Densities of Self-Similar Measures on the Line

Robert S. Strichartz, Arthur Taylor and Tong Zhang

CONTENTS

Strichartz was supported in part by the National Science Foundation, Grant DMS-9303718. Taylor and Zhang were supported by the National Science Foundation's Research Experiences for Undergraduates Program.

We describe algorithms to compute self-similar measures associated to iterated function systems (i.f.s.) on an interval, and more general self-replicating measures that include Hausdorff measure on the attractor of a nonlinear i.f.s. We discuss a variety of error measurements for these algorithms. We then use the algorithms to study density properties of these measures experimentally. By *density* we mean the behavior of the ratio $\mu(B_r(x))/(2r)^\alpha$ as $r \to 0$, were α is an appropriate dimension. It is well-known that a limit usually does not exist. We have found an intriguing structure associated to these ratios that we call *density diagrams*. We also use density computations to approximate the exact Hausdorff measure of the attractor of an i.f.s.

1. INTRODUCTION

We study a large class of measures associated to an *iterated function system* (i.f.s.), which is just a finite set of one-to-one continuous maps $S_j : [0,1] \to [0,1]$, where $j = 1, \dots, m$. (Many of the ideas presented here can be extended to compact subsets of $\mathbb{R}^n$, but there are many places where we take advantage of the simplifications that are special to one dimension.) It is common to impose a contractivity hypotheses on the maps, such as $|S_j x - S_j y| \le \rho|x-y|$ for $\rho < 1$, but this is unnecessarily restrictive. For the most part we will assume that the maps are *nonoverlapping*, meaning that the images $S_j I$ have disjoint interiors.

A measure μ on $[0,1]$ is said to be *self-similar* with respect to the i.f.s. if it satisfies an identity

$$\mu = \sum_{j=1}^{m} p_j \mu \circ S_j^{-1} \tag{1.1}$$

for some positive *weights* p_j that satisfy the probability condition $\sum p_j = 1$ [Hutchinson 1981]. More

1058-6458/96 \$0.50 per page

generally, we will consider *self-replicating* measures, those that satisfy an identity

$$\mu = \sum_{j=1}^{m} (p_j \mu) \circ S_j^{-1}, \tag{1.2}$$

where the weights p_j are now allowed to be positive functions. More precisely, (1.2) means that

$$\mu(A) = \sum_{j=1}^{m} \int_{S_j^{-1}A} p_j \, d\mu$$

for any Borel set A, or, more generally,

$$\int f \, d\mu = \sum_{j=1}^{m} \int (f \circ S_j) p_j \, d\mu$$

for any integrable function f.

Such measures are investigated in [Barnsley et al. 1989] under the probability hypothesis

$$\sum p_j(x) \equiv 1,$$

but we do not wish to make this assumption. One important example where this hypothesis does not hold is the Hausdorff measure on the attractor K of the i.f.s. (the unique closed set such that $K = \bigcup S_j K$). If the maps S_j are assumed to be C^1 but not necessarily linear, the α-dimensional Hausdorff measure on K satisfies (1.2) with $p_j(x) = |S_j'(x)|^\alpha$, where α is the Hausdorff dimension of K.

It is perhaps useful to think of a self-replicating identity (1.2) as something analogous to a differential equation. Just as there are different types of differential equations, so there are different types of self-replicating identities. Our restriction to a one-dimensional domain is analogous to considering o.d.e.'s. The analog of a linear o.d.e. would be to restrict to a linear i.f.s. The more restrictive form (1.1) is analogous to autonomous o.d.e.'s (or constant-coefficient o.d.e.'s in the linear case).

The first problem we consider is to develop effective algorithms for numerical approximation of solutions to a self-replicating identity. The algorithm we propose may be thought of as the analog of Euler's method for numerical solution of o.d.e.'s. It is a straightforward implementation of what the self-replicating identity says, and it seems to work well enough for our purposes, especially in the special case (1.1) that we take up first in Section 2. In this section we set the framework for the general problem: "How do we compute an approximation to a measure, and how do we estimate the error?" The answer we seek involves a finite set of statements "$\mu(J)$ equals (or approximately equals) certain values", where J ranges over a nonoverlapping set of intervals $\mathcal{J}$ that cover the support of μ. The main idea in our approach is that we do not attempt to specify the collection $\mathcal{J}$ in advance, but rather compute it adaptively for the particular measure, along with the computations of $\mu(J)$. We provide a careful and perhaps overly pedantic exposition of these ideas in Section 2.

In Section 3 we consider the more general form (1.2). The question of existence and uniqueness (up to a normalization constant) is not completely settled. It is convenient to work with a projectivized version $\lambda\mu = \sum (q_j \mu) \circ S_j^{-1}$ in which the eigenvalue λ is unknown, as well as the measure. We are able to prove that a solution exists and the eigenvalue is unique under suitable hypotheses, but uniqueness of the measure seems more difficult. We are able to adapt the algorithms to this set-up, but we no longer have an effective way to estimate the error (indeed, if the solution is not unique it is not even clear what the error is). Nevertheless, we have done some computations using our algorithm that seem reasonably reliable, and form the basis of some of our subsequent experiments. Another approach to the material in this section (and some of the computations in Section 5) would be to use the thermodynamic formalism of Bowen (see [Bowen 1975] and [Ruelle 1983]). This would certainly lead to different algorithms, and it is not clear how well this approach would work.

The remainder of the paper is devoted to the investigation of densities of our measures. Let $B_r(x)$ denote the ball of radius r about x—in our case the interval $[x - r,\, x + r]$. The general density problem is the behavior of the ratio $\mu(B_r(x))/(2r)^\alpha$ as

$r \to 0$. Since it is known [Falconer 1985] that the limit does not usually exist, we seek various substitutes. Bedford and Fisher [1992] consider an average in r, which they call a *second-order density*, and this approach has been widely investigated [Falconer 1992; Patzschke and Zähle 1993]. From our point of view, this average is too crude. In the linear case, a more recent approach [Bandt 1992; Graf 1993] suggests that a much richer structure exists. In Section 4 we propose such a structure, which we call a *density diagram*, defined to be essentially the superposition of all graphs of all functions $\mu(B_r(x))/(2r)^\alpha$ (on a logarithmic scale) as x varies over the attractor. For simple examples, such as the usual Cantor measure, the density diagram reveals a striking self-similar structure, which we are able to fully explain. For more complicated examples the density diagram appears more chaotic, but we are able to "decipher" it in some cases as a superposition of self-similar sets.

