arbitrarily higher dimensional spaces. When dealing with one or two dimensional fractals, we simply specified the y and/or z coordinate to be some arbitrary value, usually 0.5. Using a three dimensional histogram presented some unique difficulties. For example, in three dimensions, a grid with $2^9 = 512$ cells on a side contains a total of 134,217,728 possible cells, which places great demands on computer memory. Our strategy was to store a list of the occupied cells (ignoring the zero cells) in a hash table, using built-in functions in Python. Most attractors are almost entirely full of empty space, so this allowed us to easily achieve a very scalable implementation. All our examples are one or two dimensional, but all our modifications are carefully phrased in a manner that can be extended to an arbitrary number of dimensions.

TABLE 1.1. A sample calculation of $\sum p^q$ for a single q value, in order to find τ. The numbers used here are the numbers from figure 1.6. Assume that each side of the histogram in that figure had an overall length of 1. For this calculation, let $q = 3$.

$\ell = 1$:	$\sum p^q = (40/40)^q$	$= 1$
$\ell = 1/2$:	$\sum p^q = (10/40)^q + (17/40)^q + (12/40)^q + (11/40)^q$	$= 0.140187500\ldots$
$\ell = 1/4$:	$\sum p^q = (1/40)^q + (4/40)^q + (1/40)^q + \ldots$	$= 0.007187500\ldots$
$\ell = 1/8$:	$\sum p^q = (1/40)^q + (1/40)^q + (1/40)^q + \ldots$	$= 0.001562500\ldots$

another plot, and find its slope. Once the lists of numbers are generated from figure 1.6, the repetitive calculation of $\sum p^q$ and the least-squares fit to get $\tau(q)$ are very fast.

In this process, the step that takes the most computation time is the re-binning step. Once the histograms are found, the position information can be discarded, reducing the histograms to nothing more than lists of numbers. Raising these lists to the qth power and performing the automated log-log fit to find τ is extremely fast. We can thus compute $\tau(q)$
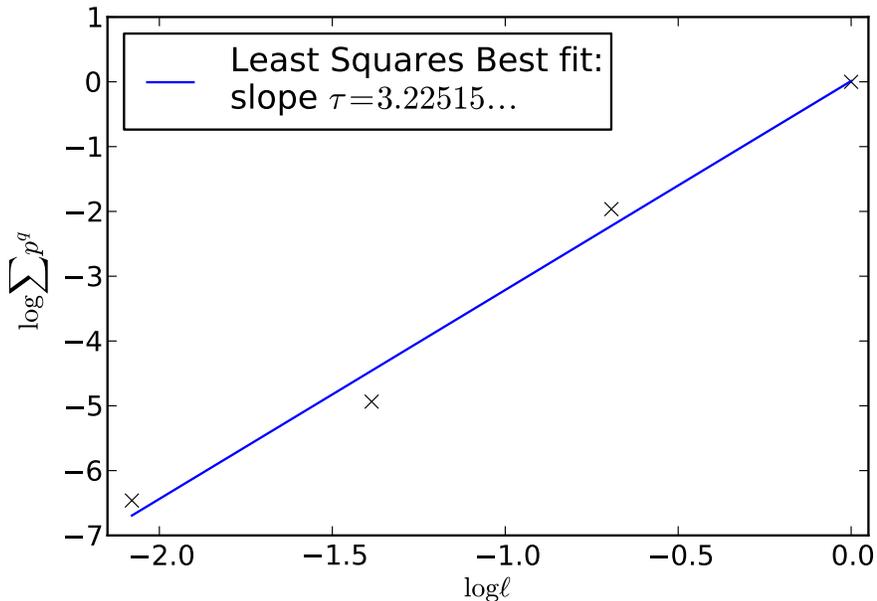
FIGURE 1.7. Sample plot of $\log \sum p^q$ versus $\log \ell$. The slope of this line should be τ. The data here is the data from figure 1.6 and table 1.1. Apparently, for this set, $\tau(3) = 3.23\ldots$, so that $D_3 = \tau/(q-1) = 1.61\ldots$. The logarithms are natural logs.

and D_q for a surprisingly large number of q values without much additional computation time, making a nearly continuous D_q curve. However, the smoothness can be deceptive; even curves representing spectra that have not fully converged yet will be nearly continuous in q.

The family of modifications that we examine consists of using various replacements for p in $\sum p^q$. The rest of the procedure is usually left unchanged. It is worth mentioning here that although we attack D_q in the manner described above, there are other frameworks out there that are entirely different. For example, the so-called 'fixed-mass' approach [14] calculates $f(\alpha)$ from the data, then applies a Legendre transform to find D_q, if D_q is needed. For the most part, our experiments use the procedure described above as a starting point, so we will not go into any detail describing any other approaches.

1.4. PROBLEMS WITH CALCULATION OF THE MULTIFRACTAL SPECTRUM

The multifractal spectrum is difficult to compute, requiring large amounts of data. This slow convergence is especially pronounced for the negative q portion of the spectrum. This is because the negative q part of the spectrum is controlled by the sparsest parts of the attractor, which by definition are the parts that are populated the slowest for most processes.

As described in section 1.2.4, one interpretation of the multifractal spectrum D_q is that the value of q picks out which areas of the attractor to focus on. This is an oversimplification, but it is phenomenologically correct. The negative q part of the spectrum is controlled by the sparse areas of the attractor, since the sum in $\Gamma = \sum p^q/\ell^\tau$ will be dominated by the terms with the smallest values of p.

Clearly, if the attractor is populated by a process that randomly places points according to the natural measure of the attractor, then these sparse areas will be rarely visited. This makes it hard to get an accurate estimate of the measure of the attractor contained in a box

which only encompasses sparse parts of the attractor. Further complicating matters is the fact that for any real measurement of a system, the populating process is not continuous. Discrete measurements are made, so that sparsely occupied cells in a histogram will usually have either zero, one, or two points in them. This means that when computing the partition function Γ, instead of having a collection of terms that have small p values that are not fully converged, but still change smoothly, there will be a collection of cells that have only one point in them. This will mean that once q becomes sufficiently large and negative, Γ will no longer be sensitive to anything except the singly occupied cells, and will cease to reflect the properties of the actual attractor. We can estimate the size of that critical q value. Let the number of singly occupied cells be denoted by $N_{\{n=1\}} = N_1$ and the number of doubly occupied cells be denoted by N_2. Imagine beginning to compute $\Gamma = \sum p^q / \ell^\tau$. When the p^q / ℓ^τ term contributed by the $n = 1$ cells is of the same size as the term contributed by the $n = 2$ cells, q is at its critical value, where the $n = 1$ terms are on the edge of dominating the sum in Γ. Each of the probabilities in the $n = 1$ group will be $1/N$ and each of the probabilities in the $n = 2$ group will be $2/N$. There are N_1 identical $n = 1$ terms and N_2 identical $n = 2$ terms. Thus the $n = 1$ contribution to Γ will be:

$$N_1 \frac{(1/N)^q}{\ell^\tau}$$

and the $n = 2$ contribution to Γ will be:

$$N_2 \frac{(2/N)^q}{\ell^\tau}.$$

32

As q becomes more negative, the $n = 1$ contribution begins to take over. The two contributions are of approximately equal size when:

$$N_1 \approx N_2 2^{q_c} \qquad \Rightarrow \qquad q_c \approx \frac{\ln(N_1/N_2)}{\ln 2}.$$

Notice that if we are computing τ for q more negative than q_c, we will effectively have $\sum n^q \approx N_1 1^q = N_1$. This quantity will depend on ℓ since changing ℓ changes the number of cells, but it is completely independent of q. When finding our slope of the $\ln \sum p^q$ versus $\ln \ell$ curves, we will get $\tau \approx \tau_0$ with τ_0 independent of q, giving $D_q = \tau/(q-1) = \tau_0/(q-1)$. Even if it so happens that N_1 varies in just the right manner so that $\tau_0 = D_{-\infty}$, then this will still produce a D_q spectrum that asymptotically approaches 0 as $q \to -\infty$. We have found empirically that in many cases, when N is small, the estimate of $D_{-\infty}$ starts low, then overshoots before converging on its final value. However, this is by no means truly general, especially when considering our modifications.

A phenomenological demonstration of this behavior is shown in figure 1.8. Various D_q curves were computed for the logistic map, with different numbers of points used to populate the histogram. When the histogram is too sparse, the curves asymptotically approach a wide spread of values for $q \to -\infty$. These values are not related to the actual $D_{-\infty}$ result. Note that although in general the trend is that $D_{-\infty}$ is too low when there is not enough data, this is by no means a universal rule. When the data is too sparse, $D_{-\infty}$ is sometimes overestimated.

This type of error is known as a clipping error, and there are several existing methods that attempt to provide better estimates of D_q for negative q. We are most concerned with the Extended Box Algorithm (EBA).
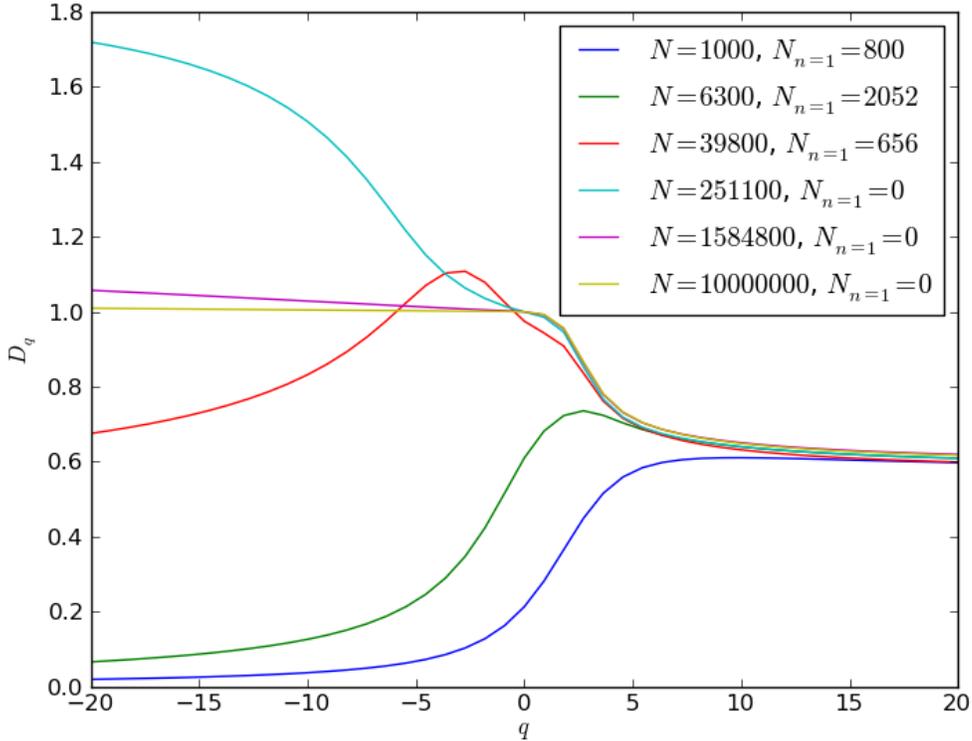
FIGURE 1.8. Plots of the D_q curve for the logistic map as the total number of points is varied. There were 10,000 bins in the histogram. Notice that D_q settles down to the expected curve in the general vicinity of 100 points per cell, although in reality this number is slightly higher, since some cells are empty. Also notice that although $N_{\{n=1\}} = 0$ appears to be a necessary requirement, it is not sufficient. All these curves were calculated with the naive algorithm.

1.5. Detailed description of the existing EBA

Although the basic idea of the extended box algorithm (EBA) proposed by Reidi [15] is quite simple, it is worth a moment to describe here in detail exactly how this idea has been put into practice.

The core idea of the EBA is that in order to more accurately estimate the measure of the attractor contained in each of the boxes in our histogram, we should use the value in that cell, and values in the immediate neighbors of that cell. This extended sum is called n^*. We only collect neighbors for cells that are already populated. If a cell was initially empty,

then we do not calculate an n^* for that cell, we leave it empty. The EBA can be thought of as replacing the occupation of each nonzero cell (or probability contained in that cell) with the average of the occupation of (or probability in) that cell and those of its neighbors (including the neighboring cells that have zero points in them). However, each cell has the same number of neighbors ($3^D - 1$ neighbors for a total of 3^D cells in the average, where D is the dimension of the space containing the attractor). This means that using the actual average would mean using $\sum(n^*/3^D)^q$ in the place of the EBA's $\sum(n^*)^q$. When finding τ, we calculate the slope of $\ln\sum(n^*)^q$ as a function of $\ln\ell$. This extra factor of $(3^D)^q$ would change the offset of that function, but not its slope, so that we can interchange the sum and the average. This is just like the overall normalization factor needed to change between p and n that was shown to be irrelevant in section 1.3.

This entails a very simple modification to our existing algorithm from the previous section. Instead of collecting the values of each cell and raising them to the q, we add a step where we find the neighbors of each cell, sum those neighbors to give n^*, and use n_i^* in place of n_i or p_i. Apart from gathering neighbors to compute n^*, the process is identical to the one described in section 1.3. We place points into a fine-grained histogram, rebin to many scales, then find the slope of $\ln\sum(n^*)^q$ versus $\ln\ell$ for all the desired q values. A basic example of some calculations of n^* for a one-dimensional histogram can be found in table 2.1, although that table is intended to demonstrate a different aspect of the EBA.

The advantage of the EBA is that it allows us to find smoother estimates of the probability contained in each box. The main problem encountered with the naive algorithm is that cells in sparse areas can only be occupied in discrete amounts. A cell can only be occupied an integer number of times. In sparse areas, that integer is usually 0, 1, or 2. As q becomes very negative, the smallest nonzero entries in the p^q sum begin to dominate the sum very strongly.

By considering neighbors, the EBA can be thought of as putting in more options between those integer values (especially if one thinks of the EBA as performing a local average). Cells that are singly occupied can now be counted differently depending on how occupied the neighborhood surrounding them is. This extra variability gives a multifractal spectrum that converges much better.

Another picture of how the EBA works is that it can be thought of as replacing the nonempty cells with expanded overlapping boxes. This has the effect of filling in single-cell-sized holes in the attractor, and moving all the edges outward by one cell, but keeping the edges of the attractor sharply defined. This interpretation is not strictly true, but it helps explain why the algorithm works so well. Sparse regions that are supposed to have filled some space but didn't because of lack of data are smoothed out to become space-filling regions, just like if we had used any other smoothing method, but the fractal edges of these regions are not smoothed away, so that it is conceivable that we could reproduce the right fractal results.

Clipping errors (the discretization errors described in section 1.4) still occur, because if the set is sparse enough there will still be cells that are singly occupied and have no neighbors. We usually have to populate the histogram until there are none of these cells left. However, the EBA only requires us to populate the histogram until there are no singly occupied cells without any neighbors, while the naive algorithm would require us to populate until there are no singly occupied cells at all. Of course this is a rough indication of whether the attractor is populated enough for the algorithm to produce reliable results. A more detailed analysis would again examine the ratio of the contributions due to the smallest possible p (or p^*) values. However, this analysis would be considerably more complicated without changing the phenomenological nature of the result.

Finally, note that just as with the naive algorithm, we can use p and n interchangeably when finding τ. Technically, finding Γ with the new definition, we should compute

$$\Gamma = \sum \frac{(p^*)^q}{\ell^\tau} = \sum_{\substack{nonzero \\ cells}} \frac{(\sum_{neighbors} p)^q}{\ell^\tau},$$

but each p is $p = n/N$, so that this is

$$\Gamma = \sum_{\substack{nonzero \\ cells}} \frac{(\sum_{neighbors} n/N)^q}{\ell^\tau} = N^{-q} \sum_{\substack{nonzero \\ cells}} \frac{(\sum_{neighbors} n)^q}{\ell^\tau}$$

$$= N^{-q} \sum \frac{(n^*)^q}{\ell^\tau}.$$

In section 1.3, we showed how the factor of N^q difference between $\sum p^q$ and $\sum n^q$ did not affect the calculation of τ. We have just shown that even though we are now summing up neighbors, the difference between $\sum (p^*)^q$ and $\sum (n^*)^q$ is that same factor of N^q, so it still follows that the two sums can be used interchangeably.

1.6. IMPROVEMENTS

This concludes our survey of the existing methodology, and our description of the basic problems encountered when applying this methodology to small data sets. The most important features of this chapter are the description and motivation of the multifractal spectrum itself, the exploration of the problems encountered with small data sets, and the framework of the EBA. Now we begin our investigations of ways to improve these calculations. Our general strategy is to consider new possible definitions of n^* within the general framework of the EBA, although the Patchwork method in chapter 5 is a considerably different approach. From this point on, unless otherwise noted, all the ideas and modifications are original work for this dissertation.

IMPROVEMENTS TO THE EBA

2.1. THE NEED FOR NORMALIZATION

As mentioned previously, the calculation of τ for many values of q is very fast, so it is easy to make a very detailed investigation of the region near $q = 1$. When this is done, one usually finds that the EBA produces a narrow spike where $D_q \to \pm\infty$ near $q = 1$. This spike is caused by the fact that with the EBA, although convergence of τ has been improved for most q values, it has been ever so slightly worsened right at $q = 1$, so that τ no longer crosses exactly through zero at that point. Dividing by $q - 1$ thus gives a divergent result for D_1, since $D_q = \tau/(q - 1)$. Clearly, no matter which interpretation we use for the multifractal spectrum, it must be a dimension-like scaling exponent of some kind, and for any kind of reasonable set, it should be finite. The particular dimension D_1 is the information dimension, which is infamous for being one of the most difficult dimensions to calculate. The authors of [15] and [16] make no mention of this spike, perhaps because of a different implementation of the algorithm that does not produce D_q at as many q values as ours does.

Understanding the resolution of this problem is simple if one understands the real cause. When $q = 1$, the sum $\sum p_i^q$ becomes $\sum p_i$ which is just the sum of all the probabilities in the histogram. This should be exactly 1, regardless of the size and number of bins that have been used. For other values of q, having a different number of entries in the list of p values will change the value of the sum, since the individual terms are raised to the qth power. However, $q = 1$ is the one spot where that should not matter. It is usually simpler to use n_i or n_i^* rather than p_i or p_i^*, but the principle still holds; instead of $\sum p_i^1 = 1$, we expect to see $\sum n_i^1 = N$. The actual value of the sum isn't important. What matters is that for this

particular q value, we cannot tolerate any ℓ dependence in the sum. Using n^* or p^* will no longer maintain a consistent normalization at different length scales. When finding n^* or p^*, cells will often be counted more than once.

More concretely, consider the one-dimensional example in table 2.1. We begin with a one-dimensional histogram of different n values. To generate n^*, we replace each nonzero cell with the sum of itself and its left and right neighbors (if those neighbors exist). Then, we rebin the n values and compute a new n^* list with the same procedure. As we rebin to coarser and coarser grids, there are fewer empty cells. If the histogram has no empty cells at all, then each cell will be counted three times: once when n^* is computed for the neighbor to the left, once when n^* is computed for the neighbor to the right, and once when n^* is computed for the actual cell in question. If that happened consistently at every length scale that was examined, there would be no problem. The algorithm would always overcount the total by a factor of 3, so that there would be no ℓ variation, and the slope of $\ln \sum n^*$ versus $\ln \ell$ would be zero as required. However, as long as we have some empty cells, we will not have a universal factor of 3. Cells with all nonzero neighbors will still be counted thrice, but cells with one zero neighbor will only be counted twice and cells with two zero neighbors will be counted once. This inconsistency in the amount of overcounting will change the total $\sum n^*$ as we change ℓ.

Knowing that normalization is the issue suggests the simple fix of renormalizing the histogram. However, there are multiple ways of accomplishing this, and the different procedures do not perform equally. Before discussing the performance of the methods, we should introduce one of our metrics for measuring convergence.

Table 2.1. A one-dimensional demonstration of how the EBA does not preserve normalization. The problem is not that the total $\sum n^*$ does not equal $\sum n$, it is that the value of $\sum n^*$ depends on which length scale we consider. Cells adjacent to the edge are considered to have a neighbor of zero on that side. Imagine an infinite field of empty cells extending past the edges of the histogram shown here.

original n values:	1	0	1	1	0	0	2	0	1	0	1	1	0	0	2	0	$\sum n = 10$
n^* values (add neighbors):	1	0	2	2	0	0	2	0	1	0	2	2	0	0	2	0	$\sum n^* = 14$
rebinned once n values:	1		2		0		2		1		2		0		2		$\sum n = 10$
n^* values:	3		3		0		3		5		3		0		2		$\sum n^* = 19$
rebinned twice n values:	3				2				3				2				$\sum n = 10$
n^* values:	5				8				7				5				$\sum n^* = 25$

2.2. CONVERGENCE TESTS

To judge the convergence of a calculation of the D_q spectrum, we choose to investigate $D_{-\infty}$ because that part of the spectrum is usually the slowest to converge, and thus gives us the most sensitive measure of the convergence of the entire curve. The curves that we will produce in this manner will not measure the width of the spike at $q = 1$, if it exists. However, we have empirically found that the width of the spike is closely related to the convergence of the rest of the curve, which is best measured by checking $D_{-\infty}$. Furthermore, when attempting to remove the spike, the method seems to either remove the spike entirely or leave it intact. The removal is not a continuous process that gradually reduces the size of the spike as more points are added. There is no convergence associated with the removal of the spike. Thus, we examine how the removal of the spike affects the rest of the spectrum.

As $q \to -\infty$, the D_q curve should asymptotically approach $D_q = \frac{q}{q-1}D_{-\infty}$ (see appendix A for details). Thus, $\tau = (q - 1)D_q$ should approach $\tau = qD_{-\infty}$. That is to say, it should approach a straight line whose slope is $D_{-\infty}$. Finding that slope of the τ curve helps to correct for any residual curvature, giving an estimate of $D_{-\infty}$ at values of q for which D_q might not be perfectly level.

The basic problem that we will encounter with every method is that although we sometimes have an analytic D_q spectrum to compare our results to, this is not always the case. We do occasionally have trustworthy calculations of certain specific dimensions for various sets from the literature. For example, [17] numerically calculates the box-counting dimension of the Lorenz attractor at a particular set of parameter values, which is D_q at $q = 0$ for that set, [18, 19] compute the correlation and Hausdorff dimension for the Hénon map, which are D_2 and D_0 for that set, and [20–23] analytically (or pseudo-analytically) compute the entire spectrum for relatively general sets of cases. However, these calculations are generally based

on perfect knowledge of the probability distribution. Our focus is on building a method that handles an actual stream of data measurements. So we often find ourselves trying to determine whether a method has converged, without knowing what value it was supposed to converge to. To overcome this problem, we used the following procedure.

We produced many instances of histograms populated with the same number of points N as each other through probabilistic processes. These could be produced by integrating a chaotic flow started with some random initial conditions, or perhaps by using a random number generator to place points into a Cantor-like fractal probability distribution. We then computed the multifractal spectrum for each instance of the histogram, and obtained an estimate of what value the spectrum was asymptotically approaching as $q \to -\infty$. These estimates were found by computing the slope of $\tau(q) = (q-1)D_q$ for very negative q, usually in the range of -20 to -30.

Once an asymptotic estimate was gathered for each histogram instance, we found the average and standard deviation of the collection of $D_{-\infty}$ estimates and stored the results. Then another set of histograms was populated, at a larger value of N, and the whole process was repeated. The process of convergence usually seemed to be more or less linear in $\log N$. As N increased, these estimates would eventually saturate. As mentioned above, for some fractal sets we had exact results available to confirm that the saturation value was the correct one, but in most cases we do not have solid proof of that. In general (but not always), when N is low, the set is closer in character to a disconnected set of points, which would be a zero-dimensional set, so as N increases, the estimate of $D_{-\infty}$ tends to rise toward the value that correctly characterizes the completely filled-in set. A typical $D_{-\infty}$ versus $\ln N$ convergence curve is shown in figure 2.1.
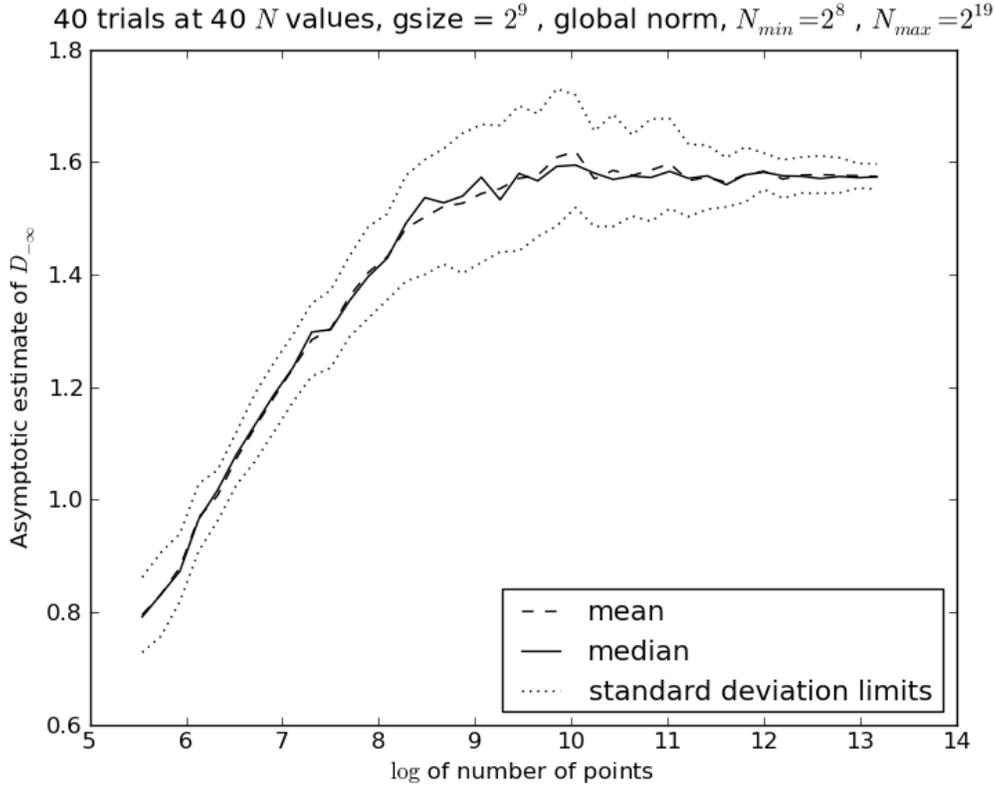
FIGURE 2.1. A typical convergence plot. This plot shows the convergence of the 'global norm' modification to the EBA, applied to our standard Cantor-like probability distribution. As N increases, the asymptotic estimate of $D_{-\infty}$ converges to a single value. Most plots of this type will use the standard deviation of the mean, which is $\sigma/\sqrt{N}$ (N being the number of trials). For this plot, we have exaggerated the error bars for clarity by using just the standard deviation σ. Further, all logarithms throughout this dissertation are natural logarithms.

2.3. GLOBAL NORM

The most effective and simplest way to remove the spike without changing the basic definition of the EBA is to normalize the histogram by dividing each n^* by a global factor of $N^* = \sum n^*$. Doing this converts our process from solving $\ln \sum (n^*)^q = \tau \ln \ell + A$ for τ to solving the following for τ:

$$\ln \sum (p_{GN}^*)^q = \tau \ln \ell + A$$

$$\ln \sum \left(\frac{n^*}{N^*} \right)^q = \tau \ln \ell + A$$

$$\ln \frac{\sum (n^*)^q}{(\sum n^*)^q} = \tau \ln \ell + A.$$

Clearly, at $q = 1$, this extra renormalization factor will ensure that τ is exactly zero, since the left side will be exactly zero, independent of ℓ, making its slope (with respect to ℓ) exactly zero. At first glance, it may appear that the factor of $(\sum n^*)^q$ could be grouped with the other unknown constants, just like we did in section 1.3 to change from p to n with an ignorable factor of $N = \sum n$, but recall that the problem we are attacking with this modification is that the sum $\sum n^*$ depends on ℓ, so that this term is not a constant.