In Section 5 we discuss the problem of the correct normalization of Hausdorff measure. Since our self-replicating identities are homogeneous, the solutions can only be determined up to a constant multiple. We usually normalize our measures to be probability measures. However, the definition of Hausdorff measure is exact, so the construction of Hausdorff measure on the attractor K by solving the self-replicating identity is incomplete without the determination of $\mu(K)$. Fortunately, there is a well-known theorem that comes to the rescue: if $0 < \mu(K) < \infty$, the upper convex density

$$\limsup \frac{\mu(I)}{|I|^\alpha}$$

as $|I| \to 0$ and $x \in I$ tends to one for μ-almost every x [Falconer 1985]. The algorithm we use involves computing this $\limsup$ for the incorrectly normalized Hausdorff measure, and using the result to correct the normalization. We illustrate this method by computing the Hausdorff measure of the attractor K of an i.f.s. for some families of examples. We also compute the Hausdorff measure for the family of Julia sets of the mapping $z^2 - c$, for $c > 2$ real. Although the accuracy of the method cannot be quantified, the results seem to indicate that the relative measure (Hausdorff measure divided by diameter to the Hausdorff dimension power) has a discontinuity at $c = 2$.

In Section 6 we discuss two other densities and dimensions that we obtain by averaging pointwise quantities. The entropy dimension is

$$\lim_{r \to 0} \frac{\int \log \mu(B_r(x))\, d\mu(x)}{\log r}$$

and the L^2 dimension is

$$\lim_{r \to 0} \frac{\log \int \mu(B_r(x))\, d\mu(x)}{\log r};$$

in each case there are corresponding densities. (It may seem strange to call this latter an L^2 dimension, but in fact it is quadratic in μ. For a full justification, see [Strichartz 1993b].) Both these densities are easy to compute using our algorithms, and we present evidence for more regular behavior than has been proved so far.

The algorithms described in this article are presented in schematic form. Of course, they were actually coded in computer programs: see the section on Electronic Availability at the end. The programs were written in C and Pascal and run on a Sun Sparcstation using Unix during the summer of 1993.

This article contains some conventional mathematical theorems, proofs and definitions, as well as experimental results, conjectures and problems suggested by the experiments. But the conventional mathematical content was all developed in conjunction with the experiments, either to justify the algorithms or to explain the results. In particular, the material in Sections 4 and 5 is almost entirely serendipitous discovery in response to the experimental results.

The reader is encouraged to consult [Barnsley 1988] and [Falconer 1990] for the general theory of iterated function systems and self-similar measures.

2. NUMERICAL APPROXIMATION ALGORITHMS

Let $\mathcal{P}$ denote the space of probability measures on the Borel subsets of $[0,1]$. For each such measure μ, let $g(x) = \mu([0,x])$ denote its distribution function (or integral). Then g is monotone increasing, right-continuous, and $g(1) = 1$. For nonatomic measures, which include all the measures we will work with, g is continuous and $g(0) = 0$. We will call such functions *ramp functions*. It is well-known that the correspondence between μ and g is one-to-one and onto. Thus to know a measure it suffices to know its ramp function.

We will adopt the point of viewthat an algorithm for computing a measure is equivalent to an algorithm for computing its ramp function. (For example, the ramp function gives us an immediate algorithm for choosing a point at random with respect to μ: simply choose a point at random from $[0,1]$ with respect to Lebesgue measure and take the inverse image under the ramp function.) This point of view leads naturally to a family of metrics on $\mathcal{P}$ of the form $\tilde{d}(\mu_1,\mu_2) = d(g_1,g_2)$, where d is a metric on ramp functions. We will use two such metrics, called the *maximum error metric*

$$d_{\max}(\mu_1,\mu_2) = \sup_x |g_1(x) - g_2(x)|$$

and the *average error metric*

$$d_{\text{ave}}(\mu_1,\mu_2) = \int_0^1 |g_1(x) - g_2(x)|\,dx.$$

These are special cases ($p = \infty$ and $p = 1$) of the L^p *error metric*

$$d_{L^p}(\mu_1,\mu_2) = \left(\int_0^1 |g_1(x) - g_2(x)|^p\,dx\right)^{1/p}.$$

It is easy to see that all these metrics for $1 \le p < \infty$ are topologically equivalent, although they are not equivalent metrics, and that $\mathcal{P}$ is complete in all such metrics.

The average error metric is equal to the Hutchinson metric [Hutchinson 1981; Dudley 1966], defined by

$$d_H(\mu_1,\mu_2) = \sup\left|\int_0^1 f\,d\mu_1 - \int_0^1 f\,d\mu_2\right|,$$

where f runs over Lipschitz functions with Lipschitz constant 1. This is a well-known result in probability theory (see [Rachev 1991, p. 28], for example), but we include the short proof for the convenience of the reader.

Lemma 2.1. *Let μ_1 and μ_2 be nonatomic measures in $\mathcal{P}$. Then $d_{\text{ave}}(\mu_1,\mu_2) = d_H(\mu_1,\mu_2)$.*

Proof. It is well-known that f is Lipschitz with Lipschitz constant 1 if and only if $f' \in L^\infty$ with $\|f'\|_\infty \le 1$ (the derivative exists Lebesgue almost everywhere and in the distribution sense). Then, by integration by parts for Stieltjes integrals, we have

$$\int_0^1 f\,d\mu_j = f(1) - \int_0^1 f'(x) g_j(x)dx$$

(since $g_j(0) = 0$ and $g_j(1) = 1$), hence

$$\left|\int_0^1 f\,d\mu_1 - \int_0^1 f\,d\mu_2\right| = \left|\int_0^1 f'(x)(g_2(x) - g_1(x))\,dx\right|,$$

and the result follows by the L^1–L^∞ duality. □

Because the Hutchinson metric comes into many existence and uniqueness theorems for self-similar measures, we will pay attention to the average error metric estimates in our computations. (Later, we will present some existence and uniqueness theorems using the maximum error metric.) Still, we will be more concerned with maximum error estimates, since they give stronger control of errors in the density computations we will be doing. Also, for many of the measures we will be dealing with, there will be a countable union of subintervals of measure zero, called *lakes*, whose total lengths sum to one, and such that we can compute the ramp function exactly on a large subset of lakes. Thus the error in the ramp function will be entirely confined to a set of small Lebesgue measure, and so the average error will automatically be small, even if the ramp function does not do a particularly good job of approximation on the remainder of the interval (which is where most of the interesting behavior is to be found). Thus average errors will tend to be considerably smaller than maximum errors.

To describe our algorithms we introduce the notion of a set of *interval partition data.* This is just a finite collection $\mathcal{J}$ of nonoverlapping subintervals J of $[0,1]$ and a function $\nu(J)$ in $\mathcal{J}$ that satisfies the probability conditions $0 \le \nu(J) \le 1$ and

$$\sum_{J\in\mathcal{J}} \nu(J) = 1.$$

By *nonoverlapping* we mean that the interiors are disjoint; since our measures are nonatomic we do not have to distinguish between open and closed intervals. We say that a measure μ in $\mathcal{P}$ *matches* the interval partition data *exactly* if $\mu(J) = \nu(J)$ for all J, or *approximately with error* ε if

$$\sum_{J\in\mathcal{J}} |\mu(J) - \nu(J)| \le \varepsilon.$$

All our algorithms compute interval partition data for a measure μ either exactly or approximately, and we take for our approximation to μ any measure that matches exactly the computed interval partition data. The simplest way to do this is using linear interpolation: the ramp function is taken to be piecewise linear on each of the intervals J and piecewise constant on the intervals that form the complement of $\bigcup_{J\in\mathcal{J}} J$. The approximating measure is thus just a multiple of Lebesgue measure on each interval J and zero elsewhere. We can easily write a formula for the ramp function of the measure. First, order the intervals J in increasing order, say $J_1, J_2, \dots, J_N$; all storage of $\mathcal{J}$ in files will use this ordering, of course. Write $J_j = [a_j, b_j]$, and set

$$g(b_j) = \sum_{k=1}^{j} \nu(J_k)$$

and $g(x) = g(b_j)$ for $b_j \le x \le a_{j+1}$ (also $g(x) = 0$ for $0 \le x \le a_1$). Finally, for any x in J_j, set

$$g(x) = \frac{x - a_j}{b_j - a_j}\, g(b_j) + \frac{b_j - x}{b_j - a_j}\, g(a_j). \tag{2.1}$$

This *linear interpolation algorithm* is easy to implement, but it does require a search for the interval containing x.