Note that this is a renormalization after the rebinning step. It is not the difference between using p^* and n^*, which was discussed earlier, and shown to be irrelevant. Instead, it is a renormalization after the counting of neighbors in the step that generates n^*, to produce a normalized probability distribution. Relative to the basic EBA, this renormalization introduces small changes away from $q = 1$, but removes the infinite spike at $q = 1$. We will refer to this particular renormalized version of the EBA as the Global Norm method.

In the perfect data limit, with $N \to \infty$, the grid infinitely fine, and each cell well populated, the Global Norm method, the EBA, and the naive method should all produce the same results. The naive method should find $\Gamma = \sum p^q/\ell^\tau$. In particular, at $q = 1$, we will have $\Gamma = \sum p/\ell^\tau = 1/\ell^\tau$. The EBA will use $\Gamma = \sum (n^*)^q/\ell^\tau$, but in this limit of perfect population, almost every cell will have all its neighbors populated, so that each cell will be counted the same number of times. This will lead to an almost universal overcounting factor, which will reduce the width of the infinite spike towards zero. Thus, adding the normalization step introduced for Global Norm will become irrelevant. The change is that with the extra normalization factor, the spike is exactly removed, at any value of N, not just in the $N \to \infty$ limit.

Logistic map, 4096 points
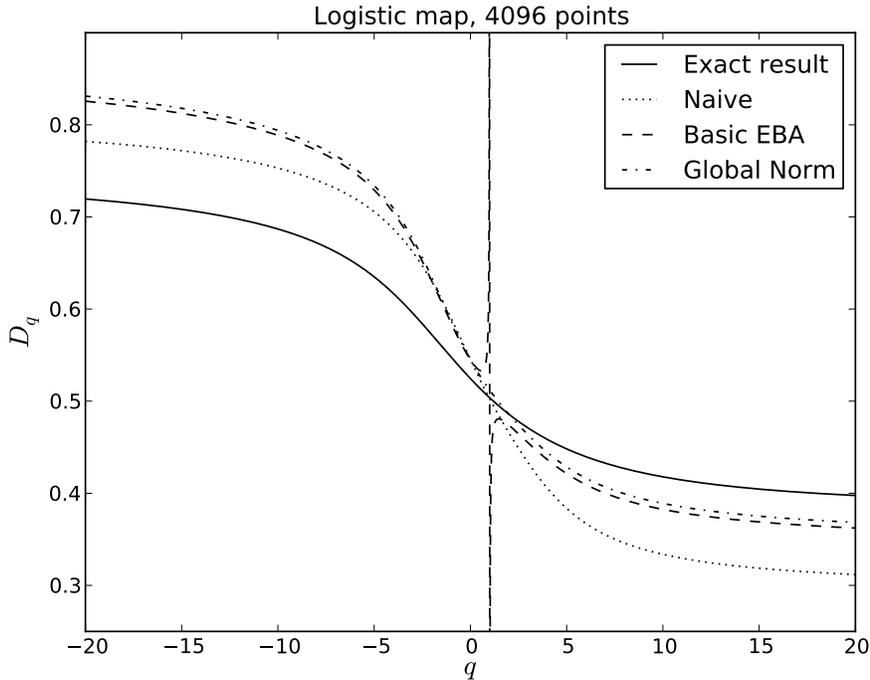
FIGURE 2.2. A D_q plot for a typical instance of the logistic map, comparing various methods. The basic EBA shows the normalization spike at $q = 1$. The Global Norm curve is very similar to the basic EBA, except for the removal of the spike. The exact result curve is an approximation calculated by treating the logistic map limit set as a 2-scale Cantor set with probabilities $\frac{1}{2}$ and fractions $1/\alpha$ and $1/\alpha^2$, which may not be perfectly accurate. The various methods are clearly approaching different $D_{-\infty}$ values, but that is not the main feature here. The main feature is the spike at $q = 1$.

We found that Global Norm performed about as well as the basic EBA for most q values, but removed the spike at $q = 1$. See figures 2.3, 2.4, and 2.5 for convergence curves of Global Norm compared to other methods, on various test sets. See figure 2.2 for a typical D_q spectrum of the Global Norm method, showing that the spike is removed.

2.4. LOCAL NORM

This brings us to our first substantial modification of the EBA. If we are considering the EBA to be a type of mean, why not take a local average instead of a global average? To be more precise, instead of adding or averaging all the neighbors, average only the nonzero

neighbors. This means that if a cell i has one nonzero neighbor, which we'll call j, then we replace n_i with $n_i^* = (n_i + n_j)/2$. The basic EBA would instead use $n_i^* = n_i + n_j$, which, if we think of the EBA as taking an average, could be interpreted as $n_i^* = (n_i + n_j)/3$ (divide by 3 if in a one-dimensional space, 9 in a two-dimensional space, or 3^D in D dimensions). Notice that if we are in a dense region where no cell has any neighboring cells that aren't occupied, then the two variations will produce the same n^* values; only the cells that are on the boundaries will have different n^* values.

Recall that we described how the EBA could be thought of as an average over neighbors. We could have replaced the sum over neighbors by a sum $n^* = \sum_{neighbors} n/N_{cells} = \sum_{neighbors} n/3^{D_S}$ with D_S the dimension of the space containing the attractor. (We showed that that so long as that factor was universal, we could ignore it.) One could argue that using the dimension of the space in this manner introduces a bias. By allowing the experimenter to choose the number of dimensions available, perhaps the values of n^* are not changing correctly. One way to remove that bias would be to use the Local Norm method, and divide by the number of occupied neighbors instead of the number of possible neighbors. That number of occupied cells would naturally scale like some sort of box counting dimension, independent of the dimension of the space the attractor is placed in.

Notice that the Local Norm does not exactly renormalize n^*. Since not all cells have the same number of nonzero neighbors, not all the cells will be renormalized by the same amount, so that as we rebin and change the distribution of occupied cells, we will have some small variations in $\sum n^*$. See table 2.2 for a demonstration. This method is introduced not as a way of removing the spike at $q = 1$, but as a modification to the EBA that could potentially improve convergence; to remove the spike at $q = 1$, we apply a global normalization to the

TABLE 2.2. A one-dimensional demonstration of how Local Norm does not preserve normalization. Notice that cells on the edge of the attractor (cells with neighbors that are zero) are counted less in the final sum than cells on the interior of the attractor (cells with all neighbors nonzero). More importantly, notice that the total of the n^* values changes when rebinned.

Original cell values	0	a	b	c	d	0	$\sum n = a+b+c+d$
Local Norm n^* values	0	$\frac{a+b}{2}$	$\frac{a+b+c}{3}$	$\frac{b+c+d}{3}$	$\frac{c+d}{2}$	0	$\sum n^* = \frac{5a+7b+7c+5d}{6}$

rebinned once values	a		$b+c$		d	$\sum n = a+b+c+d$
new Local Norm values	$\frac{a+b+c}{2}$		$\frac{a+b+c+d}{3}$		$\frac{b+c+d}{2}$	$\sum n^* = \frac{5a+8b+8c+5d}{6}$

locally normalized data in all of our calculations. This variation of the method will also have strong parallels with certain classes of our other modifications.

The benefit of the Local Norm method is that it makes the variations between occupations take on values that vary more smoothly, closer to being continuous. In the naive method, a cell can have occupations 0, 1, 2, etc. In the basic EBA, the occupations are still the sums of integers, so the possible values will still be 0, 1, 2, etc. (although, hopefully there will be fewer cells with $n^* = 1$ in the final list that is generated). The Global Norm will make all these numbers noninteger, but will do so with a universal renormalization factor, so that the occupations will become, for example, $1/100, 2/100, 3/100$, etc. The values are evenly spaced, so that when the final renormalization is applied, the occupation values in all three methods are equivalent. (This is exactly the same as the arguments showing that the universal normalization factor that changed between p and n was irrelevant.) For all these methods, regardless of the types of renormalization used, the gaps between possible occupation values are all integer multiples of the smallest p^* value. The Local Norm changes this by adding new possibilities for the occupations of the sparsely occupied cells. The minimum occupation is still 1, but the next possibility is the n^* formed from a field of ones with a single two in the mix. The n^* value for a cell with one doubly occupied neighbor will

be

$$n^* = \frac{1(N_{neighbors} - 1) + 2(1)}{N_{neighbors}} = 1 + \frac{1}{N_{neighbors}}.$$

The next possible value for n^* is $1 + 2/N_{neighbors}$, and so on. Note that $N_{neighbors}$ is the number

of neighbors that the cell has, including itself. In one dimension, $N_{neighbors}$ can take any value

up to 3, in two dimensions the maximum $N_{neighbors}$ is 9, and in D dimensions the maximum

is 3^D. The second-smallest occupation will occur when $N_{neighbors}$ is at its maximum, but

notice that different cells could have different numbers of neighbors, especially in higher

dimensional spaces, which could lead to gaps smaller than that minimum occupation value.

The important feature here is that the ratio of the size of a gap compared to the size of the

smallest occupation value is much smaller.

This gap is actually the most important feature. Recall our analysis at the end of

section 1.4, where we found the critical value of q beyond which the singly occupied cells

dominated the sum in Γ. The core of that analysis was a comparison of the contributions

of the singly and doubly occupied cells. Expanding that analysis to cover the Local Norm

method, we should compare the largest terms in Γ with the next largest terms. These will

be the $n^* = 1$ terms and the $n^* = 1 + 1/N_{neighbors}$ terms, which I will write as $1 + \epsilon$. It's

harder to tell how many $n^* = 1$ terms there will be, but call the number of those terms N_1.

Label the number of $1 + \epsilon$ terms as N_ϵ. The contribution to Γ from the $n^* = 1$ terms will be

(recall $\Gamma = \sum p^q / \ell^\tau$, but we demonstrated that it was safe to use n in place of p):

$$N_1 \frac{1^q}{\ell^\tau}.$$

The contribution from the $1 + \epsilon$ terms will be

$$N_\epsilon \frac{(1 + \epsilon)^q}{\ell^\tau}.$$

The two contributions will be of approximately equal size at some critical $q = q_c$, determined by

$$N_1 \approx N_\epsilon (1 + \epsilon)^{q_c},$$

which gives:

<div style="text-align:right">(2.1)</div>

$$q_c \approx \frac{\ln(N_1/N_\epsilon)}{\ln(1 + \epsilon)} = -\frac{\ln(N_\epsilon/N_1)}{\ln(1 + \epsilon)}.$$

For comparison, the naive method gave $q_c \approx \ln(N_1/N_2)/\ln 2$. Recall that when $q > q_c$, the two contributions to Γ are of equivalent size, and we conclude that our estimate is smoothly varying, but when $q < q_c$, Γ is dominated by the lowest occupation, which is called a clipping error. In general, if a method produces a minimum occupation n_1 and has N_1 cells with that occupation, and N_2 cells with the next-smallest occupation n_2, we will find that $q_c = -\ln(N_1/N_2)/\ln(n_1/n_2)$. For Local Norm, since ϵ is small, it is possible that this could mean that the critical q (beyond which the sum is not smoothly varying) is quite large in magnitude. However, we have not analyzed the number of terms at all. If the number of $n^* = 1$ cells approaches the number of $n^* = 1 + \epsilon$ cells, this critical q might not be very large, and may even be positive (although a positive q would invalidate the starting assumptions of our approximation).

For another reason that the Local Norm could be an improvement, examine the edges of a set. Specifically, consider a one-dimensional set, populated with some probabilistic process that converges to a uniform probability distribution that will eventually place n_0 points in

each cell. Applying the EBA to the set, the interior of the interval will be more smoothed toward its correct value. Each of the cells on the interior will converge toward $p^* = 3n_0$. However, the edge cells will approach $p^* = 2n_0$, since each of those cells only has a single neighbor. In the negative q, large N limit, those two cells will begin to dominate the Γ sum. They will basically act like a low density pair of points (a zero-dimensional set), giving a spuriously low estimate of $D_{-\infty}$. Applying Global Norm will not change this, since it is a global renormalization factor, so that the ratio of the occupation of the edge cells to the occupation of the interior cells will still be 2/3. However, when applying the Local Norm, the edge occupations will approach the same value as the interior occupations. In the interior we will have $p^* = (n_0 + n_0 + n_0)/3 = n_0$, and on the edges we will have $p^* = (n_0 + n_0)/2 = n_0$. So Local Norm will not lower the probabilities on the edges of regions. We examined a uniform interval, but clearly the basic idea extends to any sort of region with an edge abutting an unoccupied region.

These arguments (more closely spaced possible values for p^*, and reduced edge effects) establish that the Local Norm has the potential to be an improvement. However, there are also cases in which the Local Norm method might not be an improvement.

To show a potential weakness of the Local Norm method, compare two cases: a line of singly occupied cells, with every third cell removed and replaced with a zero, and a line of singly occupied cells with every fourth cell replaced with a zero. Clearly the two cases should have different densities. However, the difference in density does not come from a difference in occupations, but from from a difference in the number of occupied cells. The effect of Local Norm is to precisely remove the influence of the number of neighbors, and only consider variations in the occupation number. The average of 2 ones is the same as the average of 3 ones: one. Rather than interpreting the region as a less dense region with some patches

missing (some unoccupied cells), Local Norm interprets the region as a series of unconnected uniform regions, and does not allow the edges of these regions to decrease the estimate of the probability there. Framed a little differently, Local Norm produces too many of the lowest occupation terms and not enough of the next-lowest terms, so that N_1/N_ϵ in equation 2.1 is too close to 1, meaning that the critical q below which D_q is dominated by clipping errors is not very negative. So for very sparse regions, the Global Norm produces a more accurate estimate of the actual density, although once the region is fully occupied, we might still see some edge effects.

2.5. PERFORMANCE

We display several convergence plots showing how our various modifications change the measurement of $D_{-\infty}$ for various systems. Calculating a single D_q curve for data sets with on the order of 10^6 points in a grid of side 2^{11} takes a time on the order of a minute on a laptop computer. This time scales with the number of occupied cells, and if a hash table is used, the time scales approximately linearly with the number of occupied cells.

Figure 2.3 shows the methods applied to an asymmetric Cantor set. As N increases, the EBA overshoots before settling to the correct value. The Local Norm method reduces the amount of overshoot. The Global Norm method closely follows the basic EBA results.

Figure 2.4 shows the methods applied to the logistic map at $\lambda = 4$. The logistic map is best-known for its period-doubling transition to chaos, and is often studied near the period-∞ accumulation point $\lambda \approx 3.57$, since at that parameter value the attractor for the logistic map is easy to model as a Cantor set. However, we found small rounding errors were causing the least visited parts of the attractor to be far more chaotic than this model would account for, so instead we studied the logistic map at $\lambda = 4$, which is deep in the

FIGURE 2.3. Convergence curves for the various improvements that we have developed to the EBA, when applied to an asymmetric Cantor set. The exact $D_{-\infty}$ value is 1, which is what the horizontal line marks. The Local Norm curve is Local Norm with a global normalization applied. Again, all logarithms are base e.

fully chaotic region. As described in appendix B, the probability distribution associated with the logistic map at that parameter value is known, and that distribution is a smooth distribution that fills the entire unit interval (knowing the probability distribution, we go on to calculate the multifractal spectrum). This test is qualitatively different from our Cantor set tests. In figure 2.4, the basic EBA converges to the correct value as $N \to \infty$, but both of our normalized modifications to the EBA converge to a value that is slightly too high. We believe this is because for this set, there are two highly important points, on either edge of the unit interval. Our averaging methods mix the occupations of those points in with the occupations of the cells on the interior of the set, which causes the final estimate of the dimension to be slightly incorrect.

FIGURE 2.4. Convergence curves for the various improvements that we have developed to the EBA, when applied to the logistic map at $\lambda = 4$. The exact result for $D_{-\infty}$ is one, and is calculated in appendix B.

Figure 2.5 shows the methods applied to the Hénon map [24], which is a two dimensional map. This map is included as a worst-case test, since the dimension of the Hénon map is notoriously difficult to compute. We have no exact result to compare to, and the different methods converge to different values, so it is difficult to draw meaningful conclusions from this plot. However, given that D_2 has been measured at $D_2 = 1.25 \pm .02$ by [18] and D_0 has been measured at $D_0 = 1.261 \pm .003$ by [19], we would expect $D_{-\infty}$ to be larger than either of those values. All the methods are above that lower limit (although the naive method has definitely not converged), but we could tentatively hypothesize that the Local Norm method might be producing a more accurate estimate of the true dimension than the other two methods.
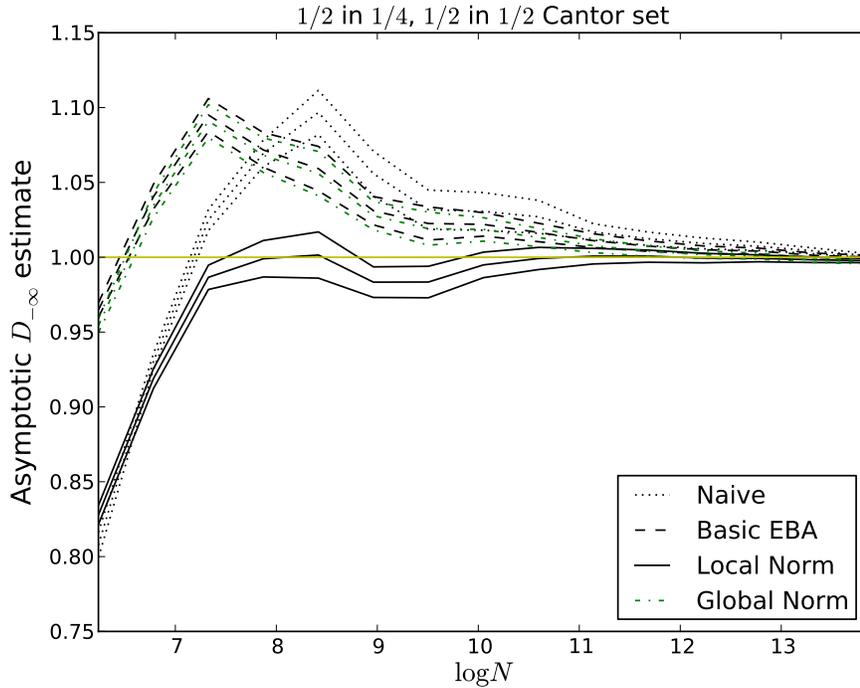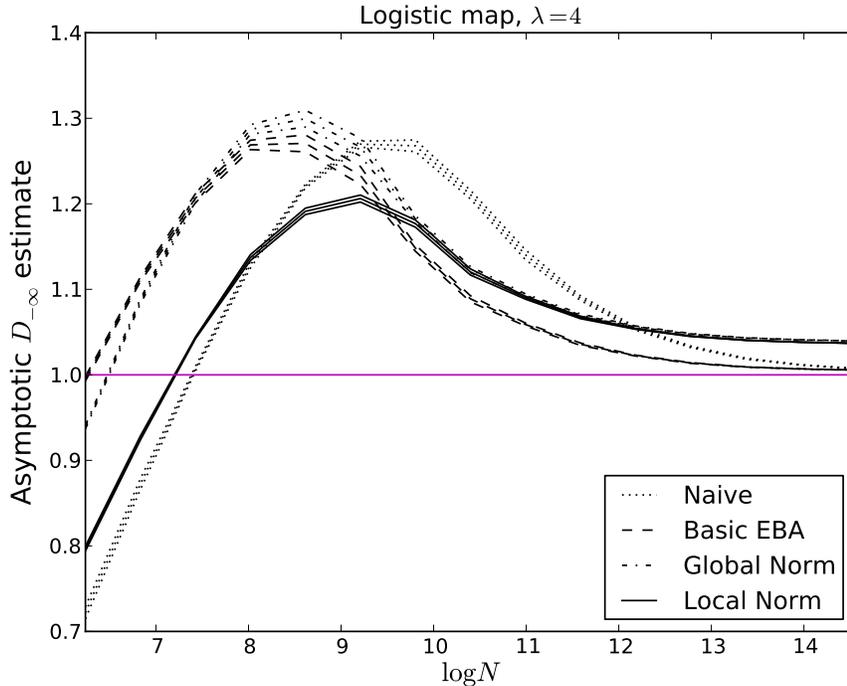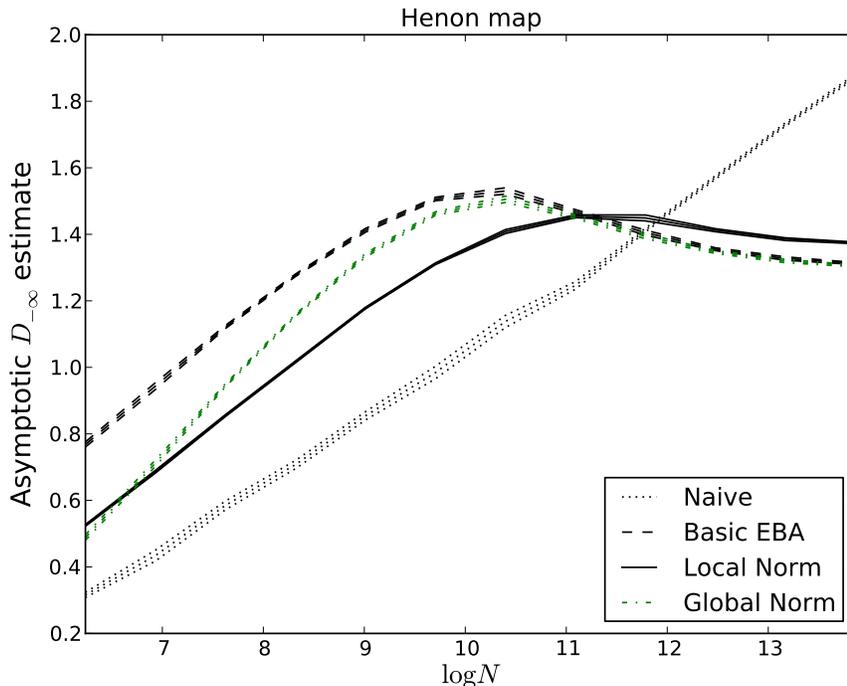
FIGURE 2.5. Convergence curves for the various improvements that we have developed to the EBA, when applied to the two-dimensional Hénon map. The exact $D_{-\infty}$ value is not known, but we can assume that it is near the value to which our methods seem to be saturating. For comparison, the Hausdorff dimension D_0 has been measured as $D_0 = 1.261 \pm 0.003$ by [19]. Note that the naive method, in more than one dimension, does not seem to be converging at all.

For all our test cases, the general trend seems to be that as N increases, the methods usually start too low, then overshoot before settling to a final value. Often, Local Norm has less overshoot than Global Norm or the original method, but converges to a larger value. This is not universal, though. For example, in our commensurate Cantor set, the Local Norm displays little if any overshoot, and converges to the correct value sooner than the other methods. To us, this shows the potential for further modifications of the Local Norm idea to display further improvements.

One of the important deciding factors controlling how much of a change our modifications make (especially the Local Norm) is whether the attractor is a rapidly varying probability

distribution on a smooth manifold or a set of disconnected fractal islands. This is because they deal with edges differently. The Local Norm version of the EBA can be interpreted as magnifying edges, by producing n^* values on the edges of the attractor that are closer to the n^* values on the interior of the attractor. The basic EBA and Global Norm EBA, on the other hand, produce n^* values in the interior that are slightly higher than the n^* values on the edges of the set. Neither option is universally better; some attractors are converging towards having sharply defined edges and some converge to have smoothly varying probabilities on their edges. However, regardless of which edge treatment is superior for a given set, the nature of the edges controls how much of a difference there will be in the $D_{-\infty}$ estimates for the various methods.

However, the convergence of $D_{-\infty}$ is certainly not the only metric by which to measure whether a method is an improvement or not. The main reason that we developed the Global Norm method (and then extended that idea to the Local Norm method) was to deal with the normalization spike at $q = 1$. Both the Global Norm method and the Local Norm method remove that spike entirely (remember that the Local Norm method has a global normalization applied after the local normalization). It is pointless to build a convergence plot for the removal of this spike, since it is not a continuous process; the method either removes the spike or does not. The rates of convergence of $D_{-\infty}$ are secondary to this more important goal. Both the modifications presented in this chapter remove that spike.

CHAPTER 3

a-NORM

3.1. DESCRIPTION OF METHOD

Consider forming a new version of n^* that averages neighbors using the weighting n^a, rather than n. This norm is usually called the p-norm or the q-norm. The p-norm of a list of numbers x_i is given by $(\sum_i x_i^p)^{1/p}$. For obvious reasons, we must use a slightly different name. We will use the variable a for the power, and call the new norm of the probabilities the a-norm.

The rationale for this method is that it provides another way of generating n^* values that vary more smoothly. The new parameter a can be used to select how strongly the values of the neighboring cells will influence the value of n^* for a particular cell. When $a = 0$, the a-norm will weight all cells equally, merely counting how many are occupied at all. As a becomes very large, the a-norm is more and more dominated by the largest value under consideration. Changing a from 0 to ∞ selects behavior anywhere within the range of those two extremes.

On closer examination, there are several ways of implementing this general idea, and the clearest strategies are closely analogous to the modifications we introduced to the EBA in chapter 2. We could take the sum of the powers, raised to the reciprocal power:

$$n_a^* = \left(\sum_{neighbors} n^a \right)^{1/a}.$$

This is analogous to the standard EBA (in fact, note that $a = 1$ exactly reproduces the basic EBA), and since it is the simplest version of this method, it will be referred to as the bare method. It does not preserve normalization, which, as described in section 2.1, leads

to a spike to $\pm\infty$ at $q = 1$. In order to preserve normalization, and thus remove the spike, we could renormalize after taking the sum. This would be analogous to the Global Norm method.

$$n_{GNa}^* = \frac{\left(\sum_{neighbors} n^a\right)^{1/a}}{\sum n^*}$$

Notice that once again $a = 1$ reproduces the EBA, but in this case $a = 1$ reproduces the Global Norm modification to the EBA that we developed. To be closest to the original definition of a norm, not just a sum, we could also take an average of the powers of the nonzero neighbors, so that locally our new n^* is a mean of some kind, not a sum. This would be:

$$n_{LNa}^* = \left(\frac{\sum_{neighbors} n^a}{N_{neighbors}}\right)^{1/a}.$$

The local normalization factor $N_{neighbors}$ is the number of occupied neighbors. This is analogous to our Local Norm method, and reduces to the Local Norm EBA at $a = 1$. Just like with our original Local Norm method, this method does not maintain normalization exactly at $q = 1$, so we apply a global normalization after the local normalization is computed in all of our analyses.

Choosing $a = 1$ is the standard EBA, or perhaps one of the variations we already investigated, but choosing $a = 2$ would be using the root mean square instead of the average to find n^* (or, thinking of the list of neighbors as the components of a vector, $a = 2$ would take the magnitude of that vector for all but the Local Norm variation of the a-norm). Taking $a = 3$ has no convenient name, but it would be taking the cube root of the mean or sum of the cubes of the occupations. In a similar manner to how the multifractal spectrum considered all moments of the distribution (even noninteger ones) to extend the definition of dimension, we are considering all moments of the local probability distribution (although

negative values of a will not be considered). These different moments characterize different aspects of the local probability distribution. For example, the variance of the local collection of probabilities would be given by $\langle p^2 \rangle - \langle p \rangle^2 = (p^*_{a=2})^2 - (p^*_{a=1})^2$ (with the averages over the group of neighbors). Perhaps the variance or some other indication of how scattered the local group of p values are would be more useful in characterizing the attractor. Different moments of the local distribution are related to p^*_a for different values of a.

3.2. Benefits and drawbacks of the Local Norm version

The Local Norm version of the a-norm method has many similarities to the Local Norm version of the EBA, first described in section 2.4.

The main theoretical benefit of the Local Norm EBA is that it introduces possible values for n^* that are more closely spaced, which allows the potential for much smoother variation in n, making it conceivable that the clipping errors described in section 1.4 due to discretization issues could be eliminated. However, the Local Norm EBA does not always fully fill in those possible n^* values, so that the method is not an improvement over other techniques in all cases.