Given a complete state of ignorance about the measure, except for the interval partition data, the linear interpolation algorithm can be justified on philosophical grounds as the choice that introduces the least extraneous information (assuming that a uniform distribution on an interval corresponds to the measure with least information). It would be interesting to investigate whether or not there is any quantitative justification for this choice. In any case, we are usually not in a state of complete ignorance concerning the measure, so we can sometimes make better choices. We will discuss some improved accuracy interpolation algorithms later.

Now consider an i.f.s. $S_1, S_2, \dots, S_m$ on $[0,1]$ with no overlap (the images of $[0,1]$ under S_j have disjoint interior). We say that an interval partition $\mathcal{J}$ is *adapted* to the i.f.s. if for every $J \in \mathcal{J}$ and S_j there exists $J' \in \mathcal{J}$ such that $S_j J \subseteq J'$. Given an adapted interval partition $\mathcal{J}$, we define a new interval partition $\mathcal{J}'$, called the *refinement* of $\mathcal{J}$, by taking $\mathcal{J}'$ to be the collection of intervals $S_j J$ as J varies over $\mathcal{J}$ and $j = 1, \dots, m$. It is easy to see that $\mathcal{J}'$ is an interval partition and that it is also adapted to the i.f.s.

Now suppose μ is a probability measure satisfying a self-similar identity

$$\mu = \sum_{j=1}^{m} p_j \mu \circ S_j^{-1} \tag{2.2}$$

for constant probabilities p_j. To avoid triviality we assume $m \ge 2$ and all p_j positive. In particular, this implies that μ is nonatomic. To see this, consider the atom of largest measure, say x_0. Then $\mu(\{x_0\}) = \sum_{j=1}^{m} p_j \mu(\{S_j^{-1} x_0\})$ implies $\mu(\{S_j^{-1} x_0\}) = \mu(\{x_0\})$ for all j, hence we must have $m = 2$ and x_0 must be the common endpoint of the images $S_1[0,1]$ and $S_2[0,1]$. But, by applying the same reasoning to $S_j^{-1} x_0$, we find $S_j x_0 = x_0$ for

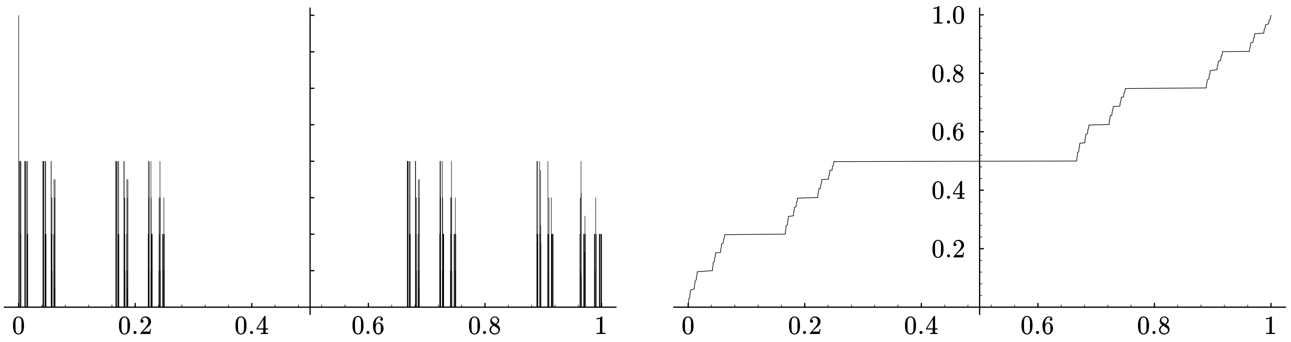

FIGURE 1. Left: a histogram representation of the self-similar measure for the i.f.s. $S_1x = \frac{1}{4}x$, $S_2x = \frac{1}{3}x + \frac{2}{3}$ with $p_1 = p_2 = \frac{1}{2}$. The errors $d_{\max} = 7.6 \times 10^{-6}$ and $d_{\text{ave}} = 3.1 \times 10^{-10}$ were computed between the next-to-last and last iteration of the algorithm. Right: The ramp function for the same measure.

$j = 1, 2$, and this contradicts the assumption that the S_j are one-to-one.

Note that the nonoverlapping assumption means that $A \subseteq [0,1]$ implies

$$\mu(S_j A) = p_j \mu(A).$$

This observation leads to a simple algorithm for passing from a set of interval partition data $(\mathcal{J}, \nu)$ that matches μ exactly, with $\mathcal{J}$ adapted to the i.f.s., to one for the refinement $\mathcal{J}'$.

Algorithm 2.2. Suppose $\mathcal{J}$ is adapted to the i.f.s. and that $\mathcal{J}'$ is its refinement. Given data $(\mathcal{J}, \nu)$ that matches μ exactly, define data $(\mathcal{J}', \nu')$ by

$$\nu'(S_j J) = p_j \nu(J).$$

Then $(\mathcal{J}', \nu')$ also matches μ exactly. Define inductively $(\mathcal{J}_0, \nu_0)$ by $\mathcal{J}_0 = \{[0,1]\}$ and $\nu_0([0,1]) = 1$, and let $(\mathcal{J}_k, \nu_k)$ be obtained from $(\mathcal{J}_{k-1}, \nu_{k-1})$ by the above algorithm.

We call intervals in $\mathcal{J}_k$ *islands of the k-th generation*, and connected components of the complement of $\bigcup_{J \in \mathcal{J}_k} J$ *lakes of the k-th generation*. It is clear that the lakes have measure zero and hence the ramp function is constant on each lake, for any measure that matches the data for the given generation. It is easy to implement the algorithm to generate the data $(\mathcal{J}_k, \nu_k)$ so it terminates when the size of the data $(3m^k)$ approaches the maximum available memory space.

In Figure 1 we display the output of the algorithm for a simple example. We display both a histogram of the measure and the graph of its ramp function. The histogram graphs the probability of each interval in a uniform partition of $[0,1]$ into 4,000 intervals. It is useful for giving a rough picture of the measure, but some of the features are artifacts due to the choice of partition.