The same issues affect the Local Norm version of a-norm. Imagine applying the different versions of the a-norm method to a sparse region. In that region, we expect all the cells to have exactly the same occupation: one point per cell. In that case, the Local Norm version of this method would give:

$$n^*_{LNa} = \left(\frac{\sum_{neighbors} 1^a}{N_{neighbors}} \right)^{1/a} = \left(\frac{N_{neighbors}}{N_{neighbors}} \right)^{1/a} = 1.$$

That is to say, Local Norm would take a type of average of the cells it sampled (disregarding the zeros), and the average of a group of ones is still one. This means that in sparse regions,

this method would make only minimal changes away from the values used by the naive method. In contrast, the bare a-norm and the Global Norm versions applied to this same group of singly occupied cells would give $n^* = N_{neighbors}^{1/a}$, so that sparser regions would have n^* values that change when the density of the region changes. This would give a smoother variation over the sparse areas of the attractor, which is exactly the behavior we seek. This is the same basic phenomenon as the one described in section 2.4; although the Local Norm version has the potential to produce n^* values that are much more finely spaced, it may not change the occupations in sparse regions so as to actually populate all these finely spaced values.

For this reason, we expect the Local Norm version of the a-norm method to perform in a manner similar to the Local Norm version of the EBA (especially in the $a = 1$ case, which exactly reproduces the EBA). Performance of the method is discussed in section 3.4.

3.3. Sup Norm

As a gets large, notice that the largest one of the neighbors will dominate the local sum more and more, regardless of the particular sub-type of a-norm being used (Global Norm, Local Norm, or the bare method). For this reason, we also investigated using the sup norm of the neighbors to build n^*.

$$\lim_{a \to \infty} n_a^* = n_{sup}^* = \max_{neighbors} n$$

This is known as the sup norm because the more extendable analog to this would be to use the supremum: $n_{sup}^* = \sup_{neighbors} n$. For our analyses, this subtle refinement is unnecessary.

Examining the bare, Global Norm, and Local Norm versions of the a-norm, notice that the different versions will converge to slightly different versions of the sup norm. The bare

a-norm will simply converge to n^*_{sup} as shown above. The Global Norm version will converge to

$$\lim_{a\to\infty} n^*_{GNa} = \lim_{a\to\infty} \frac{\left(\sum_{neighbors} n^a\right)^{1/a}}{\sum n^*} = \frac{\max_{neighbors} n}{\sum n^*} = \frac{n^*_{sup}}{\sum n^*_{sup}}.$$

The Local Norm a-norm will converge to

$$\lim_{a\to\infty} n^*_{LNa} = \lim_{a\to\infty} \left(\frac{\sum_{neighbors} n^a}{N_{neighbors}}\right)^{1/a} = \frac{\max_{neighbors} n}{\lim_{a\to\infty} n^{1/a}_{neighbors}} = \max_{neighbors} n = n^*_{sup},$$

but remember that we apply a global normalization after the local normalization, in order to remove the spike at $q = 1$. Thus the Local Norm a-norm will converge to a normalized sup norm $n^*_{sup}/\sum n^*_{sup}$.

So the bare a-norm converges to the bare sup norm as $a \to \infty$, but the Global Norm a-norm and Local Norm a-norm both converge to a renormalized version of the sup norm. We have disregarded degeneracy here, even though in sparse regions it is likely that the numbers being considered will be a group of ones. Even if there is degeneracy, our results here still hold. All that would change would be an extra factor of $N^{1/a}_{terms}$, which becomes irrelevant as $a \to \infty$.

Since the sup norm controls the $a \to \infty$ limiting behavior of all three types of the a-norm method, we used the sup norm curves as limiting curves when testing a dependence (with appropriate normalization). We found that in general changing a smoothly varied the behavior of the method. Thus, many of our plots show the smallest a tested, a medium a value, and the sup norm to replace $a \to \infty$.

3.4. PERFORMANCE

As mentioned when introducing the assortment of a-norm methods, there are special values of a that reproduce various existing methods. The question then becomes whether changing a away from those values improves convergence. To test this, we applied the a-norm, at various limiting values of a (with the sup norm replacing $a \to \infty$) as well as appropriate intermediate a values, to various test sets. We found that the a dependence was generally quite simple. Curves would smoothly change between the limiting cases as a was varied, usually monotonically.

Regardless of the particular value of a, each variation of the method requires computer time on the same order of magnitude as the EBA improvements of the previous chapter. Our most filled data sets ($\sim 10^6$ points in a grid of side 2^{11}) require on the order of a minute to compute a single D_q curve. The time scales with the number of occupied cells in the histogram, in an approximately linear manner. Using the sup norm gives a small reduction in processing time, but since the sup norm still requires a search through all the neighbors, this reduction does not reduce the time below its approximately linear scaling behavior.

There were two broad areas of parameter space to investigate: $0 < a < 1$ and $a > 1$. Preliminary investigations showed that very small a tended to reproduce results similar to the naive results. Since changing a tended to monotonically change the behavior of the convergence curve, this means that $a < 1$ tended to fill in the gap between the EBA and the naive method. Often, we found that the correct behavior was not between the EBA and the naive method. Both tended to overshoot the correct result or undershoot the correct result together. Picking a between 0 and 1 would not generate improved results. Thus, we limit our graphs here to $a > 1$.

The general trend that we observe most consistently is that the Local Norm version of a-norm produces the best convergence, and in general larger a values tend to converge the best in any version of the method investigated. This suggests that using a globally normalized sup norm to compute n^* could be an improvement. As with the methods in chapter 2, another major benefit not captured in our convergence plots is that the normalized versions of the a-norm (including the normalized sup norm) all remove the spike in D_q at $q = 1$.

For consistency, we tested each version of the method on our commensurate Cantor set, where all fractal boundaries happen to lie on histogram cell boundaries; on the logistic map at $\lambda = 4$, to test a known distribution without the fractal structure that may influence some results when using the Cantor set (see appendix B for a calculation of D_q for that set); and on the Hénon map, whose dimension is notoriously difficult to calculate, in order to check a worst-case scenario.

Figures 3.1, 3.2, and 3.3 show the bare a-norm applied to a Cantor set, the logistic map, and the Hénon map, respectively. We see in most cases that higher a values converge faster than the EBA. The exception to that trend is for the logistic map at $\lambda = 4$ (figure 3.2), for which the naive method performs the best, and as a increases the asymptotic result moves away from the correct result. We believe this is because (as discussed in appendix B) the probability distribution associated with the logistic map at $\lambda = 4$ has a singularity at each end. Our various averaging methods will mix the highly occupied cells on the edges of the attractor with the less occupied interior cells, which causes incorrect calculations of the dimension spectrum. The naive method makes no such mixture, which generates the correct behavior in the $N \to \infty$ limit. As might be expected, the commensurate Cantor set (figure 3.1) displays the best convergence, with the sup norm (unnormalized, to match the $a \to \infty$ limit of the bare a-norm) displaying almost no overshoot, and converging very

quickly to the correct result. This set is selected to provide at least one test where the set has a highly regular structure that is easily analyzed and simple enough that there are no unexpected issues affecting the convergence. For the worst-case test, the Hénon map (figure 3.3), it is unclear what the exact result is. The box counting dimension D_0 has been measured as $D_0 = 1.261 \pm 0.003$ by [19], which places our EBA asymptotic result in some doubt, since we would expect $D_{-\infty}$ to be larger than D_0 by a larger amount than the EBA shows. It seems likely that the a-norm, particularly the sup norm version shown here, is closer to the true $D_{-\infty}$ value.

FIGURE 3.1. Convergence curves for the bare a-norm method (no normalization applied), applied to a highly regular Cantor set. Recall that $a = 1$ reproduces the EBA, and that the sup norm is the $a \to \infty$ limit.

We then applied the same tests to the Global Norm version of the a-norm. Figures 3.4, 3.5, and 3.6 show the Global Norm a-norm applied to a Cantor set, the logistic map, and the Hénon map, respectively. Results were similar to the results described for the bare a-norm

FIGURE 3.2. Convergence curves for the bare a-norm method (no normalization applied), applied to the logistic map at $\lambda = 4$. Recall that $a = 1$ reproduces the EBA, and that the sup norm is the $a \to \infty$ limit.

method. The one notable exception is the logistic map. For the bare a-norm, the $a \to \infty$ moved the asymptotic value of $D_{-\infty}$ away from the exact result, so that increasing a made the convergence worse. For the Global Norm a-norm in figure 3.5, the $a \to \infty$ limit moves the asymptotic value closer to the exact value. This means that for the Global Norm a-norm, in every case shown, taking $a \to \infty$ seems to improve convergence.

Finally, we applied the same set of tests to our Local Norm version of a-norm. Figures 3.7, 3.8, and 3.9 show the Local Norm a-norm applied to a Cantor set, the logistic map, and the Hénon map, respectively. Again the results are qualitatively similar to those previously described. The only changes seem to be small improvements in all the convergence rates involved. For finite a, the Local Norm method consistently shows less initial overshoot, converges sooner, and is closer to the exact values than the other versions of the a-norm. Of
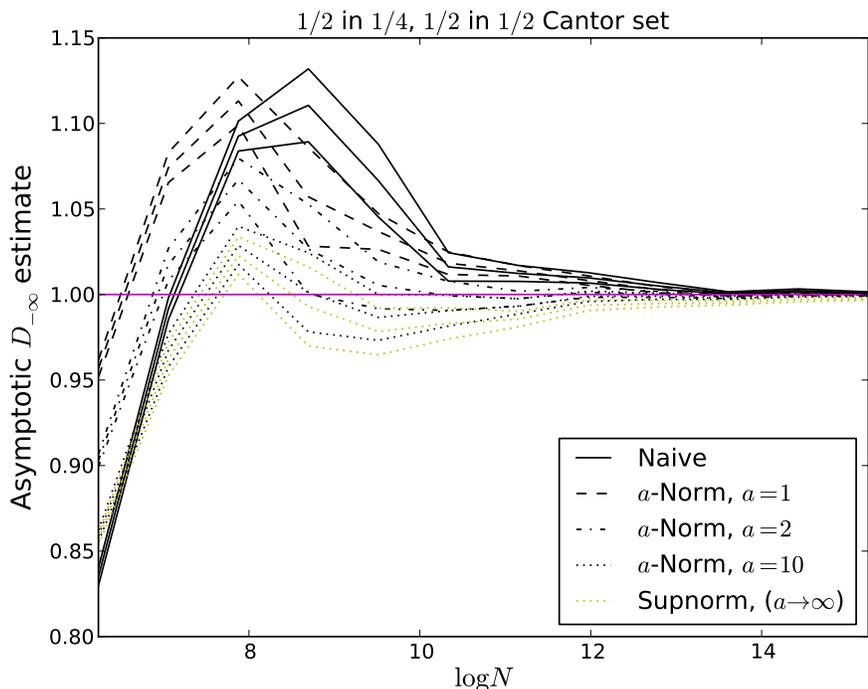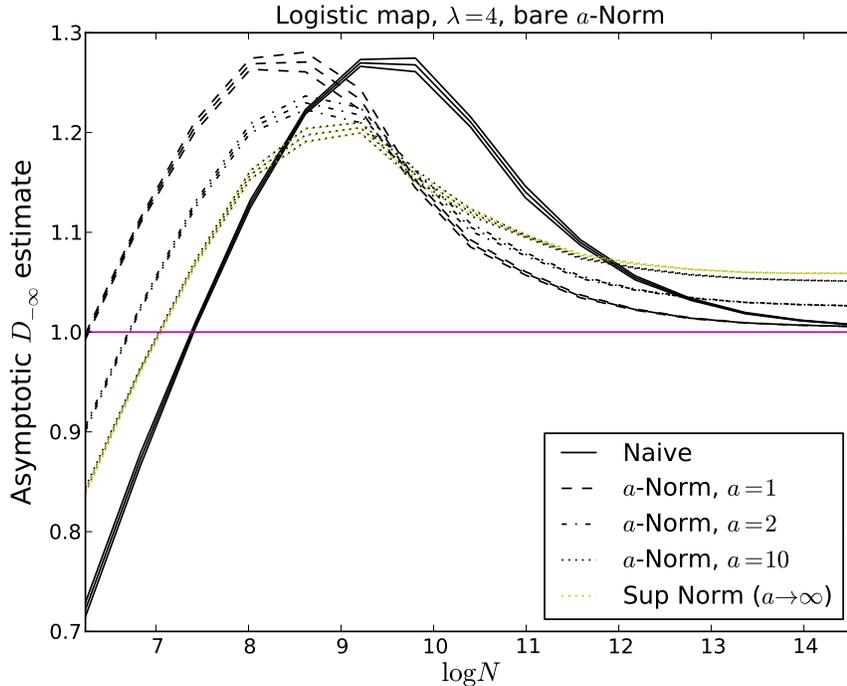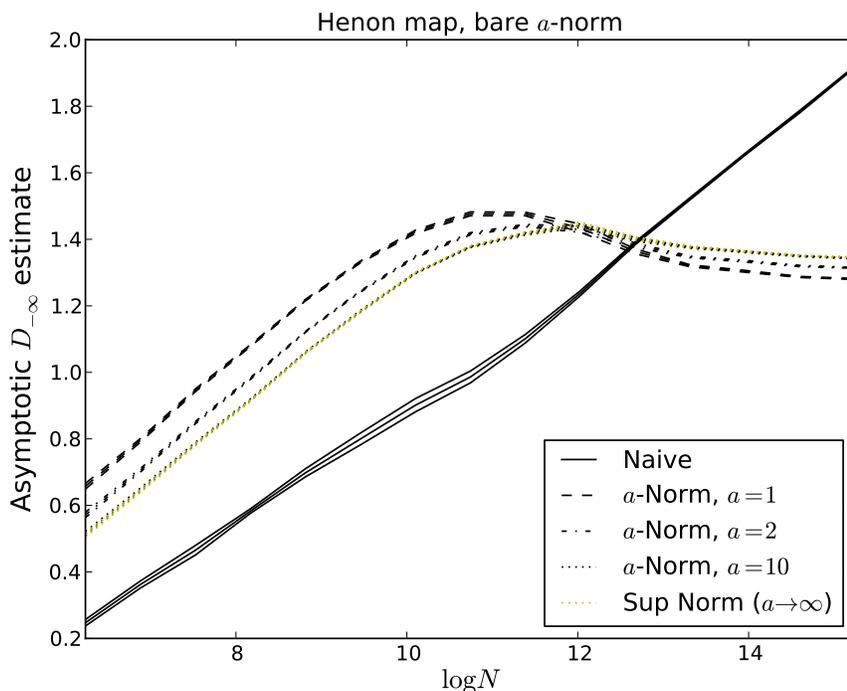
FIGURE 3.3. Convergence curves for the bare a-norm method (no normalization applied), applied to the Hénon map. Recall that $a = 1$ reproduces the EBA, and that the sup norm is the $a \to \infty$ limit.

course, the general trend is again that increasing a seems to increase convergence, and in the $a \to \infty$ limit, we have shown that the Local Norm a-norm is equivalent to the Global Norm a-norm.

These results suggest an improvement to the calculation of D_q whose simplicity begins to approach the simplicity and elegance of the basic EBA. Instead of using the actual value in any given cell, replace each cell with the maximum value found in the set of cells in a small neighborhood of that cell. Then apply a renormalization. This improvement has been shown to converge slightly faster than the EBA in some situations, and we believe that the increased convergence rate would continue to improve when applied in higher and higher dimensions, on more complex attractors.
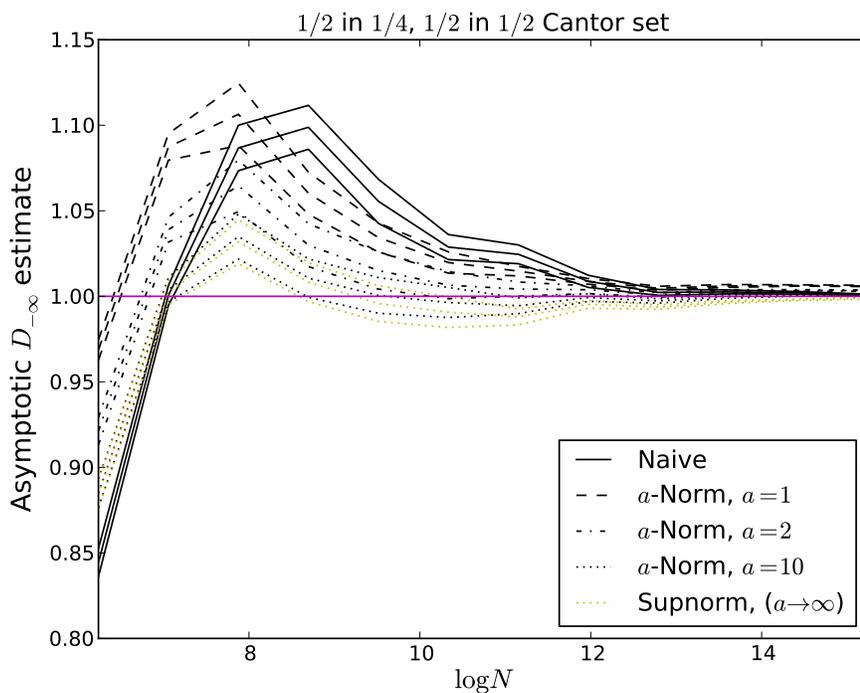
FIGURE 3.4. Convergence curves for the Global Norm a-norm method, applied to a highly regular Cantor set. Recall that $a = 1$ reproduces the Global Norm EBA, and that the sup norm is the $a \to \infty$ limit. The sup norm result plotted here has a global normalization applied, to match the Global Norm $a \to \infty$ limit.
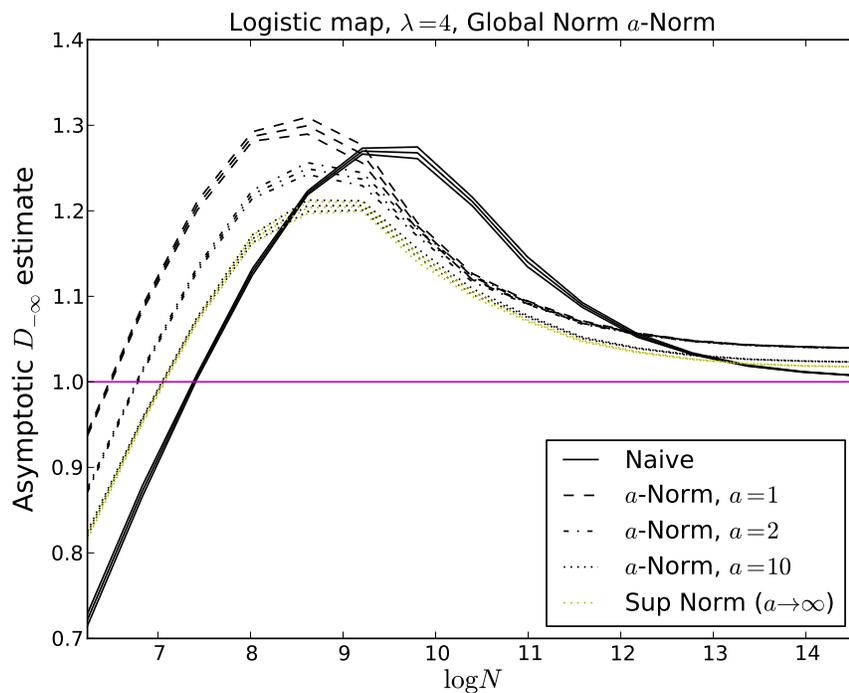
FIGURE 3.5. Convergence curves for the Global Norm a-norm method, applied to the logistic map at $\lambda = 4$. Recall that $a = 1$ reproduces the Global Norm EBA, and that the sup norm is the $a \to \infty$ limit. The sup norm result plotted here has a global normalization applied, to match the Global Norm $a \to \infty$ limit.
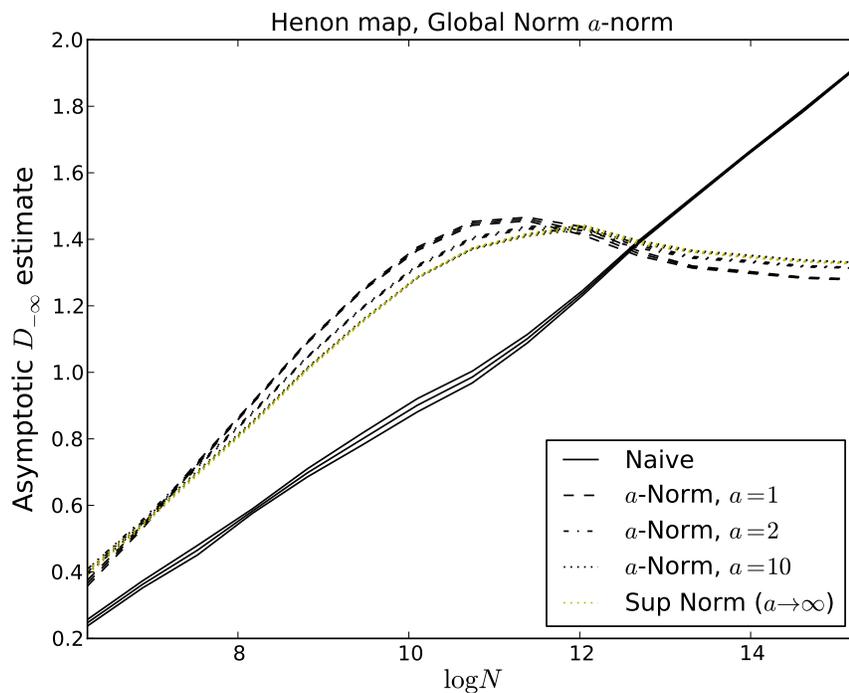
FIGURE 3.6. Convergence curves for the Global Norm a-norm method, applied to the Hénon map. Recall that $a = 1$ reproduces the Global Norm EBA, and that the sup norm is the $a \to \infty$ limit. The sup norm result plotted here has a global normalization applied, to match the Global Norm $a \to \infty$ limit.
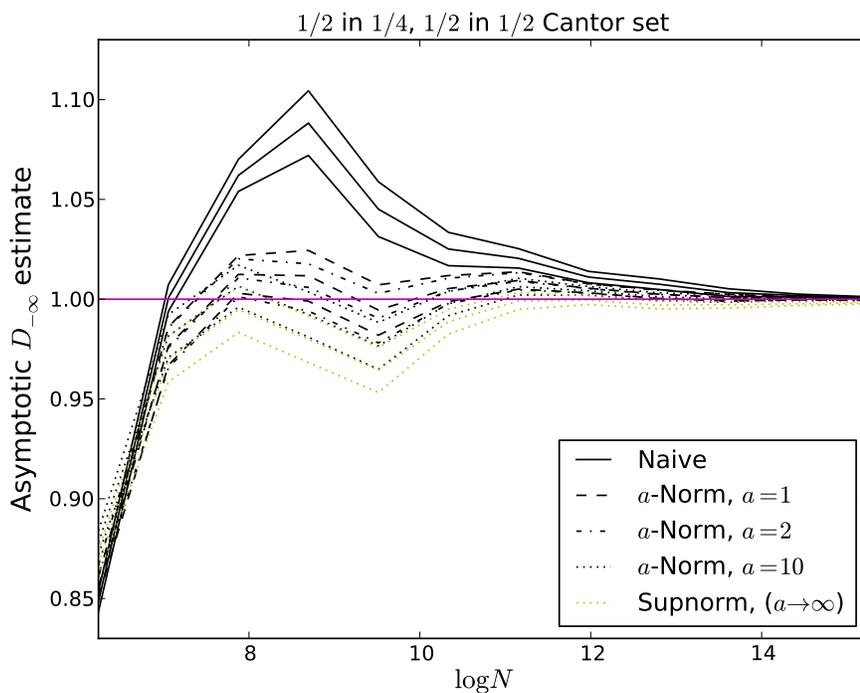
FIGURE 3.7. Convergence curves for the Local Norm a-norm method, applied to a highly regular Cantor set. Recall that $a = 1$ reproduces the Local Norm EBA, and that the sup norm is the $a \to \infty$ limit. The sup norm result plotted here has a global normalization applied, to match the Local Norm $a \to \infty$ limit normalization.
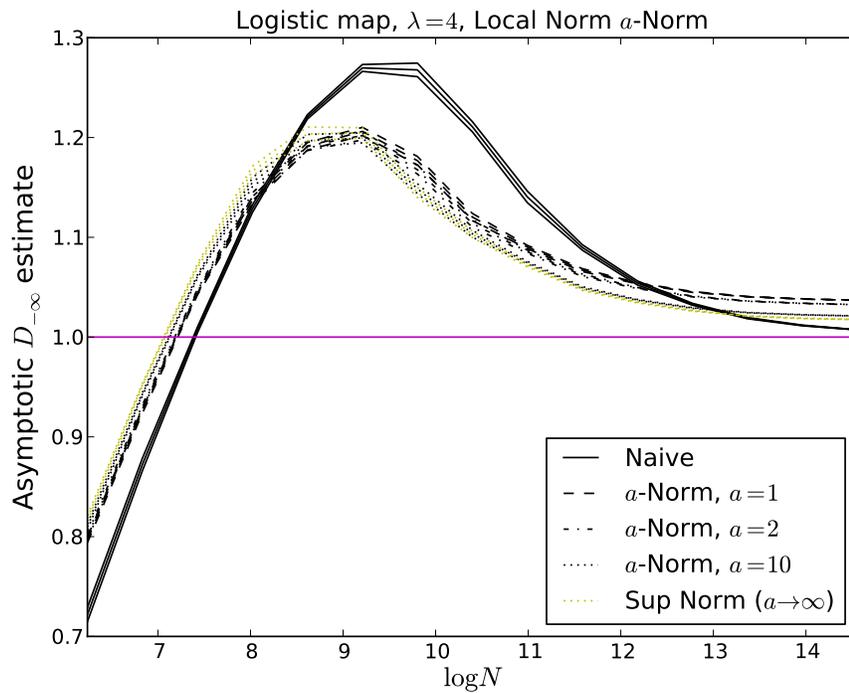
FIGURE 3.8. Convergence curves for the Local Norm a-norm method, applied to the logistic map at $\lambda = 4$. Recall that $a = 1$ reproduces the Local Norm EBA, and that the sup norm is the $a \to \infty$ limit. The sup norm result plotted here has a global normalization applied, to match the Local Norm $a \to \infty$ limit normalization.
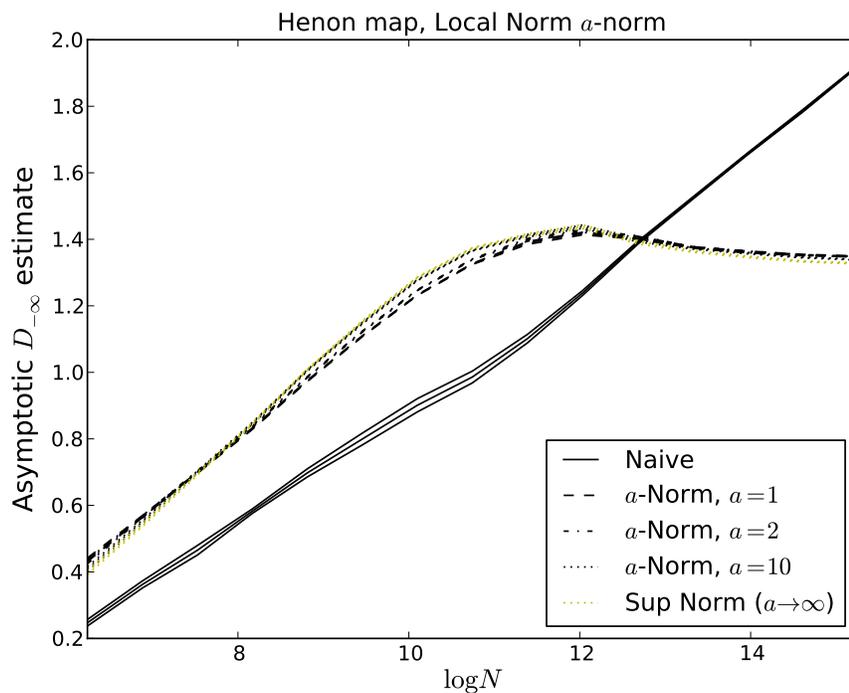
FIGURE 3.9. Convergence curves for the Local Norm a-norm method, applied to the Hénon map. Recall that $a = 1$ reproduces the EBA, and that the sup norm is the $a \to \infty$ limit. The sup norm result has a global normalization applied to match the Local Norm $a \to \infty$ limit normalization.