We can give a priori estimates for the accuracy of the algorithm based on the following elementary lemma.

Lemma 2.3. (a) *Let μ_1 and μ_2 be measures in $\mathcal{P}$ that match the data $(\mathcal{J}, \nu)$ exactly. Set*

$$\delta_1 = \sup_{J \in \mathcal{J}} \nu(J) \quad \textit{and} \quad \delta_2 = \delta_1 \sum_{J \in \mathcal{J}} |J|,$$

where $|J|$ is the length of the interval. Then

$$d_{\max}(\mu_1, \mu_2) \le \delta_1 \quad \textit{and} \quad d_{\text{ave}}(\mu_1, \mu_2) \le \delta_2.$$

(b) *Suppose μ_1 and μ_2 match the data $(\mathcal{J}, \nu)$ approximately with errors ε_1 and ε_2. Then*

$$d_{\max}(\mu_1, \mu_2) \le \delta_1 + \varepsilon_1 + \varepsilon_2 \tag{2.3}$$

and

$$d_{\text{ave}}(\mu_1, \mu_2) \le \delta_2 + \varepsilon_1 + \varepsilon_2.$$

Proof. (a) On the complement of $\bigcup_{J \in \mathcal{J}} J$, we have $g_1 = g_2$ exactly, so there is no contribution to $d_{\text{ave}}(\mu_1, \mu_2)$. If $x \in J = [a, b]$, then

$$g_1(x) - g_2(x) = \mu_1([a, x]) - \mu_2([a, x])$$

and $[a,x] \subseteq J$, so $|g_1(x) - g_2(x)| \leq \nu(J)$. Thus $d_{\max}(\mu_1,\mu_2) \leq \sup_{J\in\mathcal{J}} \nu(J) = \delta_1$ and $d_{\text{ave}}(\mu_1,\mu_2) \leq \sum_{J\in\mathcal{J}} |J|\,\nu(J) \leq \delta_2$.

(b) In this case we only have the estimate

$$|g_1(x) - g_2(x)| \leq \varepsilon_1 + \varepsilon_2$$

on the complement of $\bigcup_{J\in\mathcal{J}} J$, and if $x \in J$ we have $|g_1(x) - g_2(x)| \leq \nu(J) + \varepsilon_1 + \varepsilon_2$, so we obtain (2.3) and

$$d_{\text{ave}}(\mu_1,\mu_2) \leq \varepsilon_1 + \varepsilon_2 + \sum_{J\in\mathcal{J}} |J|\,\nu(J) \leq \delta_2 + \varepsilon_1 + \varepsilon_2.$$

□

Sometimes $\sum_{J\in\mathcal{J}} |J|\,\nu(J)$ is smaller than δ_2, but it is usually not worth the effort to compute it.

Corollary 2.4. *Let μ satisfy* (2.2) *and let μ_k be any approximations that exactly match the data $(\mathcal{J}_k,\nu_k)$ generated by Algorithm* 2.2. *Then*

$$d_{\max}(\mu,\mu_k) \leq p_{\max}^k$$

and

$$d_{\text{ave}}(\mu,\mu_k) \leq (\delta p_{\max})^k,$$

where $p_{\max} = \max_j p_j$ and $\delta = \sum_j |S_j([0,1])|$.

Proof. By our assumptions, $p_{\max} < 1$. From the algorithm it is clear that $\delta_1 = p_{\max}^k$ and $\delta_2 = (\delta p_{\max})^k$ for $(\mathcal{J}_k,\nu_k)$, and both μ and μ_k match $(\mathcal{J}_k,\nu_k)$ exactly. □

Notice that we are not claiming that the sizes of the intervals in the partitions $\mathcal{J}_k$ are going to zero. Of course, if I is an interval such that $S_j I \supseteq I$, the equality (2.2) easily implies $\mu(I) = 0$, so this is perhaps an uninteresting generalization. However, we do not require any further contractivity of the i.f.s. beyond the nonoverlapping hypothesis. We can easily turn the estimates in the corollary into an existence and uniqueness proof for solutions of (2.2), but perhaps it is more interesting to obtain existence and uniqueness out of general contractivity estimates for the transformation on $\mathcal{P}$ given by

$$T\mu = \sum_{j=1}^{m} p_j \mu \circ S_j^{-1}.$$

Theorem 2.5. *Assume the i.f.s. is nonoverlapping, $m \geq 2$, and all $p_j > 0$. Then T is contractive in the maximum error metric with contractivity constant $p_{\max}$. In particular, T has a unique fixed point μ, which can be obtained as $\lim_{n\to\infty} T^n\mu_1$ starting with any $\mu_1 \in \mathcal{P}$, and if $d_{\max}(\tilde{\mu}, T\tilde{\mu}) \leq \varepsilon$ for any $\tilde{\mu} \in \mathcal{P}$, then $d(\tilde{\mu},\mu) \leq \varepsilon/(1-p_{\max})$.*

Proof. Order the mappings S_j so that $S_j[0,1] = [a_j,b_j]$ are in increasing order. If $x \in [a_j,b_j]$ and $\tilde{\mu}_1 = T\mu_1$, we have

$$\tilde{g}_1(x) = T\mu_1([0,x]) = \sum_{k=1}^{j-1} p_k + p_j\mu_1(S_j^{-1}[a_j,x]),$$

and $S_j^{-1}[a_j,x]$ is either $[0,S_j^{-1}x]$ or $[S_j^{-1}x,1]$, depending on whether S_j is increasing or decreasing. Thus $\mu_1(S_j^{-1}[a_j,x])$ equals either $g_1(S_j^{-1}x)$ or $1 - g_1(S_j^{-1}x)$. In either case we have

$$|\tilde{g}_1(x) - \tilde{g}_2(x)| = p_j\,|g_1(S_j^{-1}x) - g_2(S_j^{-1}x)|,$$

hence $d_{\max}(T\mu_1,T\mu_2) \leq p_{\max} d_{\max}(\mu_1,\mu_2)$. If x is in the complement of the intervals $[a_j,b_j]$, the above reasoning shows that $\tilde{g}_1(x) = \tilde{g}_2(x)$ exactly. □

We consider next some modifications to Algorithm 2.2. If there is a large spread of values for the probabilities p_j, there will be an even greater spread among the values of $\mu_k(J)$ over different intervals. It can then happen that a shortage of memory space will terminate the algorithms before $p_{\max}^k$ gets sufficiently small. In that case, too much memory is dedicated to very fine information about the measure, while a certain amount of too coarse information keeps the error large. To deal with this problem one could adopt various *consolidation algorithms*. The idea is to set a threshold ε, and whenever $\nu(J)$ drops below ε, to combine J with several other intervals in the partition. We will not describe the procedure in detail since it is fairly straightforward.