CHAPTER 4

# Variable Box Size Algorithm

4.1. Description of method

This method has some similarities to the "fixed mass" approach described in [14], so we will briefly outline that method here before describing the Variable Box size Algorithm (VBA). For the fixed mass approach, instead of picking a uniform grid size ℓ and measuring p within each cell, a uniform amount of probability p is chosen, and cells are expanded until they enclose that amount of probability, so that ℓ is different for every cell. Recall that $\Gamma = \sum p^q / \ell^\tau$ should be finite along the correct $\tau(q)$ line. In our implementation of the naive algorithm and the EBA, we generate several histograms with ℓ values that are uniform across the histogram but different from each other, and plot $\ln \sum p^q$ versus $\ln \ell$ for fixed q. The slope of that plot will be τ, since finite Γ would mean $\sum p^q \sim \ell^\tau$. For the fixed mass algorithm, p is uniform instead of ℓ, so we fit a slope to $\sum \ell^{-\tau} \sim p^{-q}$, giving $q(\tau)$ instead of $\tau(q)$. There are many subtleties and refinements to the method, which are well explored in [14]. For example, the boxes formed in this manner should not overlap, since overlapping leads to counting errors. Also, it turns out to be more effective to find the $f(\alpha)$ spectrum instead of the D_q curve when using fixed-mass methods.

Now to the VBA. The VBA borrows some methodology from the fixed mass approach, but is very different. As with the fixed mass approach, we expand each occupied cell to enclose a chosen amount of probability, but rather than measuring ℓ for each of those expanded cells and treating ℓ as our primary variable, we use this expanded box to form a more precise estimate of the probability in the original box. This new estimate of the probability becomes a new type of p^*, much like the EBA.

More precisely, the procedure this algorithm uses is the following. First, a target value for n is chosen, which we call n_t. The size of each box will be chosen so that no box has fewer than n_t points in it. The box is expanded symmetrically by considering additional layers of neighbors until the number of points in the expanded box n^* is greater than or equal to n_t. Our estimate of the occupation of box i will then be the average (including zeros) of the occupations of the cells thus enclosed, specifically the number of points enclosed in that expanded box divided by the number of boxes enclosed (the volume of the expanded box): $\mu_i = n_i^*/v_i^*$. In order to ensure that the estimate obtained in this manner is still local to box i, and not measuring characteristics of distant areas of the attractor, we also introduce a cap on the size of the box. No box is allowed to expand to have a side larger than ℓ_c. The two parameters n_t and ℓ_c are the controls which fine-tune the behavior of the algorithm. If the two conflict (if n_t is not satisfied but the box has reached a size ℓ_c) then the cap on ℓ takes precedence.

A quick note on notation: it is far easier to consider the number of expansions than to measure the actual expanded length. For that reason, I will use $\tilde{\ell}$ as the number of expansions. If a cell of initial size ℓ_i has been expanded $\tilde{\ell}$ times, then its actual expanded length is $\ell_e = (2\tilde{\ell} + 1)\ell_i$. Similarly, I will often specify the length cap in terms of a number of expansions $\tilde{\ell}_c$ instead of an actual length cap ℓ_c. Note that $\tilde{\ell} = 0$ means zero expansions, not zero volume. Specifying $\tilde{\ell}_c$ is quite different from specifying ℓ_c, and that difference is explored in more detail in section 4.3.

The idea of this method is that in very sparse regions, the VBA can give a more continuous and accurate estimate for just how sparse the region is. As an example to illustrate this, imagine a very sparse collection of N points distributed along a line, where each occupied cell contains only a single point, and there is one empty cell between each pair of populated cells.

TABLE 4.1. A demonstration of how the VBA and EBA would treat sparse regions differently. There is an infinite field of zeros past the edges of the histogram. Note that the VBA produces different occupation values for the two regions of different density, since it is sensitive to the spacing between cells.

Original cell occupations	1	0	1	0	1	0	1
EBA n^*	1	0	1	0	1	0	1
VBA n^* with $n_t = 2$	2/5	0	3/5	0	3/5	0	2/5
Original cell occupations	1	0	0	1	0	0	1
EBA n^*	1	0	0	1	0	0	1
VBA n^* with $n_t = 2$	2/7	0	0	3/7	0	0	2/7

Compare this collection to another that again has N singly occupied cells, but 2 empty cells between each pair of populated cells. In both cases, the basic EBA will detect no neighbors, and each n_i^* would be 1. The new method will be able to detect that spacing between cells, and would report different numbers. If we pick $n_t = 2$ and $\tilde{\ell}_c = 10$, the VBA will expand each cell until it encompasses its two adjacent neighbors (one on either side), so that each n_i^* will be 3. In the case where there was one empty cell between occupied cells, the expanded box would have to enclose 5 cells (we count empty neighbors), giving each $\mu_i = 3/5$. When the gap is 2 cells, the expanded box must enclose 7 cells, giving $\mu_i = 3/7$. See table 4.1 for an illustration of these situations. The VBA thus produces a more finely-grained estimation of the density of the region, so that the discretization problems outlined in section 1.4 can be overcome with smaller amounts of data than with the EBA.

The tradeoff here is that in order to get a smoother estimate of the probability, we need to expand more and more, so that our estimate is less local. This is why we introduced ℓ_c to limit the size of the expanded box. Note that if the parameters are adjusted such that most cells grow until they reach the size limit, then most cells will have the same size, so that the new algorithm should mimic the behavior of the EBA, which had a fixed box size everywhere. In fact, if ℓ_c is very small, our estimates stay local, but encounter the same

discretization problems as most of the pre-existing methods. To be precise, if $\tilde{\ell}_c = 0$ we exactly reproduce the naive method, and if $\tilde{\ell}_c = 1$ and n_t is larger than the occupation of any single cell, then we reproduce the EBA.

In almost exactly the same manner as the EBA, this method does not preserve normalization. However, this can be remedied by applying a global normalization, just like our modifications to the EBA. This global normalization is implied for the rest of this chapter.

4.2. VARIABLE OR FIXED n_t

There are two parameters controlling the behavior of the VBA: n_t and ℓ_c. Investigating the convergence of the method as the total number of points N increases, it is natural to wonder what value to specify for n_t. It seems reasonable that n_t should scale somehow with N. The argument is that perhaps there is a fixed fraction of boxes that should expand, and this fraction should be independent of N. This argument expresses the idea that the need to expand a box is based on what area of the attractor it is in. Sparse areas will always be less occupied than dense areas, and so our estimates of the probability in those areas will always be less trustworthy than our estimates in the dense areas. Adding ten times as many points to the histogram will change the numbers in all the cells, but will not change what fraction of the cells are located in the sparse regions. Accordingly, one of the methods that we tested was specifying n_t as a fraction of the occupation of the most occupied cell in the histogram. We called the fraction η, so that if the highest occupation was n_{max}, then $n_t = \eta n_{max}$.

There are arguments against the use of the scheme involving this factor of η. What we really need is for the occupation numbers in the sparsest regions of the attractor to vary smoothly. The fractional variations between the most and second-most sparse areas of the attractor should look like $((n_{min} + 1) - n_{min})/n_{min} = 1/n_{min}$. For our sum to vary smoothly,

there could be an absolute size goal for this variation, regardless of the total number of points in the attractor. Using $\mu = n_e/v_e$ should give a collection of terms where n_e is just barely above n_t. For those terms, the variation between terms will be approximately $1/n_t$. Thus, arguments could be made that we should specify this number, and that if n_{max} is very large, or if picking this number means that we expand a small fraction of the cells, then this is better, because this means that we are able to have smooth variation and still generate our samples from a very local part of the attractor. Another way of describing this is that the trustworthiness of p for a cell is not based on its position in the attractor, it is solely based on the occupation of that cell. If we were to add ten times as many points to the attractor, we would make many of the sparse regions populated enough that they no longer needed to be expanded as much.

Regardless of the finer points of the behavior of a fixed or variable n_t during normal operation of the VBA, notice the more practical effect that the choice has on the convergence: for a fixed n_t, eventually as the total number of points N grows, for almost all systems, there will be some value of N beyond which even the least occupied cell has an occupation $n_{min} > n_t$. When that point is reached, n_t is met for every single cell without any expansions, so that each n_i^* will be immediately given by $n_i/1$. This means that as N increases, if n_t is a fixed value, eventually the VBA will begin to reproduce the results of the naive method. If instead of a fixed value for n_t, a fixed value of η is chosen, the method will continue to attempt to expand cells even as $N \to \infty$, although the expansions will scale with N, so that eventually a steady state will be reached, where adding more points does not change which cells are expanded or how much they are expanded.

It is also important to note that whether n_t is chosen as a flat value or as a fraction of the most occupied cell with η, once n_t is chosen, the method is unchanged. These are merely

different ways of understanding and automating the choice of the parameter n_t. We tested the convergence of this method for both a fixed n_t and for n_t controlled by η. Results are shown in section 4.4. In general, the computer time for these calculations is similar to the time for other methods: on the order of a minute per D_q curve for our most filled sets on a laptop computer. As the parameters n_t and $\tilde{\ell}_c$ are increased, the time reaches roughly double that required for most of our other EBA modifications, such as the a-norm and the Local and Global Norm. Increasing n_t of $\tilde{\ell}_c$ increases the amount of searching for neighbors, but when both parameters are suitably small, their effect on calculation time is not drastic. The scaling of that computation time is very roughly given as $\sim N_{cells}^1 n_t^1 \tilde{\ell}_c^1$, although even the scaling here is very approximate.

4.3. Variable or fixed ℓ_c

We also investigated the difference between a fixed ℓ_c and a fixed $\tilde{\ell}_c$. Much like using η to scale n_t with the occupation, using the two different options for ℓ_c will control whether the expanding cell has a size that scales with the grid, or remains an actual fixed spatial size. If we set $\tilde{\ell}_c$, then we have chosen the number of expansions that are allowed. This means that the largest expansion allowed will be $(2\tilde{\ell}_c + 1)\ell$, so that as we change the cell size ℓ, the maximum size of the expanded cell will also change. If we instead fix ℓ_c as some fraction of the size of the histogram, then the actual absolute size of the largest expanded cell will be quite close to constant even as we change ℓ. However, this means that at different length scales ℓ, we will expand that maximally expanded cell to include a different number of neighbors.

This investigation is mainly for the sake of symmetry. If the set being examined is truly a fractal, we would expect that there is no single best scale to investigate. Instead, we would

expect self-similar or pseudo-self-similar structures to repeat at all length scales. Thus, the meaning of 'local' would change depending on the size of the grid covering the set. This would support the use of a fixed $\tilde{\ell}_c$ (a fixed number of expansions). However, because we are limited to a finite amount of data, and have an actual minimum grid size, we have been forced to pick a preferred length scale. There really is a minimum and maximum size to features on our fractal. That would support the use of a fixed ℓ_c (a fixed absolute size).

The fixed $\tilde{\ell}_c$ method is the main focus of the other sections of this chapter. Let us explore briefly the behavior of a fixed ℓ_c. What ends up happening is a switch between two types of behavior. At very small length scales, a fixed ℓ_c is equivalent to a very large $\tilde{\ell}_c$. The smaller length scales are also the length scales where the cell occupations are low, meaning that these are the scales at which the algorithm expands cells frequently. As we rebin to larger and larger length scales, eventually we generate $\ell > \ell_c$, at which point even a single expansion would make the expanded box larger than the size cap. So the fixed ℓ_c method starts out generating very large $\tilde{\ell}_c$, then switches to $\tilde{\ell}_c = 0$. This behavior is demonstrated in figure 4.1.

It may seem like considering a single patch of the attractor of fixed size would generate the same results at each length scale, since our estimate of the probability contained in that expanded box would ideally not change. However, we would have a number of terms in Γ that would scale in some complicated manner with the number of boxes, giving Γ that does in fact change with ℓ.

When we tested this version of the VBA, we did not find it to be an improvement. The larger ℓ_c is, the farther the spectrum is from the correct result. See figure 4.2 for a convergence plot of this variation of the VBA. We tested a set of fixed n_t values, mainly for practical reasons. From a computational standpoint, an ℓ_c that did not scale with ℓ
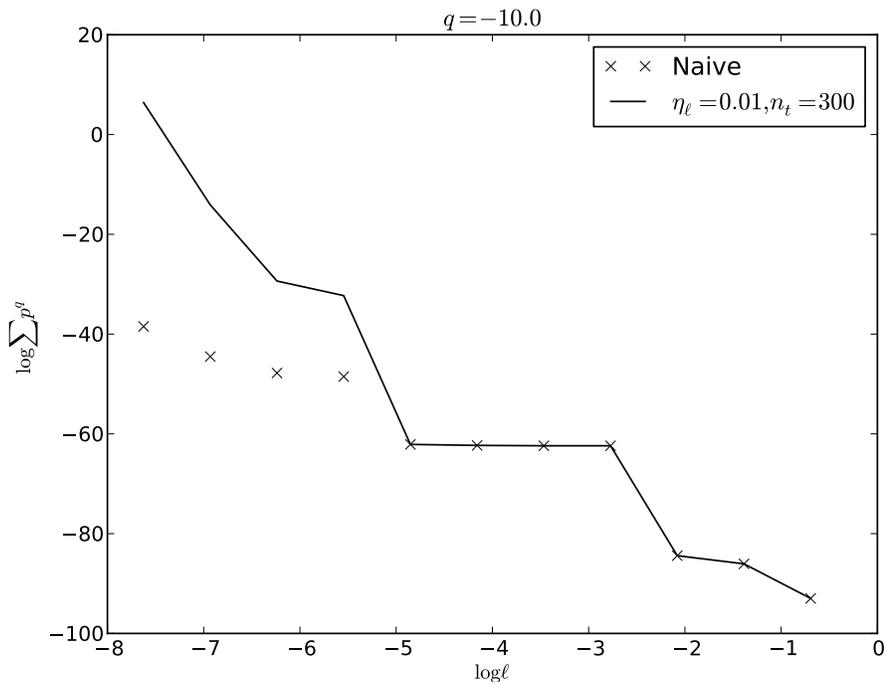
FIGURE 4.1. A plot of $\ln \sum p^q$ for a fixed q value, using a set ℓ_c instead of a set $\tilde{\ell}_c$. Recall that the slope of $\ln \sum p^q$ versus $\ln \ell$ is τ. Note that when ℓ passes the $\ell > \ell_c$ threshold, the naive method is exactly reproduced. (From a programming standpoint, specifying ℓ_c means specifying $\tilde{\ell}_c$ as a multiple of the grid size. That multiple is labeled as η_ℓ in the figure.)

required an annoyingly large amount of computation time, since at the smallest ℓ values, this would entail a large number of expansions, requiring a large number of searches through the histogram for neighboring cells. Furthermore, the most extended searches through the list of occupied cells are happening at the length scales for which the list is the longest (smaller cells mean that a larger number of cells exist). Specifying n_t in a way that scales with the occupation causes the number of these searches to grow startlingly large. The expanded boxes become large enough that the search begins to switch from being linear in the number of occupied cells to scaling (in a very rough sense) as N^2_{cells}, since for each cell, the algorithm must check a large continuous region to see how many cells it contains.
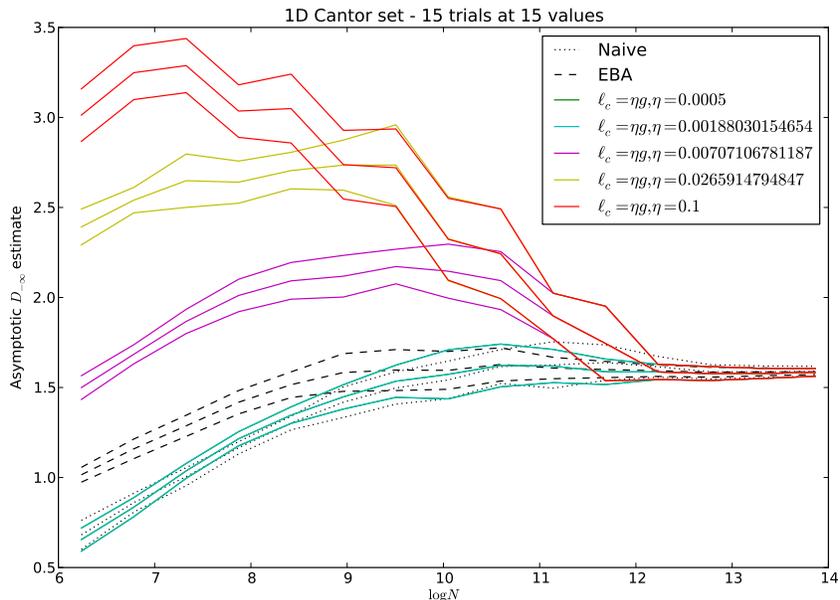
FIGURE 4.2. A convergence plot for the VBA, using a fixed ℓ_c rather than a fixed $\tilde{\ell}_c$. From a programming standpoint, specifying ℓ_c means specifying $\tilde{\ell}_c$ as a multiple of the grid size g (the number of cells per side). That multiple is labeled as η in the figure. The target occupation n_t was a fixed value for these curves.

Specifying a fixed n_t means that as N increases, fewer expansions are necessary, since the target occupation can be met with fewer adjacent cells.

The reasons why this variation are not an improvement will be more clear after our more detailed investigation of the VBA in general. The end result of that investigation turns out to be that there is a flaw at the core of the VBA, and it really works best when the ℓ limits force the VBA to reproduce versions of the EBA (obviously we were not aware of this flaw during our research of the behavior of the VBA). Using an ℓ_c that is fixed and does not scale with ℓ has the effect of removing the cap on ℓ for a large portion of the length scales under consideration, which causes the method to produce incorrect final results. Figure 4.2 is fairly generic. For various other areas of parameter space, the trend that a larger fixed ℓ_c produced worse results was very common.

4.4. Convergence of the VBA

When applying the convergence test of section 2.2 to the VBA, the trend that we see is that the algorithm works best when it expands all cells in the same manner. When n_t is lower than the occupation of the least occupied cell, then none of the cells expand, and the method reproduces the results of the naive method, where each cell is unmodified. If $\tilde{\ell}_c$ is small, and n_t is large, then all the cells are expanded until they hit the maximum size allowed, and this method becomes the EBA. Technically, it only reproduces the EBA exactly if $\tilde{\ell}_c$ is one, since the EBA only expanded cells to include the adjacent cells, and no farther. However, expanding universally by two or three cells tends to produce results similar to the basic EBA[1]. Between the no-expansion limit and the universal expansion limit, there is a region of parameter space where the cells are free to expand, all differently from each other, until they hit n_t without being limited by $\tilde{\ell}_c$. In this region, the method produces an entirely wrong value for D_q as $q \to -\infty$. See figure 4.3 for an example. We believe that it is the very variability of the box sizes that causes the algorithm to fail, for reasons that we will describe in the next section.

For a flat fixed n_t (as opposed to using η to scale n_t as the occupation changes), the algorithm jumps between methods. When N is very small, the fixed n_t is larger than the maximum occupation of any cell, so every cell expands, much like the EBA. Then as N increases, n_t is in the zone where the size freedom destroys the estimate of D_q for a while, and the estimate of $D_{-\infty}$ gets very large. Once N grows outside of that region, n_t is smaller than the minimum occupancy, and the algorithm starts to reproduce the naive algorithm. Note that this usually means that the algorithm also converges to the correct value for $D_{-\infty}$, but the way it does so is by reproducing the results of another method. We can't expect this

---

[1] We investigated the multiply-expanded EBA, but it produced no surprising or significantly improved results, so will not be presented in this dissertation.
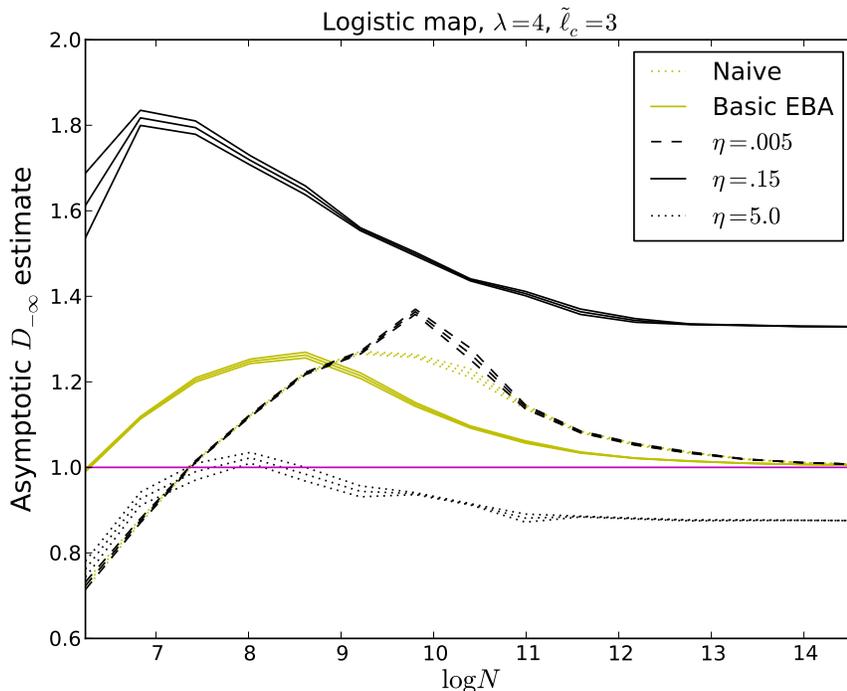
FIGURE 4.3. Convergence curves showing the general effect of varying η. For low η (in this case $\eta = .005$), the VBA closely follows the naive method. As η increases, the asymptotic value of $D_{-\infty}$ increases until it reaches a maximum. Increasing η past that value (in this case, near $\eta = .15$) causes the asymptotic value of $D_{-\infty}$ to decrease, eventually saturating at a value that closely follows a multiply-expanded version of the EBA (true saturation would not be guaranteed until $\eta = (2\tilde{\ell}_c + 1)^{D_S}$, which would be 7 for this plot, but $\eta = 5$ is nearly converged). This plot shows the extremal cases, but variation is fairly smooth between these limits.

to converge any faster than the naive method, and in fact it often converges a little slower.

Typical convergence behavior is shown in figure 4.4.

4.5. WHY THE VBA IS NOT AN IMPROVEMENT

We will show that the basic problem is that this method of estimating the measure reduces to finding the density of the attractor, and that the density is independent of the size of the box used. This independence makes our estimate of D_q become independent of q (and incorrect as well).
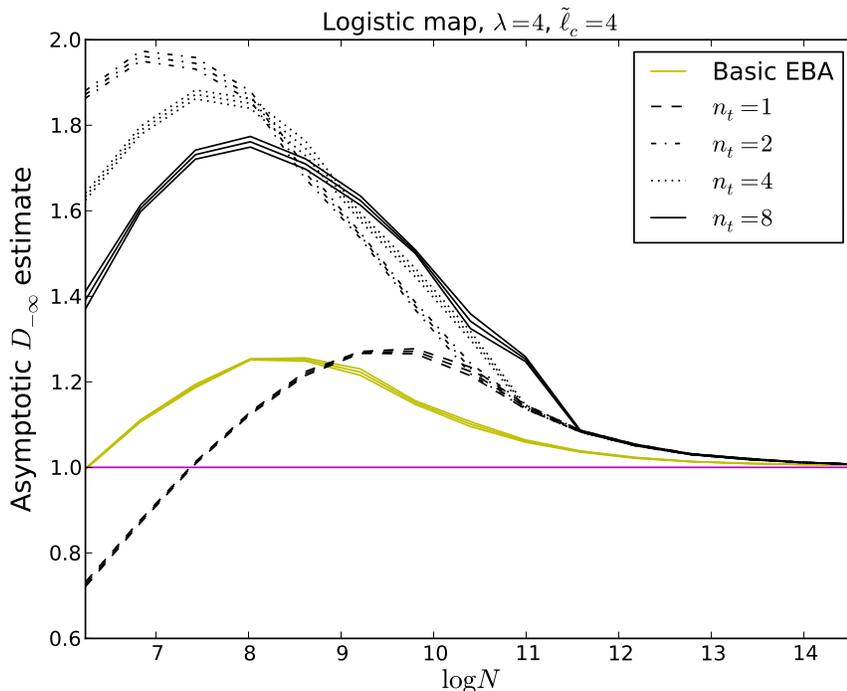
FIGURE 4.4. Convergence curves for the VBA for fixed n_t, applied to the logistic map at $\lambda = 4$. As N increases, eventually any fixed n_t will begin to reproduce the naive method. The general trend here is that as n_t is increased, the VBA continues to overshoot the expected result up to higher N values before jumping back to the naive result.

Imagine a "large" $\tilde{\ell}_c$ and a "medium" n_t, such that adjusting ℓ will consistently allow us to meet n_t for the majority of the cells in the histogram without hitting the cap imposed by $\tilde{\ell}_c$. This is the range of parameter values where the VBA is functioning as we intended it to, and it is the middle range of parameter values for which we observe convergence problems. Adopting a picture very similar to the one that we used to define $f(\alpha)$, each region of the attractor will have some local dimension D_F, so that boxes drawn in that region will enclose a number of points given by $n \sim \ell^{D_F}$. The dimension D_F could be different in different parts of the attractor. In fact, it is this distribution of different values of D_F that we are examining with the multifractal spectrum. However, to simplify our calculations, we will assume that there is a single region of the attractor that is dominating whatever sums we take, and that

this region has a more or less uniform dimension D_F. The regions that dominate the sums as $q \to -\infty$ should be the sparsest regions, and those regions are best described by $D_{-\infty}$. If we imagine perfect data, where we have such a large number of points and such a fine-grained histogram that we can expand in an almost continuous manner until we enclose n_t points, then our estimate of the density will become perfect, regardless of ℓ_i. Recall our notation measuring the number of expansions as $\tilde{\ell}$, so that the length of the side of an expanded box will be $\ell_e = (2\tilde{\ell}_e + 1)\ell_i$, meaning that the number of boxes in a cell expanded $\tilde{\ell}_e$ times will be $n_{cells} = (2\tilde{\ell}_e + 1)^{D_S}$ (with D_S the dimension of the space containing the attractor). Thus, for our estimate of the measure we will obtain

(4.1)
$$n_i^* = \frac{n_t}{n_{cells}} = \frac{n_t}{(2\tilde{\ell}_e + 1)^{D_S}}.$$

The size of any of our expanded boxes is determined by the condition that the box was expanded until it contained n_t points. That condition can be inverted to find ℓ_e. Locally, small boxes in this region should have a number of points in them that scales like $n = \rho \ell^{D_F}$. If we have expanded one particular box until it encloses n_t points, by increasing the length of the sides of the box to ℓ_e, then we should have:

$$n_t = \rho \ell_e^{D_F}$$

$$\ell_e = \left(\frac{n_t}{\rho} \right)^{1/D_F}.$$

Now, recall that we were actually looking for $\tilde{\ell}_e$, which was the number of expansions. If the original cell had size ℓ_i, then the expanded cell should have size $\ell_e = (2\tilde{\ell}_e + 1)\ell_i$. Thus:

$$2\tilde{\ell}_e + 1 = \left(\frac{n_t}{\rho} \right)^{1/D_F} \ell_i^{-1}$$

This determines $\tilde{\ell}_e$, so that we can predict n^* for any cell. Now plug this back in to equation 4.1.

(4.2)
$$n_i^* = n_t(2\tilde{\ell}_e + 1)^{-D_S} = n_t \left(\frac{n_t}{\rho}\right)^{-D_S/D_F} \ell_i^{D_S}$$

The scaling dependence of this equation is

$$n_i^* \sim \ell_i^{D_S}.$$

Apparently the new definition of n^* scales with ℓ_i, with the scaling exponent given by the dimension of the space, not the fractal dimension of any part of the set. Although the value of n^* in each of these boxes is independent of the fractal dimension we actually seek, one could hope that since the number of these boxes changes when ℓ_i changes, perhaps that dependence would introduce some correction to the dimension. The number of occupied boxes in this sparse region should scale like $N_{sp} \sim \ell_i^{-D_F}$, but it turns out that this dependence on the "correct" fractal dimension is not the dominant behavior when calculating D_q. The scaling of n^* is still the more important factor. For D_q we will calculate:

$$D_q = \frac{1}{q-1} \frac{\ln \sum (n^*)^q}{\ln \ell_i} \approx \frac{1}{q-1} \frac{\ln(N_{sp}(\ell_i^{D_S})^q)}{\ln \ell_i} \approx \frac{1}{q-1} \frac{\ln(\ell_i^{-D_F}(\ell_i^{D_S})^q)}{\ln \ell_i}$$

$$\approx \frac{qD_S - D_F}{q-1}.$$

As $q \to -\infty$, this gives $D_{-\infty} = D_S$, so that the ℓ dependence of the number of terms does not fix things. The estimate of the dimension is still not the correct D_q. The real problem here is that even in all the appropriate limits, as the various approximations involved in this result become more and more exact, our algorithm does not asymptotically approach the

correct value. We never achieve this perfect limit, so the experimental results do not always match this prediction, but this shows that as our data gets better and better, we do not expect the convergence to improve. We desire, as our data becomes perfect and as we let $q \to -\infty$, to obtain $D_q \to D_{-\infty}$.

The basic assumption that led to this problem was that we were in the part of parameter space where each box needed to expand, and that each box was free to expand until it reached its target n value. This is an important part of the parameter space to investigate, since the idea of a box expanding until it meets a target is at the core of the VBA. This means that at its very core, this method cannot work. The functional part of parameter space will be when we choose n_t and ℓ_c such that the expansion of most boxes is stopped by hitting the cap. If ℓ_c and n_t are chosen so almost all boxes expand to hit the limit on ℓ, (for example, with large n_t and small $\tilde{\ell}_c$) this more or less reproduces the EBA. Its behavior can at best match the EBA, but will likely experience some of the issues that we showed lead to the wrong $D_{-\infty}$ spectrum. If we choose n_t and ℓ_c such that very few boxes expand, (for example, with low n_t or low $\tilde{\ell}_c$) we can more or less reproduce the naive algorithm. Again, we can at best match the convergence of that method. Regardless of the parameter choices, as we gather more and more data, fewer and fewer of the cells need to expand, meaning that the VBA will always eventually act like the naive algorithm (except for very large η). However, this only guarantees that the algorithm eventually converges to the right value. It does not mean that the VBA should converge any faster than the existing methods. Our investigations have shown that in general this method converges no faster than the basic EBA, and usually converges at about the same rate as the naive method.

4.6. Local Norm VBA

Consider the following modification to the VBA. When dividing by the volume of the expanded box to form $\mu_i = n_i^*/v_i^*$, instead of counting all cells in the expanded box, count up only the cells that contain at least some neighbors. The basic method uses the volume of the expanded box as the number of cells covered by the new box, regardless of whether those cells have any points in them. Counting only the number of occupied cells encapsulated should eliminate D_S from equation 4.2, and replace it with D_F, which is hopefully D_q. This modification takes a local average over nonzero neighbors, which is equivalent to the Local Norm version of the EBA. In fact, notice that if $\tilde{\ell}_c = 1$ and n_t is large, this will exactly reproduce the Local Norm EBA.

Results when this version of the VBA is applied to various sets are quite promising. Figure 4.7 shows the convergence curves for the logistic map at $\lambda = 4$ with fixed n_t, and figure 4.8 shows the convergence curves for the same set with fixed η instead of n_t. As $N \to \infty$, the fixed n_t version converges to the naive method, as expected, but for N below that convergence, the Local Norm VBA seems to reduce the amount of overshoot that we observe in our standard methods. Figure 4.8 displays a similar trend, but as N gets large, the VBA no longer locks on to the exact same result as the naive method the way it does for a fixed n_t. However, the general description of the behavior is quite similar. With the fixed η plots, we noticed that there was a most optimal value of η, (near $\eta = 1$ for the logistic map at $\lambda = 4$, but generally more variable), beyond which the convergence curves began to systematically undershoot the exact result. Since η and fixed n_t are both just different ways of specifying the n target for the same underlying method, we suspect that an optimal value of n_t must exist, just like it does for η, but our choice of n_t values for the plots shown did not pass that optimal value.

FIGURE 4.5. Convergence curves for the Local Norm version of the VBA, for fixed n_t, tested on a Cantor set. Recall that $n_t = 1$ reproduces the naive results. Increasing n_t past the values shown here lowers the low N estimates of $D_{-\infty}$, but there is still a sharp jump back to match the naive result once N increases past a critical value.

As a final test piece, we also applied this method to the Hénon map, to see if consistent and trustable results could be generated. We used the Hénon map because its dimension is notoriously difficult to calculate, and we wanted a test that would really explore the limits of the new method. Results are in figures 4.9 and 4.10. This test shows clearly that the improvements shown in our other test cases are not completely general. Increasing n_t or η seemed to reduce the amount of overshoot in the lower dimensional tests. For the Hénon map, for fixed n_t, in figure 4.9, increasing n_t seems to prolong and exaggerate the amount of undershoot. We suspect that with increased computer time, we would see that as N increases, the various n_t curves would climb to meet the naive result, which would eventually converge to the correct $D_{-\infty}$ value. However, given that this suspected convergence is not yet apparent
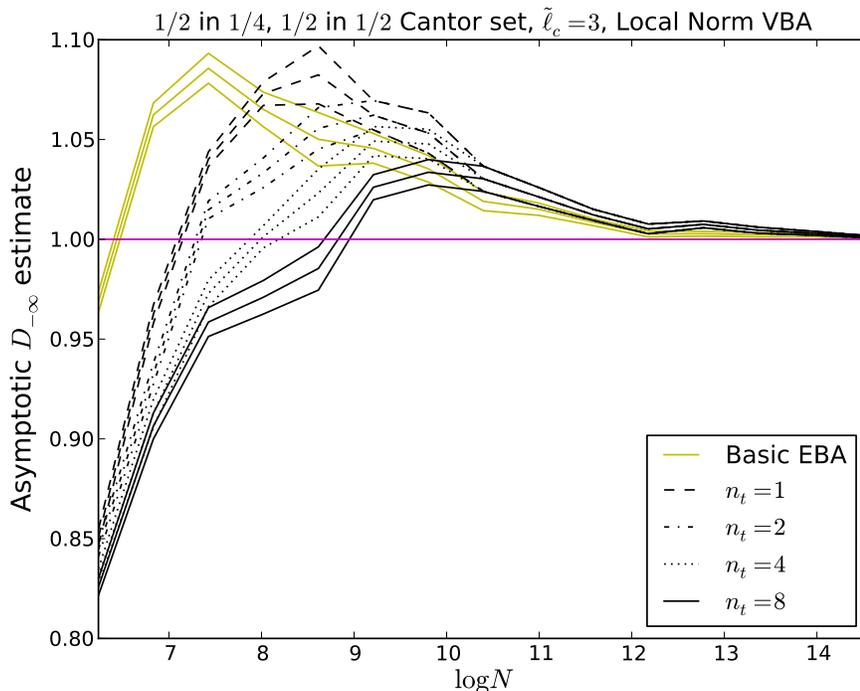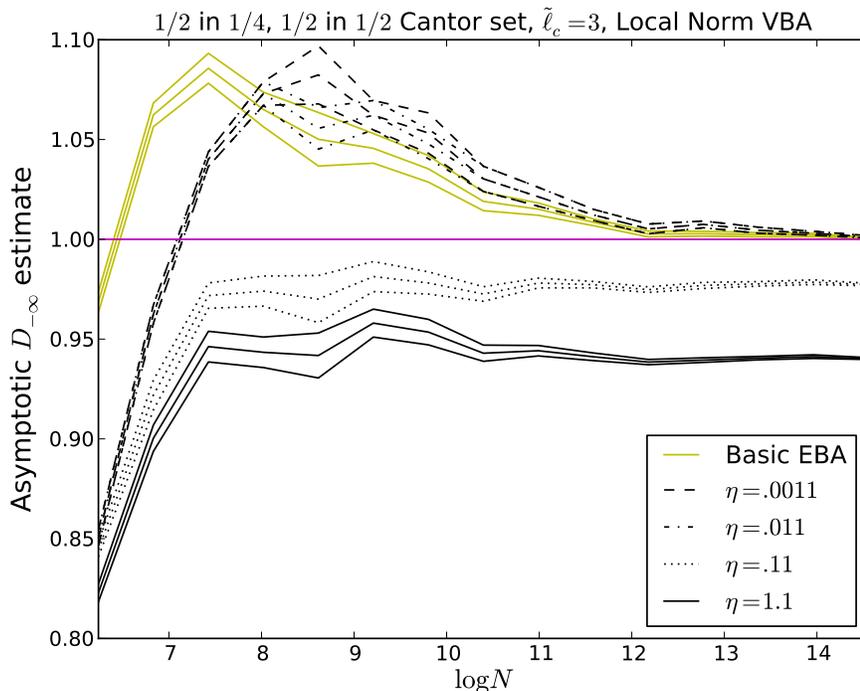
FIGURE 4.6. Convergence curves for the Local Norm version of the VBA, for fixed η, tested on a Cantor set. Once again, small η reproduces the naive results. Notice that once η passes some optimal value, the method begins to reach the wrong asymptotic value. In this case, that asymptotic value is between $\eta = .011$ and $\eta = .11$.

at these N values, but the EBA has settled to its final value (although that value may be incorrect), we can conclude that in this case the VBA is not a strong improvement. For fixed η instead of fixed n_t, in figure 4.10, we see a return to the behavior of the basic VBA, which seems to reach different asymptotic values as η is varied. In light of the success of the Local Norm VBA in lower-dimensional systems, it could be that these apparent asymptotic values are in fact not the true $N \to \infty$ behavior of the method. Perhaps at higher N values, all the different η curves actually converge to a single asymptotic result. However, as with the fixed n_t results, we cannot conclude that the new method is an improvement in this case.

So although we were not able to show completely universal improvement with the Local Norm modification to the VBA, we were able to demonstrate improvement in some cases.
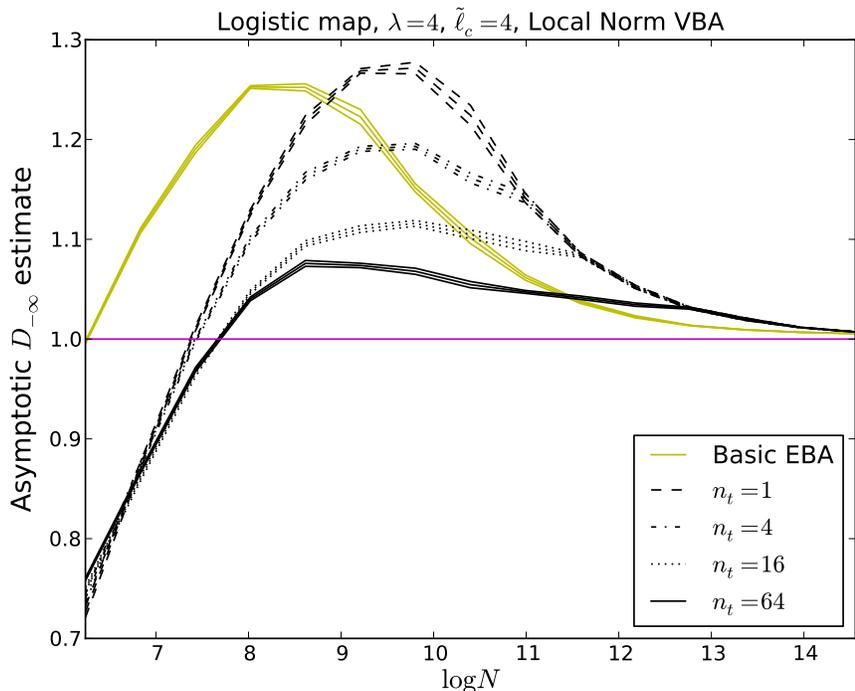
FIGURE 4.7. Convergence curves for the Local Norm version of the VBA. When n_t is very small or N is large, this method exactly reproduces the naive method since the minimum occupation is above n_t. The entire $n_t = 1$ curve in fact exactly duplicates the naive result. As we increase n_t, the Local Norm VBA switches between a multiply expanded Local Norm EBA and the naive method. The general trend here is that the larger n_t is, the more the overshoot seen in the naive method result is reduced.

The Local Norm VBA thus shows promise, and could bear further investigation. Perhaps in higher dimensional spaces (recall that the Hénon map, where the method did not appear to be an improvement, is a two dimensional map), much different regions of parameter space need to be investigated, or maybe some other simple modification along the lines of the Local Norm would improve convergence even more. It is interesting to see the continuing trend that the Local Norm modification seems to work the best in each of our modifications so far. For the basic EBA modifications, the Local Norm showed promise but was not universally better. For the a-norm, the Local Norm version of the a-norm tended to perform better in most cases. Here, for the VBA, the original method could be thought of as the extended
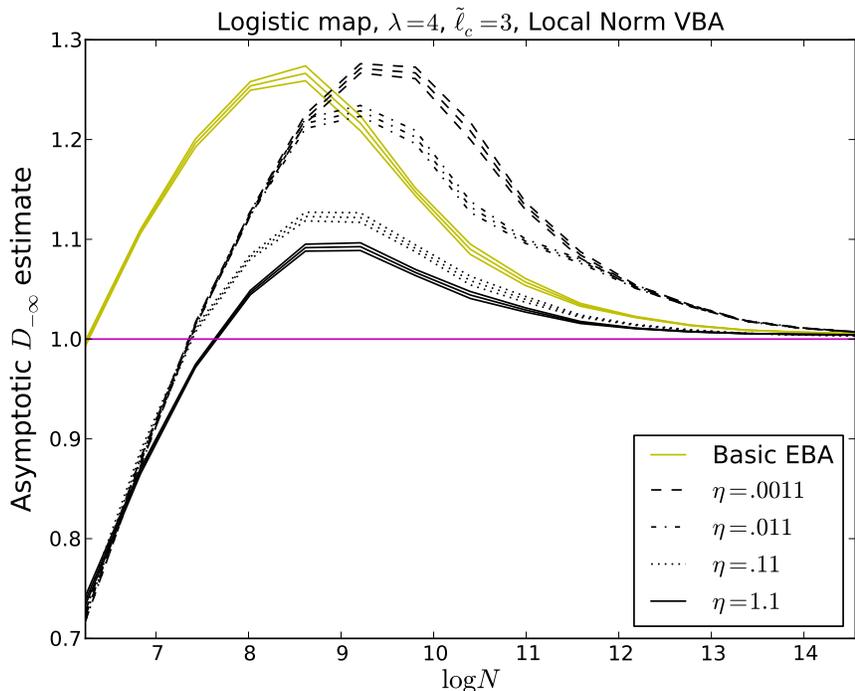
FIGURE 4.8. Convergence curves for the Local Norm version of the VBA, for fixed η instead of a fixed n_t. When η is very small or N is large, this method reproduces the naive method since the minimum occupation is above n_t. The entire $\eta = .0011$ curve in fact exactly duplicates the naive result, although it is not guaranteed to do so for all sets. As we increase η, the Local Norm VBA switches between a multiply expanded Local Norm EBA and the naive method. The general trend here is that the larger n_t is, the more the overshoot seen in the naive method result is reduced. We found that increasing η much past 1 tended to cause the final asymptotic value of $D_{-\infty}$ as $N \to \infty$ to decrease slowly below the exact result.

analog of the Global Norm modification. This time the difference between the Global Norm version and the Local Norm version is even more pronounced. We were able to show that even in the limit of perfect data, the Global Norm version of the VBA should not work, but we have demonstrated that the Local Norm version seems to be an improvement over our other modifications.
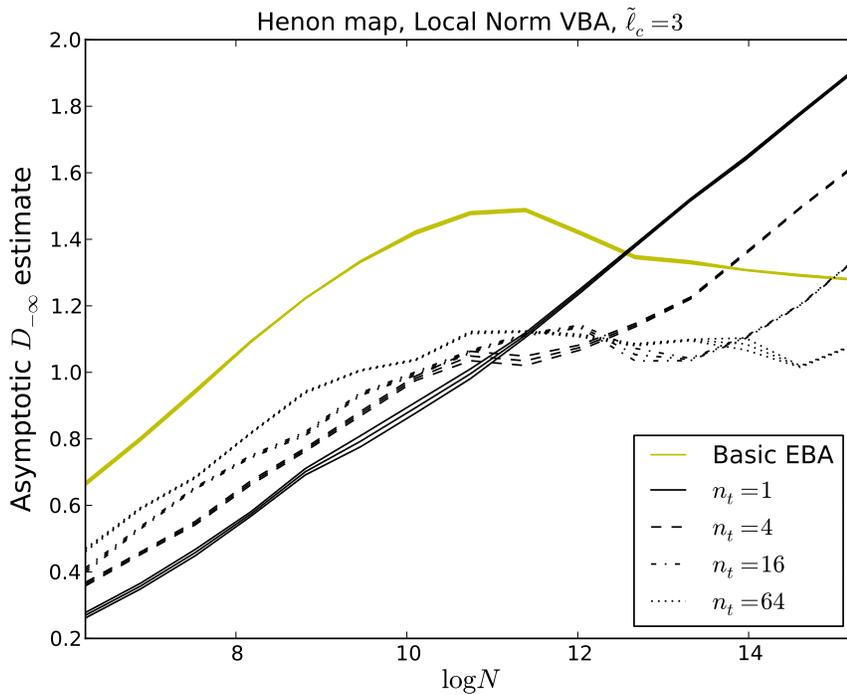
FIGURE 4.9. Convergence curves for the Local Norm VBA, when applied to the Hénon map. As before, setting $n_t = 1$ exactly reproduces the naive result. Recall that the exact spectrum is not known, but we know from [19] that $D_0 = 1.261 \pm .003$, so we know $D_{-\infty}$ must be greater than that value. The value that the various n_t curves flatten out at before climbing sharply is noticeably below that lower limit.
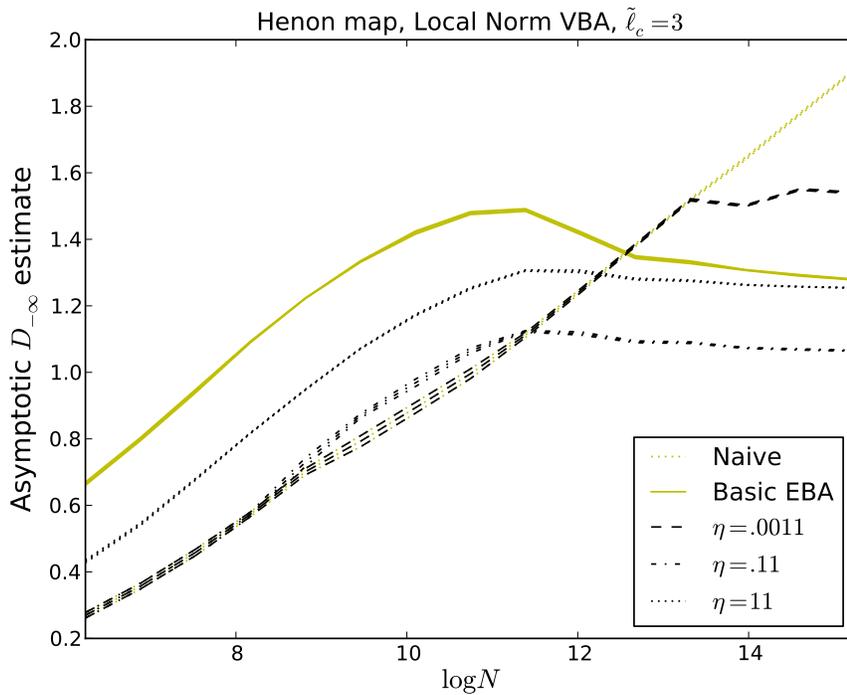
FIGURE 4.10. Convergence curves for the Local Norm VBA, when applied to the Hénon map. Recall that the exact spectrum is not known, but we know from [19] that $D_0 = 1.261 \pm .003$, so we know $D_{-\infty}$ must be greater than that value. The trend we observed here was as η increased from $\eta = .0011$ to $\eta = .11$, the curves moved monotonically downward. At some η near $\eta = .11$, the curves stopped moving down and began moving back up toward the EBA result. The $\eta = 11.0$ curve appears to asymptotically approach the measured value for D_0.

Patchwork Quilt Algorithm

5.1. Description of method

Recall that when originally defining Γ in section 1.2, we carefully allowed the possibility of a nonuniform cover, but we pointed out that nonuniform covers are usually most useful when deriving theoretical results. As before, a more rigorous derivation is covered in detail in [7]. The basic idea of this new algorithm is simple: use a nonuniform cover, in hopes of obtaining better convergence. The cover that is used will have cells of many different sizes, giving it the general appearance of a patchwork quilt. The idea is simple, but the implementation is not. Multiple realizations of this basic idea could exist.

The goal that we settled on for our implementation of this method was to ensure that even the sparse cells were occupied enough that the probability was nearly continuously changing, even in sparse regions. As described in section 1.4, it is the discontinuous jumps in probability that cause clipping errors. The difference between the $n = 1$ and $n = 2$ terms is too large, so that the $n = 1$ terms completely dominate the sum in Γ for the naive method. In order to accomplish a smooth variation by way of using a nonuniform cover, we devised the following procedure. First, choose a cutoff value c for the occupation. This cutoff is not exactly a minimum or a maximum, but it will become a general indicator of the occupation of most of the cells in the final patchwork quilt. Begin with a histogram of the attractor at a very fine length scale. Rebin this histogram to larger length scales, until the least occupied cell in the histogram is more occupied than the cutoff value, storing each intermediate histogram from this process in a stack of histograms which will be called the ladder. Once the final histogram is produced, with each cell containing $n_i > c$, split each

TABLE 5.1. A sample histogram demonstrating how the patchwork method works. For this case, we have selected $c = 5$, so the histogram is rebinned until all cells have $n > 5$, then selectively split until all cells have $n < 5$, by referring back to the list of histograms at different sizes. The final result is a histogram with cells of unequal sizes, but occupations close to c.

original histogram cells	0	1	1	0	4	3	3	1
rebinning to build ladder	1		1		7		4	
rebinning again	2				11			
final rebin (now all $n > c$)	13							
splitting cells with $n < c$	2				11			
splitting cells with $n < c$	2				7		4	
splitting cells with $n < c$	2				4	3	4	
final 'patchwork quilt'	2				4	3	4	

cell of that histogram. Splitting means referring back to the ladder to see which cells were combined together to form the cell in question. If the occupation of the current cell is larger than c, search in the ladder for the cells in the next level down, and replace the current cell with those cells. Repeat this process until all the cells either cannot be split any farther (because of the size of the original histogram) or do not need to be split any farther because their occupation is below c. The final result is a histogram with cells of many different sizes, but with most occupations near c. The final histogram composed of different sized cells is referred to as the quilt. The cells in the quilt representing sparser areas of the attractor are generally larger in size, so that they enclose more points, and the cells representing the dense areas of the attractor are smaller. Note that in usual operation, the algorithm undershoots, and each cell in the final quilt has a population less than c (apart from cells whose size was limited by the original histogram in which the data was stored). See table 5.1 for a sample operation of the patchwork algorithm.

Once that patchwork quilt histogram is formed, $\Gamma(q, \tau) = \sum p^q / \ell^\tau$ is easily calculated. However, this returns us to some of the very first problems faced during the preliminary steps of our implementation of the naive method and EBA. There should be a $\tau(q)$ curve

separating the $\Gamma \to \infty$ region from the $\Gamma \to 0$ region, but with finite data and finite grid sizes, zero and infinity will never actually be reached. So how can the regions be separated? One might hope that along the $\tau(q)$ curve, perhaps Γ would reach some fixed finite value (hopefully 1.0), but this does not seem to be the case. Instead, our strategy is to effectively take the ratio of Γ for various values of c. As we move to the limit of perfect data, we can hope that Γ would shrink toward zero in one region of the $q\tau$ plane, and grow without bound in the other region. So we choose a range of c values (the choice of c is the subject of section 5.4) to produce a set of covers, which in turn produce a family of curves $\Gamma_c(\tau)$. Of course, Γ is defined as a function on the $q\tau$ plane, but for simplicity imagine a series of plots of $\ln \Gamma_c$ versus τ at a fixed value of q. What this growing and shrinking ends up producing is a series of lines that all share a common intersection point (this intersection point is the subject of section 5.3). That intersection is the one value of τ where the region of growth $(\Gamma \to \infty)$ changes to the shrinking region $(\Gamma \to 0)$. That is the "correct" value $\tau(q)$. It generally does not happen to always be at $\Gamma = 1$, which is why we could not simply pick a value for Γ to serve as a cutoff separating the growing and shrinking regions. In order to produce multiple Γ curves to find an intersection point, we had to vary some quantity. Changing the number of points in the histogram is impractical since in real life, this would require extra data on demand. Changing the size of histogram cells ℓ as we do with our other methods has a much different meaning in the Patchwork method, since there is no single cell size to control. We briefly tested upper and lower limits on ℓ, but our investigations found that those techniques do not work well. Other methods may exist, but as we will show in section 5.3, varying c should reproduce the correct D_q spectrum. Finding the intersection of the lines is a relatively simple process to automate. We found empirically that in most cases, plotting $\ln \Gamma_c(\tau)$ produced straight lines, and the intersection point (or least-squares

best estimate of the intersection point) of straight lines is a linear algebra problem with a simple solution.

5.2. RELATION TO OTHER METHODS

Using different values of c on the same histogram can be interpreted as a more generalized version of the strategy that we use on all our other methods (methods for which the grid cells are all the same size as each other). The more generalized strategy would be to think of the histogram as a cover for the real set. By choosing different values of c in the patchwork quilt algorithm, we choose different covers, and as we will justify in more detail in section 5.3, plotting $\Gamma_c(\tau)$ curves for each cover and finding their common intersection gives us an estimate of $\tau(q)$. Interpreting our other methods in this framework, we are again generating multiple covers for the same set. Our other methods produce covers with uniform ℓ, and vary ℓ to find τ. Since there is no uniform global ℓ to vary in the Patchwork method, this procedure can no longer be applied. So how would we extend our old method to fit into our new framework? It would seem like for the old method we should have been plotting $\Gamma_\ell(\tau)$ for each ℓ value, and looking for the intersection of those plots, instead of our actual procedure, which is to plot $\ln \sum p^q$ versus $\ln \ell$ and take a slope.

We will now show that these two apparently different procedures are actually equivalent. For simplicity, we will examine the naive method. We can extend this analysis to any of our other modified methods by changing each p to a p^*. To recap some of the salient points of section 1.3, our usual method is to compute $\sum p^q$ for each different value of ℓ, and then find the slope of $\ln \sum p^q$ versus $\ln \ell$, which should be τ. To keep our calculations compact, let $\sum p^q = \chi$ (note that there is a hidden q dependence in χ, which Rényi [11] denoted with χ_q, but since we repeat these series of calculations many times at different fixed values of q, it is

not worth including this in our notation). The slope is found with a least-squares fit. Recall that the least-squares fit to the slope of $y = ax + b$ is given by:

$$a = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2},$$

with all the averages being averages over the data set. If we are fitting for τ in $\ln \chi = \ln \sum p^q = \tau \ln \ell + A$, then this means that our existing method produces the following expression for τ:

(5.1)
$$\tau = \frac{\langle \ln \ell \ln \chi \rangle - \langle \ln \ell \rangle \langle \ln \chi \rangle}{\langle (\ln \ell)^2 \rangle - \langle \ln \ell \rangle^2}.$$

Note that since we had one data point at each ℓ value, the averages are averages over ℓ.