One way to improve accuracy at very little cost is to replace the linear interpolation algorithm, which in effect chooses the measure μ_k to match the data $(\mathcal{J}_k,\nu_k)$, with one that exploits the self-similarity of the measure. This is only feasible in the special case when all the transformations S_j are linear. In that case, we know that the exact ramp function g for the measure μ, restricted to any island J in $\mathcal{J}_k$, is equal to a scaled version of the entire ramp function g on $[0,1]$. If $J = [a_j, b_j]$, the horizontal scaling factor is just $b_j - a_j$, while the vertical scaling must be adjusted so that g assumes the correct values at a and b. Note that Algorithm 2.2 computes the exact values for $g(a_j)$ and $g(b_j)$ (except for round-off error, which can be made extremely small). Thus the self-similarity translates into

$$g(x) = g(a_j) + (g(b_j) - g(a_j))\, g\Big(\frac{x - a_j}{b_j - a_j}\Big) \tag{2.4}$$

for x in $[a_j, b_j]$. Since the unknown function g appears on both sides of the identity, we cannot use (2.4) in place of (2.1) directly. However, we can use a two-step procedure, computing first g_1 by (2.1) and then g_2 using (2.4) with g_1 on the right side. Since the error in g_1 is multiplied by $g(b_j) - g(a_j)$, which is bounded by $\delta_1 = \sup_{J\in\mathcal{J}} \nu(J)$, we obtain an error estimate of δ_1^2 for g_2.

We conclude this section with a brief discussion of how we modified the algorithm to handle i.f.s.'s with overlap. The overlaps create problems that we have dealt with in an ad hoc fashion. The resulting computations seem reasonably accurate, but we have no rigorous method to estimate the error. We want to obtain a sequence of interval partition data $(\mathcal{J}_1, \nu_1), (\mathcal{J}_2, \nu_2), \dots$, where each one is determined from the previous one by a procedure we will describe. The partitions will not necessarily be adapted to the i.f.s., and $\mathcal{J}_k$ will not necessarily be the refinement of $\mathcal{J}_{k-1}$. Also, the measure μ will not match the data exactly, and there does not seem to be a realistic way to estimate the error (we can give worst case estimates that are undoubtedly too large). The procedure described can be modified by interspersing consolidation routines.

To simplify the notation, suppose $(\mathcal{J}, \nu)$ is a set of partition data, and that we want to pass to the next set $(\mathcal{J}', \nu')$. We let $\mathcal{J}'$ be the set of all intervals obtained from the collection S_jJ, for $J \in \mathcal{J}$ and $1 \le j \le m$, by the following *splitting procedure*: if two intervals $[a,b]$ and $[c,d]$ overlap, say $a < c < b < d$, replace them by the three intervals $[a,c]$, $[c,b]$ and $[b,d]$. To pass from ν on $\mathcal{J}$ to ν' on $\mathcal{J}'$ we first assign weight $p_j\nu(J)$ to S_jJ and split the weight proportionate to length if S_jJ is split. That is,

$$\nu'([a,b]) = \sum \frac{b-a}{|S_jJ|}\, p_j \nu(J),$$

where the sum is taken over all S_jJ that contain $[a,b]$. (By the construction, if $[a,b]$ in $\mathcal{J}'$ has nontrivial intersection with S_jJ it must be contained in S_jJ.) The splitting in proportion to length introduces an error. On the other hand it represents a least biased approach, in the absence of any information, and it has the advantage of being independent of the order of splitting for intervals that are split more than once.

We applied this algorithm to the famous example of Bernoulli measures for the i.f.s. $S_1x = \rho x$, $S_2x = \rho x + 1 - \rho$ with $\frac12 < \rho < 1$, and weights $p_1 = p_2 = \frac12$. Figure 2 shows histograms of the computed measure for two values of ρ. The value $\rho = \frac12(\sqrt5 - 1)$ used on the left is far from typical, since ρ^{-1} is the golden ratio, the simplest P.V. number. (Recall that a *Pisot–Vijayaraghavan number*, or P.V. number for short, is an algebraic integer of absolute value greater than 1 with all its conjugates having absolute value less than 1.) See [Lau 1992; 1993] for some interesting properties of this measure and related examples. The histogram reveals striking details. In this case there is an alternate procedure for generating interval partition data that μ matches exactly. We define three transformations:

$$\begin{aligned} T_0x &= S_0S_0x = \rho^2x,\\ T_1x &= S_1S_0S_0x = S_0S_1S_1x = \rho^3x + \rho^2,\\ T_2x &= S_1S_1x = \rho^2x + 1 - \rho^2, \end{aligned}$$

which satisfy the nonoverlapping condition (note that $\rho^2 = 1 - \rho$), with attractor equal to the whole unit interval. The defining identity for μ yields,

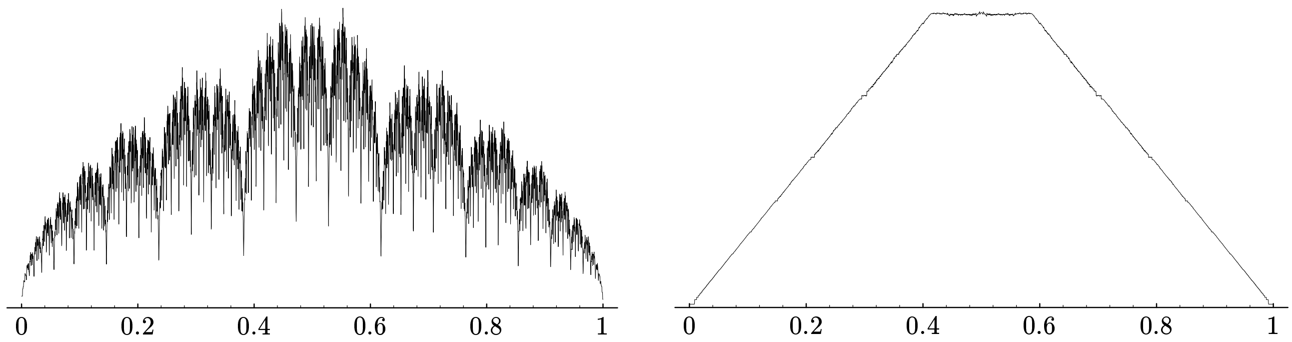

FIGURE 2. Histograms of the Bernoulli measure for the i.f.s. with overlaps $S_1x = \rho x$, $S_2x = \rho x + 1 - \rho$, with $p_1 = p_2 = \frac{1}{2}$. On the left we take $\rho = \frac{1}{2}(\sqrt{5} - 1)$ (with $d_{\max} = 3.2 \times 10^{-5}$ and $d_{\text{ave}} = 4.6 \times 10^{-6}$); the identities (2.5) can be clearly visualized in this case. On the right we take $\rho = .705$ (with $d_{\max} = 6.8 \times 10^{-5}$ and $d_{\text{ave}} = 1.1 \times 10^{-6}$). Note that .705 is close to $1/\sqrt{2}$, for which value of ρ the measure is absolutely continuous and the histogram is the piecewise linear function this one resembles.