Now let's examine what would happen if we used our new method on our old covers. If we instead consider each of these length scales as a different cover for the attractor and plot $\ln \Gamma$ as a function of τ for each ℓ value, we will need an explicit expression for finding the intersection. This is done by treating each $\ln \Gamma(\tau)$ line as a straight line with least-squares fitting (for the uniform ℓ case, $\Gamma = \sum p^q / \ell^\tau = \ell^{-\tau} \sum p^q$, so that $\ln \Gamma$ is an exactly straight line, but for our various values of c in the Patchwork method, $\ln \Gamma(\tau)$ is not exactly straight), and then finding the best estimate of the common intersection point, in a least-squares sense. This is a fairly straightforward linear algebra problem, which we will not describe here; details can be found in [25]. The end result is that if each line is of the form $\ln \Gamma = m_i \tau + b_i$, our best estimate of the x coordinate (or, more precisely, the τ coordinate) of the point closest to the common intersection of all those lines is:

$$\tau_0 = \frac{\langle m \rangle \langle b \rangle - \langle mb \rangle}{\langle m^2 \rangle - \langle m \rangle^2}$$

With the average being an average over all the various lines. For this case, all the covers would use a uniform grid, so that we could write $\Gamma = \sum p^q/\ell^\tau = \ell^{-\tau}\sum p^q = \ell^{-\tau}\chi$, and the different lines are merely different values of ℓ. (Note that the averages are thus averages over ℓ, just like in equation 5.1.) This means that the slopes and intercepts of all these $\ln\Gamma$ curves can be found by inspection, with no least-squares fitting needed. The slopes and intercepts will be will be $-\ln\ell$ and $\ln\chi$, respectively (each time, $\ln\Gamma = \tau(-\ln\ell) + \ln\chi$). This means that our point of intersection of all the lines is given by:

$$\tau_0 = \frac{(-\langle\ln\ell\rangle)\langle\ln\chi\rangle - \langle(-\ln\ell)\ln\chi\rangle}{\langle(-\ln\ell)^2\rangle - \langle-\ln\ell\rangle^2}$$

$$= \frac{\langle\ln\ell\ln\chi\rangle - \langle\ln\ell\rangle\langle\ln\chi\rangle}{\langle(\ln\ell)^2\rangle - \langle\ln\ell\rangle^2}$$

This exactly matches the result we get from the least-squares fit, shown in equation 5.1.

Thus, although our methods for finding $\tau(q)$ for the patchwork quilt algorithm seem very different from our methods in every other method investigated, both styles can be cast into the same framework. Although this comparison leads to a deeper understanding of both methods, it has not suggested any specific avenues for improvement.

5.3. EXISTENCE OF AN INTERSECTION POINT

When describing the patchwork quilt method, we motivated the existence of a point of intersection of the various $\Gamma_c(\tau)$ curves (formed by covers constructed with different c values) by describing it as the boundary separating the τ region where $\Gamma \to \infty$ from the region where $\Gamma \to 0$. We argued that at that boundary, the switch between growing and shrinking could produce a common intersection point. Let us now examine that intersection more carefully.

Recall that $\Gamma = \sum p^q/\ell^\tau$. Consider an idealized version of our algorithm, where the histogram is very very finely binned, and has a near-infinite amount of data in it (N points

total). Now consider a fixed q value, and assume that there is a collection of cells that have the same density as each other that are dominating the Γ sum. In our idealized case, each of these cells will have been diced until they contain almost exactly c points (c is the cutoff value that we have chosen). The probability in a cell will be c/N. Since the cells in this region all have the same density by assumption, and have the same occupation by construction, they must also have the same value of ℓ, and we can determine what value of ℓ the algorithm must have expanded the cells to without any additional assumptions.

We have a nonuniform grid, but let's assume that our sum is dominated by some more or less uniform set of cells which share the same occupation c and cell size as each other. The number of cells in this special region should be given by $n_{cells} = A_s \ell^{-d}$, where d is the dimension of this (possibly fractal) region. The factor A_s is an algebraic prefactor. That spatial prefactor should be independent of c, so it will not end up affecting our intersection calculations. Note that d will depend on q. It is the dimension of the subset of the attractor with the particular density that supplies the terms that dominate Γ at our chosen value of q. As $q \to -\infty$, the least dense regions will dominate, but for other values of q, different constant-density subsets will be selected. This behavior is exactly how we originally defined D_q in section 1.2, so we conclude that d is in fact D_q. To keep the following steps as clear as possible, we will continue to refer to this dimension as d, and imagine the $q \to -\infty$ limit, so that the dominant region is easily described and easily pictured as the least dense region. However, the dominant region, and the corresponding D_q, change with q.

Each cell has been diced until it contains very close to c points, so the average occupation of a cell will be c. However, the average occupation must also be given by the number of points in the sparse region divided by the number of cells in the sparse region (we argued earlier that the number of cells in the special region should scale with the dimension of that

region: $n_{cells} = A_s \ell^{-d}$). The number of points in the sparse region will be a fraction of the total number of points N, which we will write as $N_{sparse} = A_p N$. Just like A_s, the factor A_p is also independent of c. This means that our estimate for the average occupation is $N_{sparse}/n_{cells} = (A_p N)/(A_s \ell^{-d})$. Setting this equal to the average occupation c specified by our Patchwork algorithm gives a condition allowing us to find ℓ:

$$\frac{A_p N}{A_s \ell^{-d}} = c.$$

Solving for ℓ gives

$$\ell = \left(\frac{c}{N} \frac{A_s}{A_p} \right)^{1/d}.$$

Now let's assemble Γ.

$$\Gamma = \sum \frac{p^q}{\ell^\tau} \approx \sum \frac{(c/N)^q}{\{[(c/N)(A_s/A_p)]^{1/d}\}^\tau}$$

The number of terms in this sparse region should scale as described above, since each cell contributes one term in Γ; $n_{terms} = n_{cells} = A_s \ell^{-d} = A_s[(c/N)(A_s/A_p)]^{-1}$. Thus

$$\Gamma \approx A_s \left(\frac{c}{N} \frac{A_s}{A_p} \right)^{-1} \frac{(c/N)^q}{(c/N)^{\tau/d}(A_s/A_p)^{\tau/d}} = A_s^{-\tau/d} A_p^{\tau/d+1} \left(\frac{c}{N} \right)^{(q-1)-\tau/d},$$

so that apparently we expect

$$\ln \Gamma \approx \left((q-1) - \frac{\tau}{d} \right) \ln \left(\frac{c}{N} \right) + \ln(A_s^{-\tau/d} A_p^{\tau/d+1}).$$

Our scheme was to pick the point of intersection of the plots of $\Gamma(\tau)$ for many different values of c. This would be the one point where the value of Γ is independent of c, which would only

happen at

$$(q - 1) - \frac{\tau}{d} = 0 \qquad \rightarrow \qquad \tau = (q - 1)d.$$

Now recall that although we described the Γ sum as being dominated by the least dense part of the attractor, and pictured the highly negative q limit, our analysis is actually more general than that. For different values of q, different subsets of the attractor will dominate the sum in Γ, without changing our scaling arguments, so that in fact d is D_q. This means that we have just shown that if we find the τ coordinate of the intersection of the multiple $\ln \Gamma_c(\tau)$ lines, that τ value should be

$$\tau = (q - 1)D_q,$$

which is exactly what we hope for if this intersection point is the correct $\tau(q)$ value. This shows that searching for the common intersection point will produce the correct multifractal spectrum.

It should come as no surprise that the method is expected to recreate the true multifractal spectrum. In a way, this method is even more simple than the naive method. It is merely difficult to implement.

5.4. Identical covers and choosing c

The choice of c can be tricky. There is a different quilt (a different cover) generated for each value of c. If those c values are too close together, the different choices of c might construct the exact same covers as each other. If the covers are the same, then the $\Gamma(\tau)$ curves will be the same as well, and trying to find an intersection point will be meaningless, which means that the algorithm will fail to find $\tau(q)$.

The simplest case that would generate identical covers would be two values of c that are both below the minimum occupation of any of the cells in the original histogram (for example, if least-occupied cell has n_{min} points in it, choose $c = n_{min} - 1$ and $c = n_{min} - 2$). For both those values of c, the algorithm would not rebin at all on the first step, and the patchwork quilt covers that would be generated would both be copies of the original histogram. If we increase one of the c values until the cover corresponding to that c value changes, the exact condition for the change is difficult to state succinctly, apart from the obvious statement that we require at least one cell to change. For example, the condition is not so simple as setting $c > n_{min}$. In that case, the algorithm would rebin until all cells had occupations larger than n_{min}, which might be one step, but might be several, depending on the spatial distribution of the cells. It would then split each of those rebinned cells until they all had occupations less than c, which again might be just one split, but might split until the cover was all the way back to the starting configuration. Because of this ambiguity, we were unable to construct a simple condition for the minimum value of c, but generally found that having the smallest c be slightly larger than the occupation of the least occupied cell worked well, since the next larger c value would be large enough to be across the transition to a different cover.

Another simple case that could produce identical covers would be two c values just less than the occupation of the entire histogram (if the total is N, try $N - 1$ and $N - 2$). For those two c values, the algorithm would rebin until the histogram was a single cell with all N points in it, then split that cell until all the subcells had occupation less than c, which for reasonable sets and large N would be once. The final quilt would be the same for both our choices of c, yet again. As with choosing c too small, the exact value of c that would produce different covers depends on the spatial distribution of the points in the attractor and

cannot be expressed simply in terms of easily measured metrics such as total N, minimum cell occupation, or maximum cell occupation.

Of course, apart from these upper and lower limits on c, it is also possible that certain choices of c might turn out to be too close together for certain sets. This is easy to imagine for very large N. Changing c by 1 might not have any impact on which cells need to be combined and split in order to produce a quilt if every occupied cell has many hundreds of points in it. However, the upper and lower limits are the important limits, since we can tolerate a list of covers that contains some duplicate entries, so long as not all of them are identical. A single cover that is different from the rest is enough to give us a result, even if that result is slightly dubious. The problems we encountered during the first stages of development of this algorithm were usually caused by inadvertently selecting c values that all fell above N or below n_{min}.

After much experimentation, we found that the best procedure was to choose the smallest c as one larger than the minimum occupation, the largest c as a multiple η of the most occupied cell, and to pick η near 6. We then chose a small number of c values logarithmically spaced between those limits. Using values for c that do not scale with N (what we call 'flat' values) works well until the minimum occupation grows larger than c_{max}, at which point the algorithm begins to encounter the problems described here. Choosing c as a multiple of the occupation makes the cutoffs scale with the occupations, so that we encounter these problems much more rarely. Our procedure is not rigorously proven to always select unique covers, but it takes a highly unnatural set to fall into the special cases where our choices of c produce identical covers.

Investigation of the choice of c also led us to consider a modification to the Patchwork method. Occasionally, usually when the grid boundaries are incommensurate with the actual

fractal boundaries, we encounter very sparse cells near very dense cells in what should be a uniform region. For example, if the fractal should be a uniform probability over the interval from zero to 1/3, and the histogram cells are divided in half at every division step, then no matter how finely the histogram is divided, there will always be one cell that spans from slightly left of the edge of the probability distribution at 1/3 to slightly to the right of that edge. This one cell will always have a lower occupation than its neighbors (a lower occupation than it should). This problem persists for almost all choices of c, since cells are split until every member of the cell has an occupation less than c. On the second to last split, the problem cell is still included in a larger cell with occupation above c. After the last split, the problem cell is far below c, but its neighbor, entirely within the uniform region, might be quite close to c. To fix this problem, we simply stop splitting each cell one step earlier, so that instead of splitting until every cell has an occupation less than c, we split until every occupation is just barely more than c. Notice that this does not simply take back up globally to one level before the stop point in our brief example in table 5.1. It instead stops each individual cell just before its last split. A demonstration of the new method is shown in table 5.2.

We call this modification 'overshooting,' since the normal Patchwork method undershoots and produces occupations just below the chosen cutoff, but this modification produces occupations just above the cutoff. Note that this is a quite minor modification whose real purpose is to bring all occupations closer to c. Thus all of our work in the previous sections, showing that the intersection of the various $\Gamma_c(\tau)$ lines produces the correct τ value and showing that the intersection of the Γ plots is equivalent to our more conventional methods, is still valid. We tested this modification to the Patchwork method as well as our original Patchwork method, to see if there were situations where one version outperformed the other.

TABLE 5.2. A demonstration of the 'overshoot' modification to the Patchwork method. In this case, let $c = 15$. The overshoot method does not split the cell with occupation 21 because doing so would produce a cell with occupation 1. The original method would continue to split that cell (and others) until all occupations were below c. Notice that the final results for both methods give quilts with occupations near c, but the overshoot method has all occupations just above c and the normal method has all occupations just below c. Notice also that the overshoot method avoids placing the singly occupied cell in the final output.

original histogram	5	5	5	5	5	5	5	5	5	5	5	5	1	0	0	0
rebinning	10		10		10		10		10		10		1		0	
rebinning	20				20				20				1			
final rebin	40								21							
splitting cells	20				20				21							
'overshoot' result	20				20				21							
original method result	10		10		10		10		10		10		1			

5.5. PERFORMANCE

Before even investigating the $D_{-\infty}$ convergence behavior of this method, examine figure 5.1. Both the overshoot modification and the original undershoot version of the Patchwork method have a spike where $D_q \to \pm\infty$ at $q - 1$. This is again because the method actually generates a $\tau(q)$ curve, not a D_q curve, and at $q = 1$ convergence is not completely perfect. Thus, when $D_q = \tau/(q - 1)$ is calculated, we divide by zero to get a divergent result in the neighborhood of $q = 1$. For the EBA and our modifications thereof, the slight inaccuracies in τ at $q = 1$ were due to a loss of normalization, which was easily rectified. However, note that neither version of the Patchwork method fails to preserve normalization. The Patchwork method does not use surrogate values for p, it uses the actual completely unchanged values for p, simply combined together in a nonuniform cover. The problem arises because of small numerical errors in our fit procedure that finds the various $\Gamma_c(\tau)$ curves, then finds the intersection point of all those curves. Since the problem is more subtle than a simple lack of normalization, we have not been able to find a way to eliminate the spike.
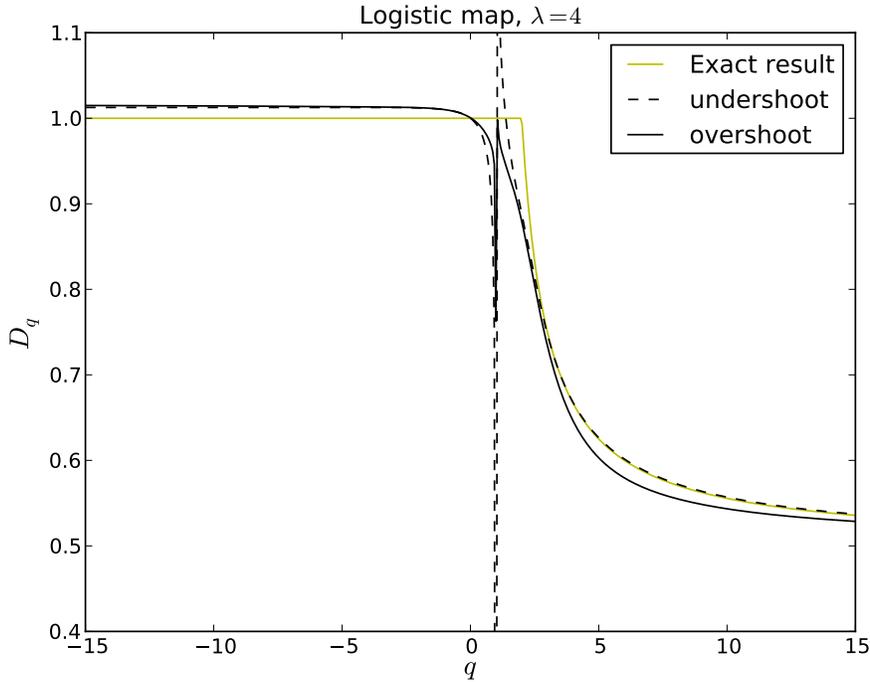
FIGURE 5.1. Numerically calculated D_q curves for logistic map at $\lambda = 4$, generated with the two versions of the Patchwork method. The exact spectrum is the expression derived in appendix B. Notice that both versions of the Patchwork method have a spike to infinity at $q = 1$.

As the D_q curve converges to the correct result, the spike becomes narrower and narrower, but we have no way to completely remove the spike. For the rest of this analysis, we do not discuss the spike explicitly, but the general trend is that better convergence of $D_{-\infty}$ indicates a narrower spike.

The computation time required for this method is generally longer than that required for our other techniques. However, direct comparison can be difficult. For our other modifications, the rebinning process was identical in every case, and the differences in computer time came from the different searches for neighbors. For the Patchwork method, the entire process is a series of rebinning steps, followed by searches through different histograms. We will not attempt to determine the exact scaling behavior, but we assert that it at least scales with the number of occupied cells. On a laptop computer, the computation of a single D_q
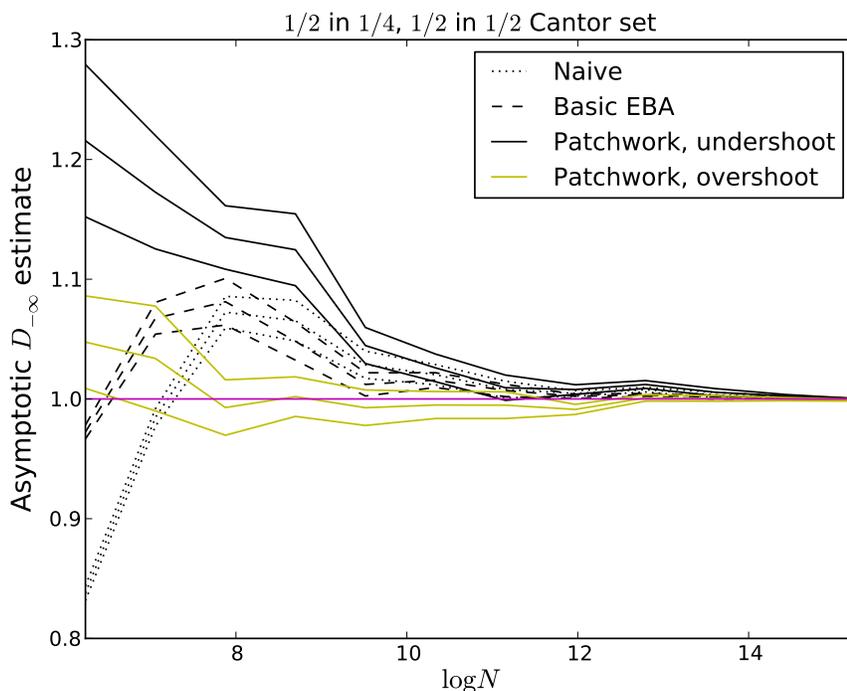
FIGURE 5.2. Convergence curves for our two versions of the Patchwork method, when applied to a Cantor set where the boundaries of the fractal and the histogram are commensurate. The horizontal line indicates the correct asymptotic value, found with the mother-daughter formula (see section 1.2.2). Note that when the method is set to overshoot c, the method converges to the correct $D_{-\infty}$ value more rapidly than the naive method or the EBA.

curve for our most occupied histograms still takes on the order of a minute, although like the VBA, this time is often roughly twice as large as for the basic EBA.

This method seems to find the correct asymptotic $D_{-\infty}$ value at much smaller N values than the EBA or the naive method, for some cases. Figure 5.2 shows the convergence curves for one of these highly successful cases - specifically a Cantor set for which the histogram cell divisions are known to match actual features on the fractal. However, this initial convergence is a relatively weak convergence; the result comes with large error bars.

The fast convergence is probably an artifact of the high symmetry of the Cantor set being examined in figure 5.2. In that case, the fractal is constructed by dividing the unit interval into powers of two, and the histogram cells happen to follow that same division scheme. To

check a less ideal case, we investigated a Cantor set where the divisions were powers of 3 (but with asymmetric probabilities). The support of this fractal (the set of points containing at least some part of the fractal) is the classic Cantor set. The cell divisions and the fractal features will not match in general. In fact, it can be shown that the point 1/4 is in the Cantor set, which means that no matter how many times the self-similar construction rule is applied, the cell division at 1/4 will always fall in an interval that is still in the set. It will never fall into the empty space between two intervals. Thus, that cell division will always be splitting an interval in two parts, which means that our algorithm cannot be forming a perfect cover. This will reproduce the problem illustrated in table 5.2, where a nearly empty cell is adjacent to a very occupied cell. We expect the overshoot modification of the Patchwork method to be a substantial improvement over the undershoot version of the Patchwork method. The question is how these new methods compare to the EBA and the naive method. The results are shown in figure 5.3. It turns out that the EBA converges better than the Patchwork method for this set.

Comparing figures 5.2 and 5.3 shows that the Patchwork method is strongly affected by whether a set in commensurate with the grid imposed by our method. This suggests a direction for future work, but let us put that aside for now. To investigate a set where the grid should not matter, we investigated the logistic map at $\lambda = 4$. As described in appendix B, we know the probability distribution for the logistic map at $\lambda = 4$ (as well as the multifractal spectrum D_q at that parameter value), and the distribution is smoothly varying, without the Cantor-like band structure that would make large changes to the results of the Patchwork method. At smaller λ values, the logistic set goes through a period-doubling cascade, in which the iterates of the map trace out more and more complicated patterns. The probability distribution associated with the periodic motion is in general a Cantor-like
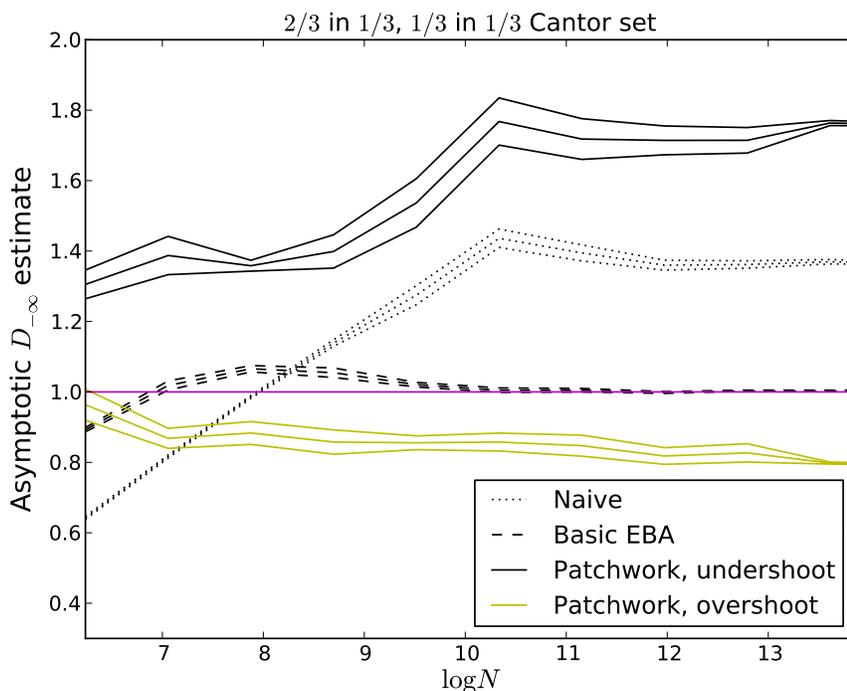
FIGURE 5.3. Convergence curves for our versions of the Patchwork method, applied to a Cantor set whose natural divisions are incommensurate with the cell divisions (cells were divided by a factor of 2 at every step). Notice that although neither version of the Patchwork method converges as quickly as the EBA, the overshoot method is closer.

structure. However, as λ in increased, the iterates of the periodic orbits eventually transition into chaotic bands, and then the structure of those bands changes. At $\lambda = 4$, all the chaotic bands have merged into a single chaotic band. Iterates of the map wander over the entire unit interval, and the probability distribution they converge to is described in appendix B. Convergence curves for our Patchwork method variations when applied to the logistic map at $\lambda = 4$ are shown in figure 5.4. As expected, the two methods disagree by less in this case. However, the overshoot version converges a little faster than the undershoot version. For this set, neither method converges faster that the basic EBA, although the convergence rate is comparable.
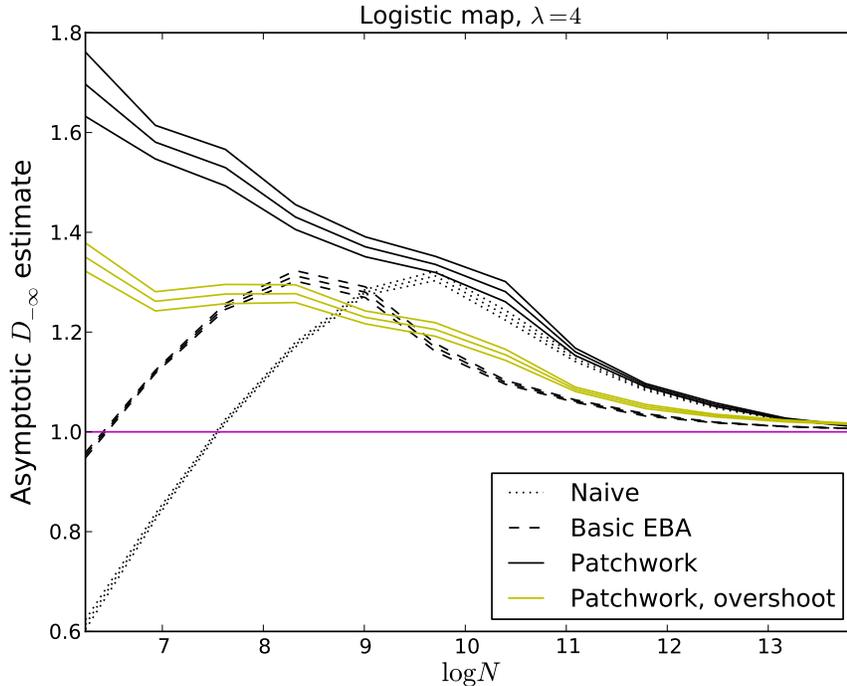
110

FIGURE 5.4. Convergence curves for our two versions of the Patchwork method, when applied to the logistic map at $\lambda = 4$ (the probability distribution associated with the logistic map at $\lambda = 4$ is a continuous distribution that fills the unit interval - there is no Cantor set structure to it). Note the close agreement between the 'overshoot' version of the Patchwork method and the basic EBA.