after iteration, the following nine identities (for arbitrary Borel sets A in $[0, 1]$):

$$\begin{aligned}
\mu(T_0T_0A) &= \tfrac{1}{4}\mu(T_0A)\\
\mu(T_0T_1A) &= \tfrac{1}{4}\mu(T_1A)\\
\mu(T_0T_2A) &= \tfrac{1}{2}\mu(T_1A)\\
\mu(T_1T_0A) &= \tfrac{1}{8}\mu(T_0A) + \tfrac{1}{4}\mu(T_1A)\\
\mu(T_1T_1A) &= \tfrac{1}{4}\mu(T_1A)\\
\mu(T_1T_2A) &= \tfrac{1}{8}\mu(T_2A) + \tfrac{1}{4}\mu(T_1A)\\
\mu(T_2T_0A) &= \tfrac{1}{2}\mu(T_1A)\\
\mu(T_2T_1A) &= \tfrac{1}{4}\mu(T_1A)\\
\mu(T_2T_2A) &= \tfrac{1}{4}\mu(T_2A).
\end{aligned} \tag{2.5}$$

It follows from this that μ assigns equal probability to the three intervals $[0, \rho^2]$, $[\rho^2, \rho]$ and $[\rho, 1]$ that make up $\mathcal{J}_1$. We then choose $\mathcal{J}_k$ inductively to be the refinement (for T_0, T_1, T_2) of $\mathcal{J}_{k-1}$, and the nine equations above give an algorithm for computing ν_k on $\mathcal{J}_k$ to match μ exactly.

There is a sense in which the identities (2.5) are analogous to differential equations of second order. An extensive computer search conducted by Alan Ho during the summer of 1994 (after the first draft of this paper was completed) turned up one other P.V. number, namely the root of $\rho^3 - \rho^2 + 2\rho = 1$, or approximately 0.56984, with an analogous set of second order identities (in this case 81 equations involving 9 transformations). The search failed to find such identities for any of the other P.V. numbers defined by low-order polynomials. Therefore, such identities are either rare or extremely large.

3. VARIABLE WEIGHTS AND HAUSDORFF MEASURE

In this section we want to consider the more general self-replicating identity

$$\mu = \sum_{j=1}^{m} (p_j\mu) \circ S_j^{-1}, \tag{3.1}$$

where the weights $p_j(x)$ are variable functions. We cannot expect there to exist solutions without imposing some consistency condition on the weights, analogous to the probability condition for constant weights. In [Barnsley et al. 1989] the condition

$$\sum_{j=1}^{m} p_j(x) \equiv 1 \tag{3.2}$$

was imposed, but this is too restrictive for our purposes. Instead, we reformulate the problem as an eigenvalue problem and projectivize the weights: given *projective weights* $q_j(x)$, we look for solutions of

$$\lambda\mu = \sum_{j=1}^{m} (q_j\mu) \circ S_j^{-1} \tag{3.3}$$

for μ in $\mathcal{P}$ and $\lambda > 0$. Any solution of (3.3) yields a solution of (3.1) for $p_j = \lambda^{-1} q_j$, and it turns out that we don't have to put any consistency conditions on q_j to solve (3.3). We will, however, impose the following assumptions:

Hypotheses 3.1. (a) The i.f.s. is nonoverlapping.
(b) The maximum length of intervals in $\mathcal{J}_k$ tends to zero as $k \to \infty$.
(c) The projective weights q_j are continuous.
(d) $m \geq 2$ and all the q_j are positive.

Note that (c) and (d) imply the existence of positive numbers $q_{\min}$ and $q_{\max}$ such that

$$0 < q_{\min} \leq q_j(x) \leq q_{\max} < \infty$$

for all x and j. Since all solutions to (3.3) must be supported on the attractor K of the i.f.s., it suffices to have q_j defined on K. Of course, by a standard extension theorem, we can assume without loss of generality that the q_j are defined on the whole interval $[0, 1]$, since continuity and the existence of bounds $q_{\min}$ and $q_{\max}$ can be preserved in the extension.

The following counterexamples to uniqueness justify the assumptions of continuity and positivity. In both cases (3.2) is satisfied, by the way. We take the simple i.f.s. consisting of

$$S_1 x = \tfrac{1}{2} x \quad \text{and} \quad S_2 x = \tfrac{1}{2} x + \tfrac{1}{2}.$$

Then Lebesgue measure μ_1 solves the constant-weight self-similar identity with $p_1 = p_2 = \frac{1}{2}$, but any other choice of probabilities $\tilde{p}_1 \neq \tilde{p}_2$ yields a measure μ_2 singular with respect to μ_1. That means there exists a disjoint decomposition $[0, 1] = A_1 \cup A_2$ with $\mu_1(A_2) = \mu_2(A_1) = 0$. We then may define

$$p_j(x) = q_j(x) = \begin{cases} \frac{1}{2} & \text{if } x \in A_1, \\ \tilde{p}_j & \text{if } x \in A_2. \end{cases}$$

It is obvious from the construction that both μ_1 and μ_2 (and any convex combination of them) satisfy (3.1). Of course, in this example the weight functions are everywhere discontinuous.

For the second example we use the same i.f.s., and we take

$$p_1(x) = 1 - x, \quad p_2(x) = x.$$

It is easy to verify that the delta measures at $x = 0$ and $x = 1$ both satisfy (3.1). The vanishing of the weights in this case provides hiding places for extraneous solutions (presumably there exists a nonatomic solution as well, but we have not been able to find it). It is no coincidence here that the zeroes of the weights lie at the fixed points of the mappings.

We also give a counterexample to uniqueness where the problem is with the i.f.s., not the weights. Let

$$S_1 x = \begin{cases} x & \text{if } 0 \leq x \leq \frac{1}{4}, \\ \frac{1}{4} + \frac{1}{3}(x - \frac{1}{4}) & \text{if } \frac{1}{4} \leq x \leq 1, \end{cases}$$

and let $S_2 x = \frac{1}{2} x + \frac{1}{2}$. Since the interval $[0, \frac{1}{4}]$ is invariant under S_1, if we choose $p_1 \equiv 1$ any measure supported on $[0, \frac{1}{4}]$ will satisfy (3.1).

Theorem 3.2. *Under Hypotheses* 3.1, *there exists a nonatomic solution of* (3.3) *and* λ *is unique.*

Proof. For each k and each $J \in \mathcal{J}_k$, define

$$q_{jk}^{+}(J) = \sup_{x \in J} q_j(x), \quad q_{jk}^{-}(J) = \inf_{x \in J} q_j(x),$$

so that

$$q_{jk}^{-}(J) \leq q_j(x) \leq q_{jk}^{+}(J) \tag{3.4}$$

for $x \in J$. By the integral mean value theorem, any solution of (3.3) must satisfy

$$\lambda \mu(J) = \sum_{S_j J' \subseteq J} q_{jk}(J') \mu(J')$$

for $q_{jk}(J')$ also satisfying the analogue of (3.4). We begin by trying to find approximations to $\mu(J)$ by solving

$$\lambda_k^{\pm} v_k^{\pm}(J) = \sum_{S_j J' \subseteq J} q_{jk}^{\pm}(J') v_k^{\pm}(J') \tag{3.5}$$

for $J \in \mathcal{J}_k$. Since this is a finite-dimensional matrix eigenvalue problem, the Perron–Frobenius theorem implies there is a unique positive solution, normalized by $\sum_{J\in\mathcal{J}_k} v_k^{\pm}(J) = 1$. Because the dependence of the eigenvalues on the matrix is monotone, we have

$$\lambda_k^- \le \lambda \le \lambda_k^+$$
$$\lambda_k^- \le \lambda_{k+1}^- \le \lambda_{k+1}^+ \le \lambda_k^+.$$

But we claim even more, namely, that $\lambda_k^+ - \lambda_k^-$ goes to zero as $k \to \infty$. The reason for this is that by hypotheses (b) and (c) the differences $q_{jk}^+ - q_{jk}^-$ go to zero uniformly as $k \to \infty$. This alone is not enough to conclude what we want, because the size of the matrix ($m^k \times m^k$) also increases as $k \to \infty$. The key observation is that each of the equations (3.5) involves only m nonzero terms, and this is independent of k. Thus the variational characterization of the eigenvalue shows that it depends continuously on the coefficients, uniformly in k. This shows that $\lim_{k\to\infty} \lambda_k^{\pm} = \lambda$, hence λ is unique.

Once we have the solution of (3.5), we can solve a related self-similar identity for a measure $\mu_k^{\pm}$ that is required to satisfy $\mu_k^{\pm}(J) = v_k^{\pm}(J)$ and

$$\lambda_k^{\pm}\mu_k^{\pm} = \sum (q_{jk}^{\pm}\mu_k^{\pm}) \circ S_j^{-1}, \tag{3.6}$$

with the obvious notation $q_{jk}^{\pm}$ for the function that assumes the constant value $q_{jk}^{\pm}(J)$ on J. The functions $q_{jk}^{\pm}$ may not be continuous, but this does not matter because the correct way to think of (3.6) is as a matrix version of (2.2) for the family of measures $\mu_k^{\pm}$ restricted to J (as J varies over $\mathcal{J}_k$). See [Mauldin and William 1988], [Edgar and Mauldin 1992] or [Strichartz 1993b] for a proof of existence (in these references the i.f.s. is assumed contractive, but the proof of Theorem 2.5 can be adapted to the general case).

Next we pick a subsequence of μ_k^+ that converges weakly to a measure μ in $\mathcal{P}$. We claim that μ satisfies (3.3). To see this, choose any continuous function f. Then

$$\left| \int f \circ S_j (q_{jk}^+ - q_j)\, d\mu_k \right| \le c \|q_{jk}^+ - q_j\|_\infty$$

for c independent of k, so the limit along the subsequence of $\int f \circ S_j q_{jk}^+\, d\mu_k$ exists and equals the limit of $\int f \circ S_j q_j\, d\mu_k$, which equals $\int f \circ S_j q_j\, d\mu$ by weak convergence. But since

$$\lambda_k^+ \int f\, d\mu_k = \sum_{j=1}^m \int f \circ S_j q_{jk}^+ d\mu_k$$

by (3.6), we can take the limit along the subsequence to obtain

$$\lambda \int f\, d\mu = \sum_{j=1}^m \int f \circ S_j q_j\, d\mu$$

which is (3.3). □

Problem 3.3. Under what conditions does uniqueness hold for probability measure solutions of (3.3)?

A reasonable conjecture is that uniqueness holds if the functions $q_j(x)$ are Lipschitz continuous. This would be analogous to the theory of o.d.e.'s, where the Peano existence theorem requires only continuity, while the Picard existence and uniqueness theorem requires Lipschitz continuity. It is not hard to modify the proof of existence and uniqueness in [Hutchinson 1981], under the hypothesis that S_j are strictly contractive, to prove existence and uniqueness under the perturbative hypothesis that the Lipschitz constants are very small.

The proof of Theorem 3.2 suggests an algorithm for computing λ and μ approximately, and testing the likelihood of uniqueness. We solve (3.5) for $\lambda_k^{\pm}$ and $v_k^{\pm}$ on $\mathcal{J}_k$ by iteration, for fixed $\mathcal{J}_k$ (chosen based on memory limitations). This gives us approximate interval partition data for perhaps two solutions. We know that $\lambda_k^{\pm}$ must eventually be close to the unique value of λ, and their difference gives a plausible measure of whether we are reasonably close to matching the interval partition data (we have no rigorous estimate for the error). Then we compare $v^{\pm}$; if they are close, this gives plausible evidence that the solution is unique. In practice we do not go through the trouble of computing $q_{jk}^{\pm}(J)$ to be the max and min of q_{jk} on J, but simply evaluate q_{jk} at the endpoints of J. This gives the same result if the functions $q_j(x)$ are

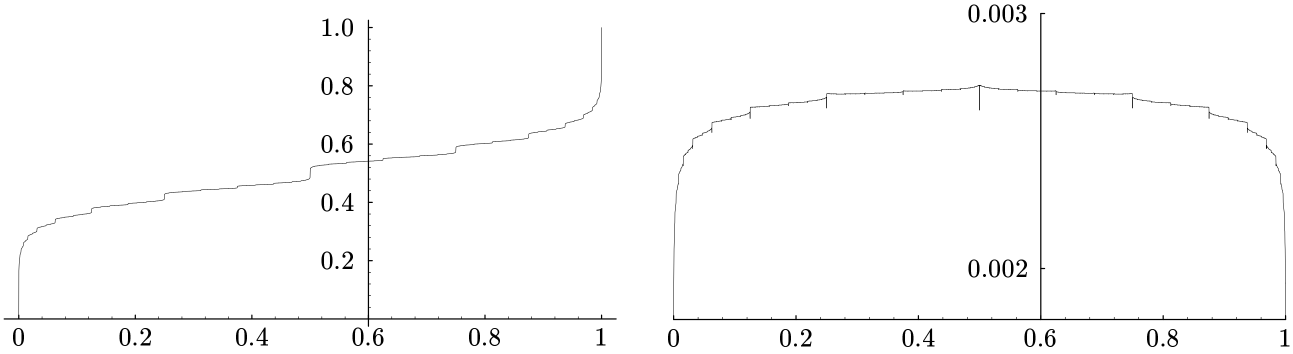

FIGURE 3. Left: The ramp function for the self-replicating measure associated with the i.f.s. $S_1 x = \frac{1}{2}x$, $S_2 x = \frac{1}{2}x + \frac{1}{2}$ with projective weights $q_1 = 1.1 - x$ and $q_2 = .1 + x$. Right: The difference between the ramp functions for approximations to the same measure, using left and right endpoint evaluation of the weights. (We have restricted the vertical range to show the detail; the graph drops abruptly to zero at both endpoints.)

all increasing (decreasing). Figure 3 shows the resulting ramp function for a typical example where we expect uniqueness, and the difference between ramp functions using $v^{\pm}$ for the same example. On the other hand, we ran several tests where we used input that was very close to the second counterexample to uniqueness, and the differences between v^+ and v^- were enormous.