In still other cases, the Patchwork method does not produce improved results. Figure 5.5 shows the convergence curves for a relatively unsuccessful case - specifically the Hénon map (a two dimensional map). In that case, it is not clear if either version of the Patchwork method will converge at all. However, given the fact that the EBA has converged for N values much less than the maximum N value shown, figure 5.5 shows that the Patchwork method is not the improvement we might have hoped for. The Hénon map is notorious for being one of the attractors whose dimension is most difficult to compute, and the multifractal spectrum is not know for that set. We believe the EBA is not converging to the correct result, but have no conclusive proof one way or the other (we do know from [19] that $D_0 = 1.261 \pm .003$, and the asymptotic value the EBA is reaching is suspiciously low in light of that result). One
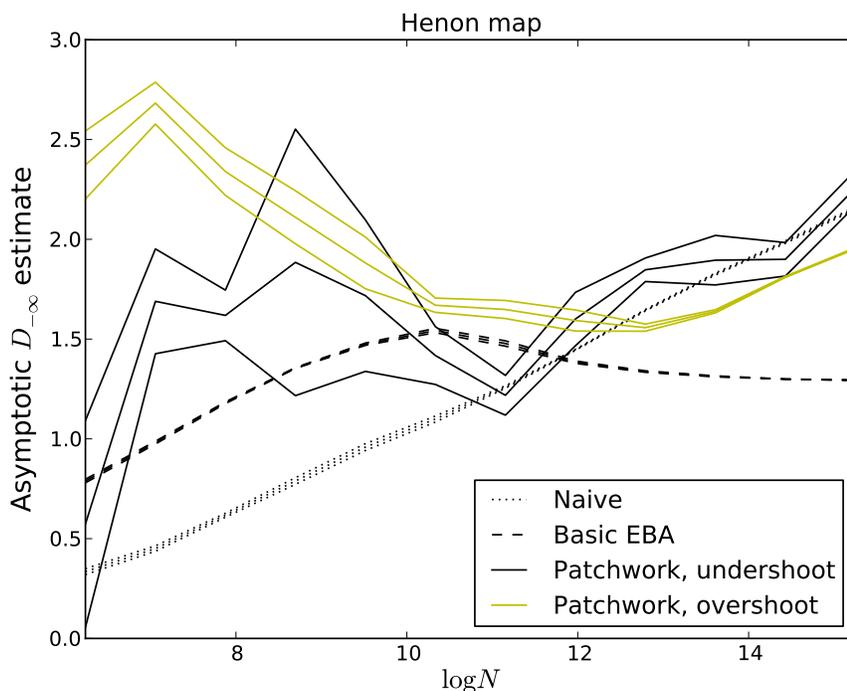
FIGURE 5.5. Convergence curves for our two versions of the Patchwork method, when applied to the Hénon map. Note that neither version of the method seems to be converging to a single value.

could argue that the fact that both variations of the Patchwork method are continuing to grow is in fact an indication that the true asymptotic value for $D_{-\infty}$ is closer to 2. However, our other modifications from previous chapters consistently found $D_{-\infty}$ close to the EBA result.

We hypothesized that the early convergence for the commensurate Cantor set shown in figure 5.2 might be attributed to the fact that if the attractor is very sparse, the Patchwork algorithm will expand cells until they become very large, so that no cell is too sparsely filled. Because of this, even for small N, the Patchwork algorithm can detect self-similarity at large scales, even if the attractor is not populated enough for the EBA and naive methods to work well. Even when N is small, the naive method and EBA still use data from all available length scales, even the very smallest ℓ limit. If the set is very sparse, cells might

be insufficiently filled at the small ℓ limit. To test this hypothesis, we tried two different tests. We tested the Patchwork method on a Cantor set that used different construction rules at different scales, and we tested the Patchwork method on a set formed from two non-overlapping Cantor sets in the same histogram.

For the Cantor set that changes construction rules, we began selecting points to fill a Cantor set according to construction rule A. A random process placed points in the center of some interval containing part of the Cantor set. We began with the point in the center of the unit interval, then using a discriminant, selected the right or left subinterval of the unit interval, then the right or left part of that subinterval, then the right or left part of that sub-sub-interval, and so on until the length of the interval was smaller than the size of a histogram cell. So long as the interval under consideration is longer than some predetermined length, construction rule A selects the interval lengths and probabilities. Once the length of the interval is below the predetermined length, the selection is made using construction rule B. The end result is a finite number of small copies of fractal B, so should have the dimension of set B. The theory is that for small N, the Patchwork algorithm will expand all cells until they detect self-similarity rule A, controlling the placement and population of the copies of set B, thus using the Patchwork algorithm's sensitivity to large-scale self-similarity to force it to generate an incorrect result. The naive method and the EBA, on the other hand, would still use data obtained on the smaller length scales, which would reflect construction rule B. Results are shown in figure 5.6. We found that although the undershoot version of the Patchwork method followed our predictions, and detected a construction rule closer to the large-scale construction rule, the overshoot version of the patchwork method was the closest to detecting the correct spectrum. In contrast, our specially constructed set managed to have the largest effect on the basic EBA. This is most likely because expanding

each box in the EBA caused the EBA to see effects from length scales one order of magnitude larger than ℓ, making it more sensitive to the large- scale behavior than we had predicted. These results were a little unexpected, since we built this set with the intent of making the Patchwork method detect the wrong rule, and in hopes that the EBA would detect the right rule. What's happening is that the EBA is generating a series of plots of $\ln \sum p^q$ versus $\ln \ell$ at different q values (recall that the slope of $\ln \sum p^q$ versus $\ln \ell$ is τ). The data points for $\ell > 2^{-5}$ (2^{-5} was the break between the two construction rules) form a line with a slope corresponding to the incorrect τ, and the data points for $\ell < 2^{-5}$ form a line with the correct slope. Since we made no a priori modifications to the EBA, it finds a slope by fitting to all the data points available, giving a final value for τ (and thus also D_q) which is a rough average of the two values. A simple tweak of some of the parameters involved would make the EBA pick out the correct result, although this modification would be difficult to automate. In the EBA, inspecting the plots of $\ln \sum p^q$ versus $\ln \ell$ by eye would show two linear regions, and the method could be modified to only find the slope of the linear region that actually extends to $\ell \to 0$. In a more realistic system showing this type of behavior, the two linear regions might have a more continuous transition instead of an abrupt break, but the principle is the same. Modifications fine tuning what ℓ values are considered are quite easy to make, and would undoubtedly be one of the preliminary steps of any application of the EBA to a real datastream. However, modifying the operation of the Patchwork method to make it converge to the correct result would be considerably more involved.

Regardless of the unexpected accuracy of the overshoot version, these results are consistent with our hypothesis that the Patchwork method detects large-scale patterns very quickly. The results are also consistent with the general trend that the overshoot method is more robust.
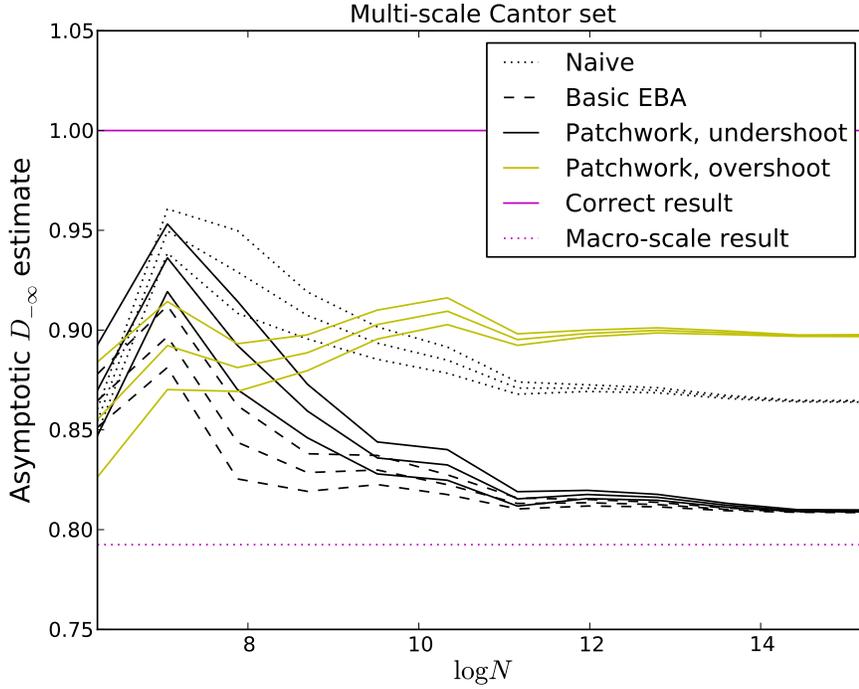
FIGURE 5.6. Convergence plot for the Patchwork method, when applied to a Cantor set made with a construction rule that changes depending on the scale under consideration. In this case, for features larger than 2^{-5}, the set followed a $\frac{2}{3}$ in $\frac{1}{4}$, $\frac{1}{3}$ in $\frac{1}{4}$ construction rule. For features smaller than that scale, the construction rule placed $\frac{1}{2}$ in $\frac{1}{4}$ and $\frac{1}{2}$ in $\frac{1}{2}$. The correct dimension comes from the smallest features of the set, since the macro-scale construction rule simply selects where to place a finite number of copies of the micro-scale set. Notice that the overshoot method is closer to the correct result.

To test multiple construction rules, we placed two Cantor sets in the same histogram.

See Appendix C for a derivation of the multifractal spectrum of the union of two Cantor sets.

As with the multiple construction rules, our theory was that the Patchwork algorithm would

be controlled by the wrong construction rule when N was small, while the naive method

and the EBA would still detect the correct scaling. The two sets used were a $\frac{2}{3}$ in $\frac{1}{4}$ and

$\frac{1}{3}$ in $\frac{1}{4}$ set, and a $\frac{1}{2}$ in $\frac{1}{4}$, $\frac{1}{2}$ in $\frac{1}{2}$ set. Our derivation in appendix C shows that the correct

dimension comes from the lower τ of the two individual sets, which corresponds to the higher

$D_{-\infty}$ (since, as shown in appendix A, a simple approximation for $D_{-\infty}$ is the slope of τ as

$q \to -\infty$). The two individual results for our adjacent sets, were they alone, would be

$D_{-\infty} = \ln 3 / \ln 4$ and $D_{-\infty} = 1$, so the correct result is apparently $D_{-\infty} = 1$, but we expect some detection of the other trend. The results are in figure 5.7. In this case, we have finally found an instance in which the overshoot modification is not more robust. Here is why. For clarity, call the two sets set A and set B. Say that set A has the higher $D_{-\infty}$, which would mean that it has the lower density sparse regions. We believe the reason that the overshoot method is detecting the wrong pattern is that it rebins in a manner that avoids creating sparse cells. When dealing with set A, it rebins those cells until they have an occupation a little closer to the occupations in the sparse regions in B. Thus, when calculating D_q, the two different sparse areas are more equivalent than they should be. For the undershoot method, there is a larger difference in density, so that the sparse areas in set A dominate, which they should.

The multiple Cantor sets result could cast the robustness of the overshoot method where other sets are concerned in a different light. Perhaps the reason that the overshoot method is so effective is that it treats all areas of the set as if they follow the same construction rule. That is not to say that it treats all areas of the set as following a symmetric self-similar construction rule, but merely that the method arranges things so that disparate sparse areas contribute a roughly equal amount to the sum in Γ. For many sets, this would tend to average out small fluctuations that would otherwise cause a few under-occupied cells to dominate the sum. If the set actually has a single construction rule, this would allow the overshoot method to detect that rule more accurately. Our tests of the two-scale Cantor set, in figure 5.6, show that even when that rule is rather strange, the overshoot method seems to detect it better than other methods. However, if there are entirely different behaviors mixed together, the overshoot method may detect the wrong multifractal spectrum. The real question is which of these types of behavior is more prevalent in nature. Feigenbaum showed

that a large family of maps follow the same period-doubling route to produce a Cantor-like structure on their way to chaos [1], which would seem to imply that for most chaotic systems the overshoot version of the Patchwork method would be appropriate. We believe that this holds in almost all chaotic systems, and it would be extremely rare to see two completely different attractors in the same chaotic system. However, it is believed that the reason that fully developed turbulence is so difficult to analyze is that there are many different types of behavior, intermixed in a very complicated manner (full turbulence is not chaos). This would suggest that the overshoot method should not be applied to turbulent systems (or other real systems of analogous complexity). This is a subtle difference. It's not merely the presence of high dimensional attractors, it's the presence of two different attractors intermixed that would cause problems for this method. Furthermore, we suspect that if we were to build a single construction rule that broke each interval into four pieces at each step (instead of two construction rules that broke each interval into two pieces at every step), the overshoot method would detect the correct spectrum. In that case, the same construction rule applies to every sparse region, allowing the averaging effect described earlier to detect the correct D_q.

In short, we found that in some situations, the Patchwork method did an excellent job of detecting multiscale self-similarity. However, this detection was not general. This suggests some refinements to the implementation, such as considering more general covers. Our implementation used the existing histogram cells to make the quilt, combining cells in preset groups to change the sizes. A more complicated implementation could slide cell boundaries, allowing the cell divisions to always be commensurate with the natural divisions of the fractal being examined. However, the cells thus generated could easily overlap without a very careful selection scheme. The number of covers generated in such a process would
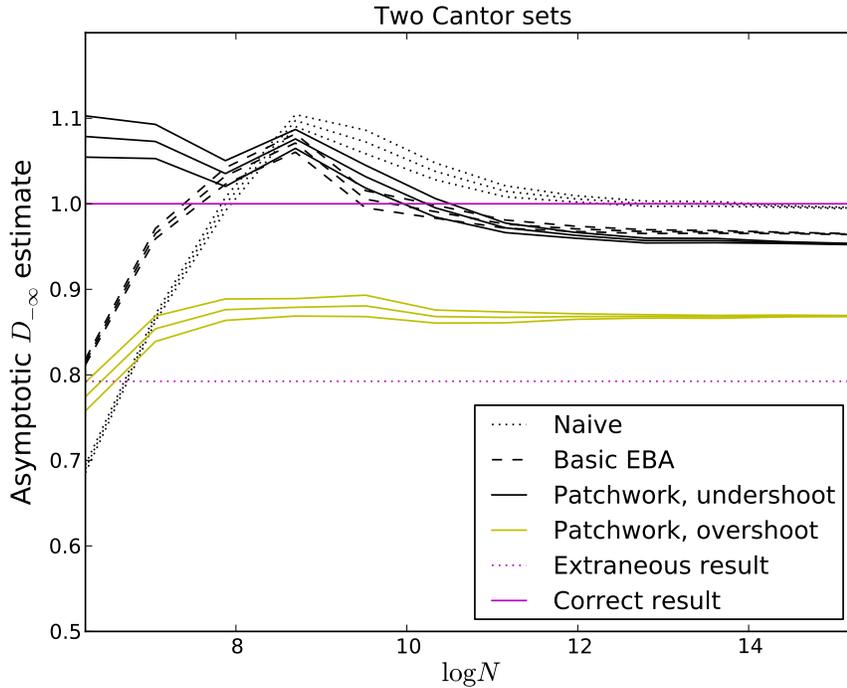
FIGURE 5.7. Convergence curves for the two versions of the Patchwork method, applied to the union of two disjoint Cantor sets. A derivation of the expected results can be found in appendix C. The two horizontal lines are the dimensions we would have calculated by considering either of the two sets alone. The correct result is the larger of the two values.

also increase by at least an order of magnitude, which could necessitate a more complicated method for determining $\tau(q)$ from the multiple covers. A simpler modification could be to make several covers with our current scheme that are shifted by small amounts, and to average, minimize, or maximize the results from that group of covers.

CHAPTER 6

Conclusions and Further Work

6.1. Conclusions

We have built several new definitions of n^*, and tried to cover as many possible types of modification as we could. The a-norm systematically investigated a whole family of different possible combinations of a cell with its nearest neighbors. The VBA investigated combinations of a cell with neighbors that were different distances away. We found that in both cases, modifications to the EBA could produce convergence for smaller data sets.

We have also investigated a completely different framework for computing D_q, with the Patchwork method. In that case, we found again that improvements are possible.

We also discussed and eliminated the normalization spike present in the basic EBA. Interpreting the EBA as an average, we found that averages that ignored unoccupied cells tend to be more effective, in every context we investigated (this is the Local Norm version of each of the methods we formulated).

6.2. Application to real systems

We have developed new methods for computing the multifractal spectrum, and refined existing methods. We tested our work on well-known systems as part of the validation and experimentation process, but a clear next step would be to begin using these new methods and refinements on real physical systems and data. What we have really established here is that a number of these methods show promise.

Throughout the process of developing these methods, we took care to ensure that all the new modifications and methods were fully extendable to an arbitrary number of dimensions, although that was never the focus of the work. For the purposes of keeping our simulations

simple enough to run on a laptop computer, most of our validations were performed on low-dimensional fractals, and on fractals placed in one or two dimensions. However, when working in higher and higher dimensional spaces, there are more and more directions possible for neighbors. Thus, the treatment of the interactions between a cell and its neighbors should become more and more important, indicating the possibility for many more interesting and surprising results when studying these methods in higher-dimensional spaces.

6.3. Other definitions of n^*

Even though we have studied a relatively broad family of different definitions of n^*, there are literally an infinite number of possible definitions, and it is completely possible that there still exists a novel and simple modification that we did not manage to investigate with our methods.

During our investigations, we encountered a few other possible n^* definitions that showed some promise.

We conducted brief experiments into using an EBA that simply expanded the box to include not just nearest neighbors, but to encompass points up to two, three, or n cells away. We found no significant improvement, and suspected that the larger expanded boxes would begin to detract from the locality of our estimates of p. However, we did not perform a fully systematic investigation of all the possible implementations of this idea. It is possible that, for example, like the VBA, the Global Norm version of this idea fails, but the Local Norm version is an improvement.

A further possibility is to use an EBA that has a larger expanded box, but weights the closer neighbors more than the distant neighbors. This method brings to mind our definition of a local dimension from section 1.2. The strength of the mixing between nearby cells and

more distant cells seems to hold a promise of many rich avenues for investigation. Is it best to think of such a method as an average of different multiply-expanded EBA methods, with non-uniform weighting, or is it more appropriate to control the weighting with an exponent of some kind? What will the normalization consequences be when using such a method? Is there a single limiting behavior that could be investigated (like the $a \to \infty$ limit with the a-norm)?

6.4. Wavelet-based analysis

Wavelet analysis [26] is a growing field, which has seen much success when applied to multifractal analysis. (For example, the study of the human heartbeat in [5] used a wavelet-based analysis.) Current techniques use scaling exponents related to the wavelet coefficients to compute the multifractal spectrum directly. This dissertation has had a completely different focus, but perhaps some ideas could successfully cross over. It could be possible to use some of the ideas of wavelet analysis to produce smoother estimates of the probability distribution, rather than using the coefficients directly.

6.5. Repeated application

During our research, we investigated the effect of applying the EBA (and our modifications of the EBA) multiple times. We found some interesting effects, but did not investigate those effects to our full satisfaction. Specifically, we predicted that for a uniform interval, we should see problems emerge with the spectrum calculations due to edge effects. The smoothed edges should produce a low dimensional set of underpopulated cells that have a drastic effect on $D_{-\infty}$, for the EBA applied multiple times as well as for the Global Norm applied multiple times. However, experimentally we did not observe these drastic effects.

Table 6.1. EBA applied to a uniform interval

original probabilities:	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	$\ldots$
p^* values:	$\frac{2}{n}$	$\frac{3}{n}$	$\frac{3}{n}$	$\frac{3}{n}$	$\frac{3}{n}$	$\ldots$

Here is a description of these effects. Consider a uniform interval broken into n cells, as is schematically illustrated in table 6.1. We expect to see $D_q = 1$ for all q, which would give $\tau = q - 1$, which in turn would mean that we expect $\sum p^q \sim \ell^{q-1}$. Let's analyze this interval with the basic EBA (applying the EBA multiple times will magnify the effect we will demonstrate here). Notice that this is the bare EBA, not renormalized, for ease of analysis. There are two end terms with $p^* = 2/n$, and the rest have $p^* = 3/n$. These two edge terms are responsible for edge effects. This gives

$$\sum p^q = 2 \left(\frac{2}{n} \right)^q + (n - 2) \left(\frac{3}{n} \right)^q$$

Since the unit interval is broken into n boxes, the length ℓ of any one box will be $\ell = 1/n$.

$$\sum p^q = 2(2^q - 3^q)\ell^q + 3^q \ell^{q-1},$$

which looks nice when written as:

$$\sum p^q = 3^q \ell^{q-1} \left\{ 1 + 2 \left(\left(\frac{3}{2} \right)^{-q} - 1 \right) \ell \right\}.$$

The expression takes the form of a perturbative expansion already. The leading behavior is ℓ^{q-1}, and ideally we could take a limit $\ell \to 0$, but in any real experiment, we have to consider finite but small ℓ and finite but large N. Instead, our focus must be on the range of validity of various assumptions. In this case, we should ask what balance of N, ℓ, and q we will need for the EBA to produce the correct result. Notice that for large negative q,

that coefficient multiplying the extra ℓ term in the expansion is quite large. Specifically, if $q = -30$, then that term is $2((\frac{3}{2})^{-q} - 1) \approx 383500.12$. As we make ℓ smaller, we can tolerate a more and more negative q value before reaching the limits of where these spurious edge effects should contribute. The edge of the allowable q values should occur when the two terms are of approximately the same size:

$$1 \approx 2 \left(\left(\frac{3}{2} \right)^{-q} - 1 \right) \ell$$

$$2\ell \approx \left(\frac{3}{2} \right)^{q}$$

The limiting q value would thus be

$$q_{edge} \approx \frac{\ln 2\ell}{\ln 3/2} = -\frac{|\ln 2\ell|}{\ln 3/2}$$

We typically had a minimum grid size around 2^{-9} to 2^{-11}. This would mean that if we were to investigate a uniform interval that filled that whole grid, we should have seen edge effects dominate for $q < q_{edge} \approx -15$. It is unclear whether these sorts of effects actually came into play or not. We would like to investigate this further, to at least come to a better understanding of the issue, if not to use it to improve some aspect of our D_q calculations.

These considerations surfaced during investigations of multiple applications of a single method. Another topic that could be investigated would be a mixture of multiple methods, such as using the EBA to generate a histogram populated with n^* values, and then applying the patchwork method to that new histogram, or considering an extended box of variable size as the VBA does, but then applying a a-norm to that extended box. It is conceivable that a mixture of that sort could use the strengths of one method to overcome the deficiencies of

another method. However, in the interest of being able to finish our investigation, we did not pursue further refinements of this type.

6.6. Further improvements to the Patchwork method

The goal with the Patchwork method was to automate a way of choosing a better cover for the fractal. The method we generated ended up using a lot of the machinery we developed for the EBA modifications. Specifically, the Patchwork method found its cell boundaries by rebinning the cells using our existing rebinning scheme. This placed severe limits on the locations of cell boundaries, which is a large part of why we had to develop the overshoot modification to the method. Returning to our first description of how nonuniform covers are used in the multifractal spectrum, back in section 1.2, recall that we have to choose a cover that maximizes or minimizes Γ in different regions of the $q\tau$ plane. The intuitive understanding that we developed for this was that maximizing and minimizing appropriately reduced to finding the cover that had the smallest amount of empty space in it, and was snug against the edges of the fractal. Our Patchwork method clearly does not generate this 'most snug' cover possible, in general. It is certain that more clever implementations exist that are able to pick a more optimal set of covers.

To be specific, consider the following one-dimensional algorithm. First choose a value c to be a target occupation, much like our own c for the Patchwork method. Begin constructing a cover for the set by placing the left edge of a very small histogram cell at the left edge of the attractor. Start moving the right edge of that cell to the right and keep track of how many points are enclosed as the cell expands. Once there are more than c points in the cell, stop expanding it, and place the left edge of the next very small cell at the right edge of the recently completed cell. Then begin expanding that new cell until it too has c points in it,

and so on. Continue this process until the right edge of the fractal is reached. This would generate a cover that had occupations much closer to c than either of our versions of the Patchwork method, and a number of areas for improvement are immediately obvious, such as the treatment of unoccupied cells (the cover could probably be improved by minimizing the number of empty cells that are included). Of course, this particular implementation is inherently one dimensional, and it is not clear how it could be extended into higher dimensions. However, it illustrates the point that improvements to the Patchwork method must exist.

Apart from entirely different methods for generating the covers in the Patchwork method, there is also the troubling issue of the spike at $q = 1$. A careful investigation of the spike at $q = 1$ for the EBA yielded a very simple solution, so there seems to be a possibility that another equally simple modification to our method could remove the spike we generate. However, we have not been able to uncover any such modification.

References

[1] M. J. Feigenbaum, "The universal metric properties of nonlinear transformations," *Journal of Statistical Physics*, vol. 21, no. 6, pp. 669–706, 1979.

[2] N. Metropolis, M. Stein, and P. Stein, "On finite limit sets for transformations on the unit interval," *Journal of Combinatorial Theory, Series A*, vol. 15, no. 1, pp. 25 – 44, 1973.

[3] B. Derrida, A. Gervois, and Y. Pomeau, "Iteration of endomorphisms on the real axis and representation of numbers," *J. Combinatorial Theory*, 1978.

[4] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980* (D. Rand and L.-S. Young, eds.), vol. 898 of *Lecture Notes in Mathematics*, pp. 366–381, Springer Berlin / Heidelberg, 1981.

[5] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, "Multifractality in human heartbeat dynamics," *Nature*, vol. 399, no. 6735, pp. 461–465, 1999.

[6] E. Bonnet, O. Bour, N. E. Odling, P. Davy, I. Main, P. Cowie, and B. Berkowitz, "Scaling of fracture systems in geological media," *Reviews of Geophysics*, vol. 39, no. 3, pp. 347–383, 2001.

[7] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, "Fractal measures and their singularities; the characterization of strange sets," *Physical Review A*, vol. 33, no. 2, pp. 1141–1150, 1986.

[8] H. Hentschel and I. Procaccia, "The infinite number of generalized dimensions of fractals and strange attractors," *Physica D: Nonlinear Phenomena*, vol. 8, no. 3, pp. 435 – 444, 1983.

[9] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Physical Review Letters*, vol. 50, no. 5, pp. 448–451, 1983.

[10] J. D. Farmer, "Information dimension and the probabilistic structure of chaos," *Zeitschrift für Naturforschung. Teil B, Anorganische Chemie, organische Chemie*, vol. 3A, pp. 1304–1325, 1982.

[11] A. Rényi, "On a new axiomatic theory of probability," *Acta Mathematica Academiae Scientiarum Hungarica*, vol. 6, no. 3-4, pp. 285–335, 1955.

[12] R. M. May *et al.*, "Simple mathematical models with very complicated dynamics," *Nature*, vol. 261, no. 5560, pp. 459–467, 1976.

[13] P. J. Stephens and R. Eykholt, "Analytical calculation of the dimension spectrum D_q for the 2^∞-cycle of the logistic map, and lack of convergence of the Feigenbaum rescaling parameter α." Unpublished paper on α convergence for the logistic map, July 2004.

[14] R. Badii and G. Broggi, "Measurement of the dimension spectrum $f(\alpha)$: a fixed-mass approach," *Physics Letters A*, vol. 131, no. 6, pp. 339–343, 1988.

[15] R. Reidi, "An improved multifractal formalism and self-similar measures," *Journal of Math Analysis and Applications*, vol. 189, pp. 462–490, 1995.

[16] R. Pastor-Sattoras and R. H. Reidi, "Numerical estimates of the generalized dimensions of the Hénon attractor for negative q," *Journal of Physics A*, vol. 2, pp. L391–L398, 1996.

[17] M. J. McGuinness, "The fractal dimension of the Lorenz attractor," *Physics Letters A*, vol. 99, pp. 5–9, 1983.

[18] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D Nonlinear Phenomena*, vol. 9, pp. 189–208, 1983.

[19] D. A. Russell, J. D. Hanson, and E. Ott, "Dimension of strange attractors," *Physical Review Letters*, vol. 45, pp. 1175–1178, 1980.

[20] G. A. Edgar and R. D. Mauldin, "Multifractal decompositions of digraph recursive fractals," *Proceedings of the London Mathematical Society*, vol. s3-65, no. 3, pp. 604–628, 1992.

[21] J. Schmeling and R. Siegmund-Schultze, *The Singularity Spectrum of Self-affine Fractals with a Bernoulli Measure.* Institut für Angewandte Analysis und Stochastik Berlin: Preprint, Inst. für Angewandte Analysis und Stochastik, 1992.

[22] P. Collet, J. Lebowitz, and A. Porzio, "The dimension spectrum of some dynamical systems," *Journal of Statistical Physics*, vol. 47, no. 5-6, pp. 609–644, 1987.

[23] K. Falconer, "The multifractal spectrum of statistically self-similar measures," *Journal of Theoretical Probability*, vol. 7, no. 3, pp. 681–702, 1994.

[24] M. Hénon, "A two-dimensional mapping with a strange attractor," *Commun. math. Phys.*, vol. 50, no. 69–77, 1976.

[25] Wikipedia, "Line-line intersection — wikipedia, the free encyclopedia," 2014. [Online at `http://en.wikipedia.org/w/index.php?title=Line-line_intersection&oldid=589297087`; accessed 25-January-2014].

[26] S. Jaffard, "Wavelet techniques in multifractal analysis," in *Proceedings of symposia in pure mathematics*, vol. 72, pp. 91–152, 2004.

[27] B. "Derrida, A. Gervois, and Y. Pomeau, "Iteration of endomorphisms on the real axis and representation of numbers," *Annales de l'institut Henri Poincaré (A) Physique théorique*, vol. 29, no. 3, pp. 305–356, 1978.

[28] E. N. Lorenz, "Noisy periodicity and reverse bifurcation," *Annals of the New York Academy of Sciences*, vol. 357, no. 1, pp. 282–291, 1980.

[29] S. Grossman and S. Thomae, "Invariant Distributions and Stationary Correlation Functions of One-Dimensional Discrete Processes," *Zeitschrift Naturforschung Teil A*, vol. 32, p. 1353, 1977.

APPENDIX A

DERIVATION OF THE ASYMPTOTIC FORM OF D_q

In order to measure how well our calculations of the D_q spectrum are converging (mostly in the context of section 2.2), we seek an asymptotic expansion of D_q for highly negative q, so that we can extract $D_{-\infty}$ without having to use immensely large negative q.

We begin by assuming that the fractal in question can be approximated by some n-scale Cantor set. This is reasonable because one would expect the system to have followed some sort of period-doubling cascade [27] or reverse bifurcation sequence [28], giving chaotic bands of different widths and occupations. To make our approximation more general, we allow degeneracy in each of the interval lengths (meaning that we include n copies of each term). We then examine the mother-daughter formula $\Gamma_0 = 1$, (the mother-daughter formula is discussed in section 1.2.2) giving:

$$(A.1) \qquad\qquad n_1 \frac{p_1^q}{f_1^\tau} + n_2 \frac{p_2^q}{f_2^\tau} + \ldots = 1$$

In the $q \to -\infty$ limit, one term will dominate this sum. To determine which term that is, it is tempting to just examine the explicit q dependence. However, remember that we are trying to solve for $\tau(q)$, so we are examining Γ in a particular region of the $q\tau$ plane, where q and τ are of the same order of magnitude. To cut through this hidden dependence, we express equation A.1 in terms of D_q instead of τ. There is still a q dependence hidden in D_q, but D_q asymptotically approaches a constant $D_{-\infty}$ as $q \to -\infty$, and, even though D_q is not constant, it is of order 1 for the vast majority of systems, so that its variations are a second order correction to this approximation.

Thus, since $D_q = \tau/(q-1)$, we write the mother-daughter formula as:

$$n_1 \frac{p_1^q}{f_1^{D_q(q-1)}} + n_2 \frac{p_2^q}{f_2^{D_q(q-1)}} + \ldots = 1$$

$$n_1 \left(\frac{p_1}{f_1^{D_q}}\right)^q f_1^{D_q} + n_2 \left(\frac{p_1}{f_2^{D_q}}\right)^q f_2^{D_q} + \ldots = 1$$

Now it is clearer that to determine which term is largest (as $q \to -\infty$), we should examine the quantity $x_i = p_i/f_i^{D-\infty}$. The term in Γ with the smallest x value will be the largest term, since q is negative. Order the terms so that the smallest x is x_1 or x_{min} (even though the term containing x_{min} is the largest), the second smallest x is x_2, and so on. For now, assume that all the x values are distinct. To condense our notation a little, let $y_i = p_i/f_i^{D_q}$. Each y_i will asymptotically approach x_i, and even for finite q, x and y will be close to each other. However, y is the q dependent variable, and x is the asymptotic value, independent of q.

As $q \to -\infty$, the largest term dominates the Γ sum entirely, and we may ignore all other terms, so we have:

$$n_{min} y_{min}^q f_{min}^{D_q} \approx 1$$

Clearly as $q \to -\infty$ this requires that $y_{min} \to x_{min} = 1$, which means:

$$\frac{p_{min}}{f_{min}^{D-\infty}} = 1$$

$$D_{-\infty} = \frac{\log p_{min}}{\log f_{min}}.$$

But this is only a limiting value for $D_{-\infty}$. We wish to determine how D_q approaches this limit. Returning to the more precise $n_{min} y_{min}^q f_{min}^{D_q} \approx 1$, we have:

$$n_{min} \left(\frac{p_{min}}{f_{min}^{D_q}}\right)^q f_{min}^{D_q} \approx 1$$

$$\log n_{min} + q \log \frac{p_{min}}{f_{min}^{D_q}} + D_q \log f_{min} = 0$$

$$\log n_{min} + q \log p_{min} - q D_q \log f_{min} + D_q \log f_{min} = 0.$$

So that our next approximation to D_q is

$$D_q \approx \frac{q}{q-1} \frac{\log p_{min}}{\log f_{min}} - \frac{1}{q-1} \frac{\log n_{min}}{|\log f_{min}|}$$

Remember that the 'min' index refers to the segment that gave the smallest x, not necessarily the largest or smallest individual f or p.

Since $D_q = \tau/(q-1)$, our corresponding approximation to τ is

$$\tau = q \frac{\log p_{min}}{\log f_{min}} - \frac{\log n_{min}}{|\log f_{min}|}$$

In their most basic form, if we let $a = \frac{\log p_{min}}{\log f_{min}}$ and $b = \frac{\log n_{min}}{|\log f_{min}|}$, these two results can be written as

(A.2) $$\tau \approx aq - b, \qquad D_q \approx \frac{q}{q-1} a - \frac{1}{q-1} b.$$

Notice that as $q \to -\infty$, D_q approaches $a = \frac{\log p_{min}}{\log f_{min}}$, so that $D_{-\infty} = \frac{\log p_{min}}{\log f_{min}}$, consistent with our earlier result. This approach to the asymptotic value is even faster if there is a single (non-degenerate) most dominant term in the mother-daughter formula (as opposed to n copies of that term), which is hopefully quite common. In that case $n_{min} = 1$ and $b = 0$.

Now let us generalize. Since the size of the contribution from the various terms was determined by $x_i = p_i/f_i^{D_{-\infty}}$, it is possible to have two different terms with exactly the same size contribution but different p and f values. This could happen if, for example, $p_j = p_{min}^2$

and $f_j = f_{min}^2$. In general it will occur whenever we have more than one segment that gives $x_i = p_i/f_i^{D_{-\infty}} = 1$. Let us analyze the asymptotic form of D_q in that degenerate case.

Now there will be two dominant terms of approximately equal size, and we cannot ignore either one. Call the two indices corresponding to these dominant terms a and b. To extend our generality a little more, let there be n_a copies of interval a and n_b copies of interval b. Our approximation of Γ will be:

$$\Gamma \approx n_a y_a^q f_a^{D_q} + n_b y_b^q f_b^{D_q} \approx 1$$

These two terms must both be considered because they have the same x value (recall that y is the q dependent expression p/f^{D_q} which asymptotically approaches x). As with the limited-degeneracy case, that value must still be $x_a = x_b = x_{min} = 1$, which, through the same few steps as before, gives $D_{-\infty} = \log p_a/\log f_a = \log p_b/\log f_b$. As $q \to -\infty$, we may write $D_q = D_{-\infty} + \epsilon(q)$, where $\epsilon(q)$ is a small (non-constant) correction. Inserting our $D_q = D_{-\infty} + \epsilon$ approximation into our truncated expression for Γ gives:

$$n_a y_a^q f_a^{D_{-\infty}+\epsilon} + n_b y_b^q f_b^{D_{-\infty}+\epsilon} = 1,$$

Note that we can write

$$f_a^{D_{-\infty}} = f_a^{\log p_a/\log f_a} = \exp(\log f_a \frac{\log p_a}{\log f_a}) = p_a,$$

and since $D_{-\infty} = \log p_a/\log f_a = \log p_b/\log f_b$, we can follow an equivalent procedure to show that $f_b^{D_{-\infty}} = p_b$. Thus our expression becomes:

$$n_a y_a^q p_a f_a^\epsilon + n_b y_b^q p_b f_b^\epsilon = 1.$$

133

Now note that $y_i = p_i/f_i^{D_q}$ is can be written as

$$y_i = \frac{p_i}{f_i^{D-\infty+\epsilon}} = \frac{p_i}{f_i^{D-\infty}} f_i^{-\epsilon} = x_i f_i^{-\epsilon},$$

and since the x values we are considering are defined by $x = 1$, this gives $y_i = f_i^{-\epsilon}$ for y_a and y_b. Thus our approximation for $\Gamma = 1$ becomes:

$$n_a p_a f_a^{-(q-1)\epsilon} + n_b p_b f_b^{-(q-1)\epsilon} = 1$$

Now for each of the f terms, we write

$$f^{-(q-1)\epsilon} = \exp\left[-(\log f)(q-1)\epsilon\right] \approx 1 - (q-1)\epsilon \log f.$$

This means that our approximation for Γ becomes:

$$n_a p_a (1 - (q-1)\epsilon \log f_a) + n_b p_b (1 - (q-1)\epsilon \log f_b) = 1,$$

or

$$(n_a p_a + n_b p_b) - (n_a p_a \log f_a + n_b p_b \log f_b)(q-1)\epsilon = 1,$$

giving

$$\epsilon = -\frac{1 - (n_a p_a + n_b p_b)}{n_a p_a \log f_a + n_b p_b \log f_b} \frac{1}{q-1} = -\frac{b'}{q-1}.$$

So that the overall approximation for the asymptotic form for D_q is:

$$D_q = \frac{\log p_{min}}{\log f_{min}} - \frac{b'}{q-1},$$

Which is consistent with

$$D_q = \frac{q}{q-1}\frac{\log p_{min}}{\log f_{min}} - \frac{b''}{q-1} = \frac{q}{q-1}a - \frac{1}{q-1}b''$$

for the appropriate choice of b''. Thus the inclusion of this more general type of degeneracy, where multiple branches have the same x value, changes the value of one of the terms in the asymptotic expansion, but does not affect the final asymptotic value that D_q approaches, and does not affect the functional form of that expansion. The asymptotic form is still the same as in equation A.2.

Now, let's refine this estimate. It is easiest to do this by modifying τ. Since degeneracy in x does not change the functional form of the expansion, but makes the algebra prohibitive, consider a set with all unique x values, with the understanding that degenerate x values may change the values of the coefficients in our first approximation. Labeling our first approximation from equation A.2 as $\tau = \tau_0(q)$, we now define our next approximation as $\tau = \tau_0(q) + \delta(q)$, with δ being a small (non-constant) correction. Plugging this back in to our original mother-daughter formula shown in equation A.1, and including one more term (the next-largest term: the $i = 2$ term), we find

$$n_{min}\frac{p_{min}^q}{f_{min}^{aq-b+\delta}} + n_2\frac{p_2^q}{f_2^{aq-b+\delta}} \approx 1$$

$$n_{min}\frac{p_{min}^q}{f_{min}^{aq-b}}f_{min}^{-\delta} + n_2\frac{p_2^q}{f_2^{aq-b+\delta}} \approx 1$$

Since $\tau_0 = aq-b$ was constructed by effectively solving $n_{min}p_{min}^q/f_{min}^\tau = 1 \rightarrow n_{min}p_{min}^q/f_{min}^{aq-b} = 1$, the first term reduces very cleanly.

$$f_{min}^{-\delta} + n_2\frac{p_2^q}{f_2^{aq-b+\delta}} \approx 1$$

Using $p^q = \exp(q \log p) = \exp(q \log f \log p / \log f) = f^{q \log p / \log f}$, this can be written as:

$$f_{min}^{-\delta} + n_2 f_2^{(\frac{\log p_2}{\log f_2} - a)q + b - \delta} \approx 1.$$

Since δ is a small correction, we can ignore it in the second term; the leading order contribution is from the first term. It is then only a few steps of algebra to solve for δ.

$$\delta = \frac{\log\left(1 - n_2 f_2^b f_2^{(\frac{\log p_2}{\log f_2} - a)q}\right)}{-\log f_{min}}$$

But of course it is sort of silly to retain this kind of precision in an estimate, so notice that $\log(1 + \zeta) \approx \zeta$, allowing us to simplify our expression to:

$$\delta = -\left(\frac{n_2 f_2^b}{|\log f_{min}|}\right)\left(f_2^{\frac{\log p_2}{\log f_2} - a}\right)^q$$

So that we have shown that the form of our next correction to τ is $c\,d^q$, where c and d could be found numerically. Combining with equation A.2, this gives our overall asymptotic form as:

$$\tau \approx aq - b - c\,d^q + \ldots$$

Keep in mind that this is a correction to τ, not D_q. Do not forget that $D_q = \tau/(q-1)$, giving our asymptotic form for D_q as:

$$D_q \approx a\frac{q}{q-1} - \frac{b}{q-1} - c\frac{d^q}{q-1} + \cdots$$

Notice that when seeking the asymptotic limit that a D_q curve may be approaching as $q \to -\infty$, this result shows that the slope of $\tau(q)$ is a very direct estimate for $D_{-\infty}$ (since

136

$a = \frac{\log p_{min}}{\log f_{min}} = D_{-\infty}$). This allows us to find $D_{-\infty}$ very easily in the convergence tests first described in section 2.2.

CALCULATION OF D_q OF THE LOGISTIC MAP AT $\lambda = 4$

Often, the logistic map [12] given by $x_{n+1} = \lambda x_n(1 - x_n)$ is analyzed at the period-∞ accumulation point near $\lambda = 3.57$. This is convenient because at that accumulation point, the infinite cascade of period doubling has produced a Cantor set structure, which is easily modeled as a two-scale Cantor set. The model is to treat the set as a Cantor set that is spatially rescaled by Feigenbaum's α and α^2 [1], with half the probability in each interval. There are some complications to the modeling [13], but they do not affect the asymptotic values, which are $\frac{\ln 2}{\ln \alpha}$ and $\frac{\ln 2}{2 \ln \alpha}$ (found by taking $\ln p / \ln f$ as described in appendix A). However, during our experiments, we found that slight rounding errors in the value of λ and in the actual values of the iterates x_n led to a map that was in the chaotic region just beyond the period-∞ accumulation point. This may not have a strong effect on the most frequently visited parts of the attractor, but it does change the probabilities in the least visited areas of the attractor, which makes the $D_{-\infty}$ asymptotic value incorrect. Accordingly, we analyzed the logistic map at $\lambda = 4$, which is as deep into the chaotic region as possible.

In a startlingly clever 1977 paper, Grossmann and Thomae [29] found (as a sidebar) the probability distribution associated with the logistic map at $\lambda = 4$. Without going into full detail, their procedure was the following. They showed that the logistic map at $\lambda = 4$ is conjugate to the tent map. The tent map is easily analyzed, and the invariant probability distribution associated with it can be found. This is the probability distribution describing where iterates of the map will land. The probability distribution can then be transformed through the conjugacy relation, to give the probability distribution for the logistic map at

$\lambda = 4$:

$$p(x) = \frac{1}{\pi\sqrt{x(1-x)}}.$$

Now we will calculate the multifractal spectrum for this probability distribution.

It is simplest to find this spectrum by finding $f(\alpha)$ first, much as we did for our example $p(x) = \beta x^{1-\beta}$ at the beginning of section 1.2.3. Consider a box of size ℓ centered on a point x. That box will contain an amount of probability given by $p = \int_x^{x+\ell} \frac{dx}{\pi\sqrt{x(1-x)}}$. So long as x is not 0 or 1, that probability will scale linearly with ℓ. This means that for the 1 dimensional interval $x \in (0, 1)$, the probability in a small box scales like $p \sim \ell^1$. Thus, matching our definitions in section 1.2.3, we have just found that $f(1) = 1$ for this set. If x is 0 or 1, it is quite simple to show that $p \sim \ell^{1/2}$ as $\ell \to 0$. This means that for the zero dimensional set of points $x = 0, 1$, the probability in a small box scales like $\ell^{1/2}$, giving us another point on our $f(\alpha)$ spectrum: $f(\frac{1}{2}) = 0$. Thus, much like in equation 1.4, our $f(\alpha)$ spectrum is given by:

$$f(\alpha) = \begin{cases} 0 & \alpha = \frac{1}{2} \\ 1 & \alpha = 1 \\ undefined & \text{elsewhere.} \end{cases}$$

However, our methods do not deal with $f(\alpha)$, but with D_q, so for any useful comparison, we must transform this spectrum to a D_q spectrum. Since our $f(\alpha)$ is so highly discontinuous, our Legendre transform (described in equation 1.6) must be the fully general Legendre transform, rather than the more common minimization performed with a derivative. Recall equation 1.6:

$$\tau = \min_\alpha \left(\alpha q - f(\alpha) \right).$$

In our case, since f is only defined at two values of α, this becomes

$$\tau = \min_{\alpha=1,\,1/2} (\alpha q - f(\alpha))$$

$$= \min\left(q - 1,\; \frac{1}{2}q - 0\right).$$

The two options are equal when $q = 2$. To the right of $q = 2$, the minimum is $q/2$, and to the left of $q = 2$, the minimum is $q - 1$. Thus

$$\tau(q) = \begin{cases} q - 1 & q < 2 \\[1mm] \frac{q}{2} & q \geq 2, \end{cases}$$

giving $D_q = \tau/(q-1)$ as

$$D_q = \begin{cases} 1 & q < 2 \\[1mm] \frac{q}{2(q-1)} & q \geq 2. \end{cases}$$

This result is plotted in figure B.1. A particularly notable consequence is that for the logistic map at $\lambda = 4$, we expect to observe $D_{-\infty} = 1$. We expect this result to be less prone to rounding problems, since instead of analyzing a Cantor set structure, but having rounding errors move our system into a chaotic region, we have analyzed the chaotic region.
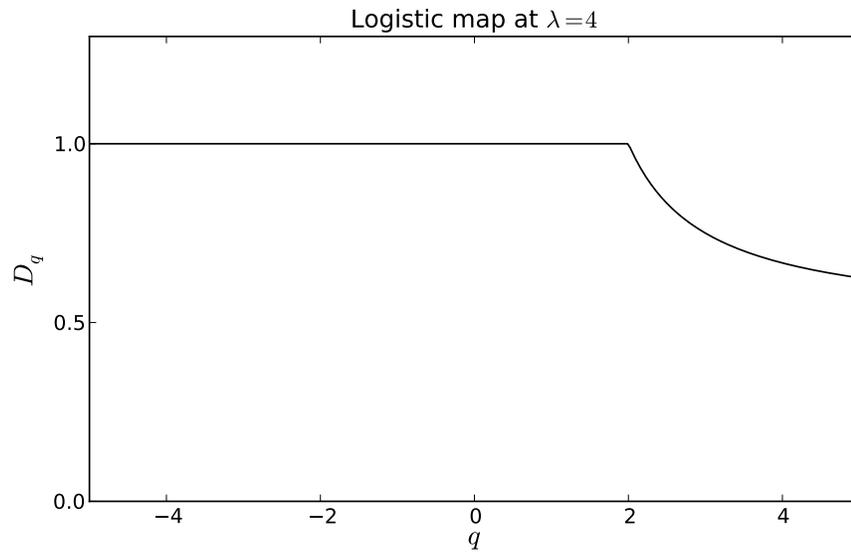
FIGURE B.1. D_q for the logistic map at $\lambda = 4$

APPENDIX C

CALCULATION OF D_q OF THE UNION OF TWO DISJOINT CANTOR SETS

Consider two two-scale Cantor sets near each other (set A and set B). Let each set be constructed from an interval that started with unit length, and place the two sets end to end, with a large enough gap between them that during the process of forming covers of the sets in order to determine the overall dimension, no box enclosing part of one of the two sets has to enclose any part of the other set (in other words, the two sets are far enough apart not to interfere with each other). Let us now find the dimension of the set $A \cup B$. Imagine constructing Γ by examining successive stages of construction. We're going to try to rebuild a mother-daughter formula. For clarity, our notation is that set A, if it were alone in the universe, would assign a fraction p_{AL} of the probability to the leftmost f_{AL} fraction of the length, and a probability p_{AR} to the rightmost f_{AR} of the length (similarly for set B). That is to say that the probabilities I refer to are the probabilities that, if we add up all four of them, add up to 2. They are not renormalized.

At the first stage, we should have

$$\Gamma_1 = \frac{(\frac{1}{2}p_{AL})^q}{f_{AL}^{\tau}} + \frac{(\frac{1}{2}p_{AR})^q}{f_{AR}^{\tau}} + \frac{(\frac{1}{2}p_{BL})^q}{f_{BL}^{\tau}} + \frac{(\frac{1}{2}p_{BR})^q}{f_{BR}^{\tau}}$$

$$= \frac{1}{2^q} \left(\frac{p_{AL}^q}{f_{AL}^{\tau}} + \frac{p_{AR}^q}{f_{AR}^{\tau}} + \frac{p_{BL}^q}{f_{BL}^{\tau}} + \frac{p_{BR}^q}{f_{BR}^{\tau}} \right)$$

$$= \frac{1}{2^q} \left(\tilde{\Gamma}_A + \tilde{\Gamma}_B \right)$$

Those $\tilde{\Gamma}$ functions are the prototypical mother-daughter formulae that we would have used if we were examining set A or B alone. The factor of one half is because half the probability

is sent to one set, half to the other. This half will later be shown to be irrelevant, because it doesn't affect how the sets are constructed, it only affects how many points go to each set.

The next stage will be

$$\Gamma_2 = \left(\frac{(\frac{1}{2}p_{AL}p_{AL})^q}{(f_{AL}f_{AL})^\tau} + \frac{(\frac{1}{2}p_{AL}p_{AR})^q}{(f_{AR}f_{AR})^\tau} + \frac{(\frac{1}{2}p_{AR}p_{AL})^q}{(f_{AR}f_{AL})^\tau} + \frac{(\frac{1}{2}p_{AR}p_{AR})^q}{(f_{AR}f_{AR})^\tau} \right)$$

$$+ \left(\frac{(\frac{1}{2}p_{BL}p_{BL})^q}{(f_{BL}f_{BL})^\tau} + \frac{(\frac{1}{2}p_{BL}p_{BR})^q}{(f_{BR}f_{BR})^\tau} + \frac{(\frac{1}{2}p_{BR}p_{BL})^q}{(f_{BR}f_{BL})^\tau} + \frac{(\frac{1}{2}p_{BR}p_{BR})^q}{(f_{BR}f_{BR})^\tau} \right)$$

Just like for the single set, this factors very cleanly.

$$\Gamma_2 = \frac{1}{2^q} \left(\tilde{\Gamma}_A^2 + \tilde{\Gamma}_B^2 \right)$$

Note that the factor of one half is not squared, even though the individual $\tilde{\Gamma}$s are. Clearly this trend will continue, so that after N stages, we should have

(C.1)
$$\Gamma_N = \frac{1}{2^q} \left(\tilde{\Gamma}_A^N + \tilde{\Gamma}_B^N \right)$$

In the original mother-daughter formulation, described in section 1.2.2, we found that the partition function was given by $\Gamma_N = \tilde{\Gamma}^N$, which was only finite when it was exactly 1, so that we could solve $\tilde{\Gamma} = 1$ to find where the transition to ∞ happened. The extension of that picture would be to examine our new partition function from equation C.1 as $N \to \infty$. In this case, the partition function does not equal 1 along the boundary between the regions where $\Gamma \to 0$ and $\Gamma \to \infty$. However, if we start below the transition and gradually raise τ, we start with both terms zero, and as soon as either one of the individual partition functions crosses the boundary that made it diverge to infinity, the entire partition function has also switched to infinity.

This means that we can completely ignore the 2^q and examine

$$\Gamma'_N = \max\left\{\tilde{\Gamma}^N_A, \tilde{\Gamma}^N_B\right\}.$$

Along the boundary that separates the $\Gamma \to 0$ region from the $\Gamma \to \infty$ region, Γ' may not equal Γ, but the important fact is that the boundary between the two regions is the same for our new Γ' as it is for Γ.

To find that boundary, find the individual solutions $\tau_A(q)$ and $\tau_B(q)$, which would be the multifractal spectra of either set when it was alone. For either set alone, $\Gamma_{A,B} \to 0$ for $\tau < \tau_{A,B}(q)$ and $\Gamma_{A,B} \to \infty$ for $\tau > \tau_{A,B}(q)$. Thus, for the combined set, $\Gamma \to 0$ when $\tau < \tau_A(q)$ and $\tau < \tau_B(q)$, and $\Gamma \to \infty$ as soon as $\tau > \tau_A(q)$ or $\tau > \tau_B(q)$, so that the curve separating the $\Gamma \to 0$ region from the $\Gamma \to \infty$ region is the curve formed by taking the $\tau(q)$ curve that has the smallest τ value. In other words,

$$\tau(q) = \min\left\{\tau_A(q), \tau_B(q)\right\}.$$

See figure C.1 for an illustration of the $q\tau$ plane showing which $\tau(q)$ curve to select.

So τ is piecewise equal to the solutions we had for the individual sets. Note that these τ curves quickly approach straight lines, so that intersections away from $\tau = 0$ are rare. This means that usually we will have τ equal the solution for one set when $q < 1$ and τ equal the solution for the other set when $q > 1$. It could even be that one $\tau(q)$ curve is always greater than or equal to the other, so that there is not even a switch between the two solutions, and $\tau(q)$ for the combined set is simply equal to $\tau(q)$ for one of the individual sets.

For our convergence tests, we will be looking at $q \to -\infty$, so the smallest τ in that region will be the $\tau(q)$ with the steepest slope (so the largest $D_{-\infty}$ estimate). Thus $D_{-\infty}$ for two
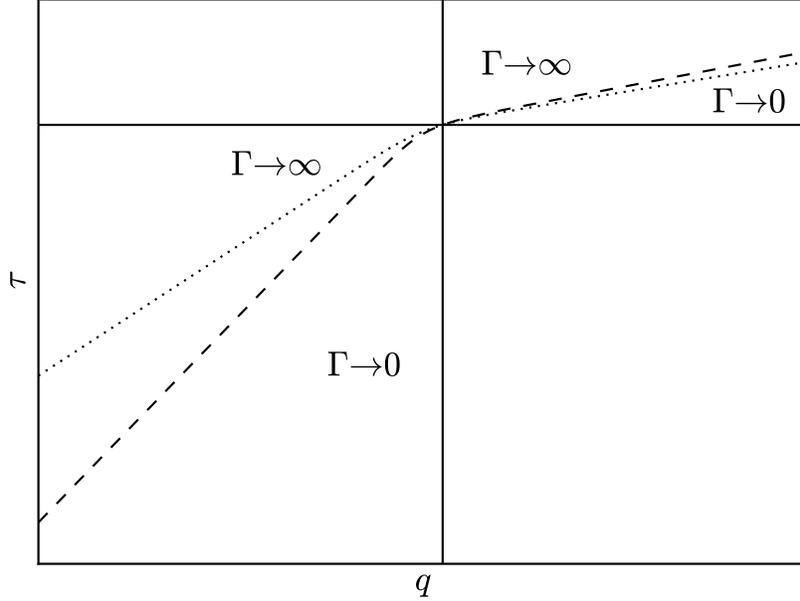
FIGURE C.1. The zero and infinite regions of the $q\tau$ plane, with two sample $\tau(q)$ curves drawn, corresponding to the spectra of each of the two sets when alone. The vertical line that separates the space is $q = 1$. The horizontal line is $\tau = 0$. According to the analysis in this section, when the two sets are considered together, we should select the dashed line to the left of the vertical divider, and the dotted line to the right of the divider.

sets is simply the $D_{-\infty}$ which happens to be largest. There is undoubtedly a quick way of

arriving at this conclusion alone, but now we have a fairly rigorous justification of a method

for finding the whole $\tau(q)$ curve, not just the asymptotic result.

All this boils down to saying that we should look at $\tau(q)$ for each of the two sets alone,

and use the lower of the two curves.