This algorithm also gives a method for approximately computing the Hausdorff measure (up to a constant multiple) for the attractor of a nonoverlapping i.f.s. We need to assume that the mappings S_j are C^1 with derivatives satisfying $0 < |S_j'| < 1$ on the attractor. Then it is known [Falconer 1985] that the Hausdorff measure satisfies

$$\mu = \sum (|S_j'|^\alpha \mu) \circ S_j^{-1}, \tag{3.7}$$

where α is the dimension of the attractor. Since $|S_j'|^\beta$ is decreasing in β under the hypothesis $0 < |S_j'| < 1$, we conclude that the unique eigenvalue $\lambda(\beta)$ associated with the equation

$$\lambda(\beta)\mu = \sum (|S_j'|^\beta \mu) \circ S_j^{-1} \tag{3.8}$$

is decreasing in β. Also $\lim_{\beta\to\infty} \lambda(\beta) = 0$ since $|S_j'|^\beta \to 0$, and $\lambda(0) = m$ (evaluate (3.8) on the entire attractor). Thus α is characterized as the unique solution of $\lambda(\beta) = 1$. Since our algorithm yields an accurate estimate of $\lambda(\beta)$, we merely compute values of $\lambda(\beta)$ until we obtain $\lambda(\beta_1) < 1 < \lambda(\beta_2)$, then divide the interval between β_1 and β_2 (to obtain a speed-up we used linear interpolation to get a value of $\lambda(\beta)$ close to 1) and iterate.

We applied this method to compute the Hausdorff dimension and a multiple of Hausdorff measure (normalized to be a probability measure) for the Julia sets of the mappings $z^2 - a$ for a real and $a \geq 2$. In this range the Julia set lies in the real axis and is the attractor for the i.f.s. given by the two transformations $\pm\sqrt{x+a}$. This attractor lies in the interval $[-b, b]$, where b is the larger root of $b^2 = a + b$, namely $b = \sqrt{a + \frac{1}{4}} + \frac{1}{2}$. As $a \to 2$, the i.f.s. becomes badly behaved; in fact it fails to satisfy $0 < |S_j'| < 1$ for $2 \leq a \leq \frac{1}{4}(5 + 2\sqrt{5})$, but at $a = 2$ the attractor is just the interval $[-2, 2]$. Nevertheless, it is known [Blanchard 1984] that the condition $0 < |S_j'| < 1$ can be restored by taking an equivalent i.f.s. obtained by iteration, and it follows from [Sullivan 1983] that (up to a constant multiple) Hausdorff measure is the unique solution of (3.7) (Sullivan calls such measures *α-conformal*). See also [Aaronson et al. 1993] and [Hofbauer 1992] for related uniqueness results.

Figure 4 shows the histogram of the measure for two typical values of a. The results of these computations are used later in this paper to explore properties of these Julia sets.

We will take up the question of the correct normalization of Hausdorff measures in Section 5.

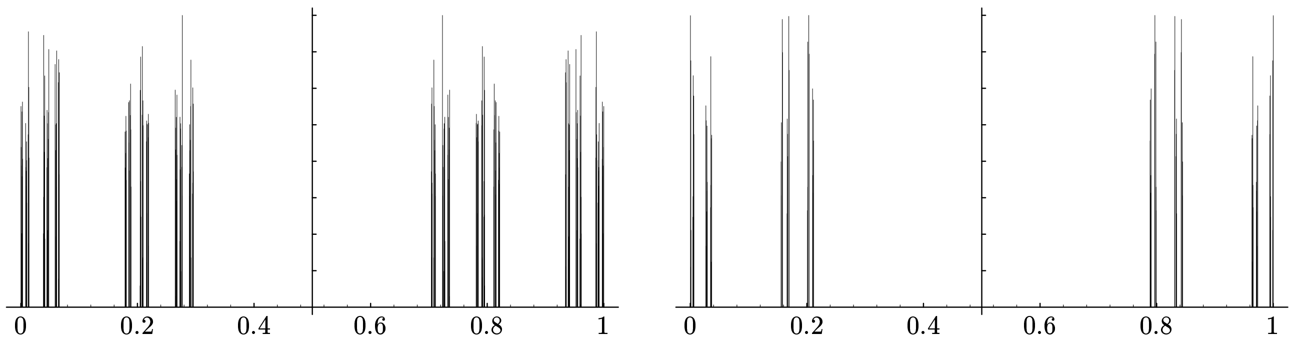

FIGURE 4. Histograms of Hausdorff measure on the Julia set for the mapping $z^2 - a$, where $a = 3.36$ (left) and $a = 6$ (right). (We have composed with an affine transformation to map the interval of interest, with endpoints $\pm(\sqrt{a + \frac{1}{4}} + \frac{1}{2})$, to the unit interval.) The Hausdorff dimensions are .584332 (left) and .451837 (right). The parameters for the computation are $d_{\max} = 4.1 \times 10^{-5}$, $d_{\text{ave}} = 4.7 \times 10^{-16}$ (left) and $d_{\max} = 4.2 \times 10^{-5}$, $d_{\text{ave}} = 9.4 \times 10^{-16}$ (right).

4. DENSITY DIAGRAMS

For μ a measure on the unit interval and α a given dimension, we define the full, left and right density functions by

$$\begin{aligned} d_\alpha(\mu, x, r) &= \mu(B_r(x))/(2r)^\alpha, \\ d_\alpha^L(\mu, x, r) &= \mu([x - r, x])/r^\alpha, \\ d_\alpha^R(\mu, x, r) &= \mu([x, x + r])/r^\alpha, \end{aligned}$$

where $B_r(x) = [x-r, x+r]$. Since we are interested in the behavior as $r \to 0$, and this will almost never be an ordinary limit, it is convenient to introduce a logarithmic scale for r, so we consider the function

$$h(x, s) = d_\alpha(\mu, x, c^{-s}) \tag{4.1}$$

for a convenient choice of the constant $c > 1$. In Figure 5 we show the graph of $h(x, s)$ as a function of s for a point x chosen at random in the Cantor set, for μ equal to the usual Cantor measure, and $\alpha = \log 2/\log 3$. Here it was convenient to choose $c = 3$. The graph was generated using Algorithm 2.2 with 17 iterations and then using the double-precision modification (2.4) once to compute the distribution function g of μ. It is evident that computation error becomes significant around $s = 19$, and by $s = 21$ the graph is totally misleading. However, in the region $0 \le s \le 15$, we have a reasonably accurate and rapid method for computing $h(x, s)$.

Figure 6 shows an overlap of all the graphs of $h(x, s)$ for a random selection of 500 points x in the Cantor set, for $0 \le s \le 2$. (The procedure we used was to choose one point at random between $2(k-1)10^{-3}$ and $2k10^{-3}$ for $1 \le k \le 500$, and then take the inverse image under the ramp function of these points.) The result is so striking it calls for a definition.

Definition 4.1. For μ supported on a compact set K, the *associated density diagram* is the subset of the first quadrant of the plane consisting of all pairs $(s, h(x, s))$ as x varies over K, and $s \ge 0$. For each

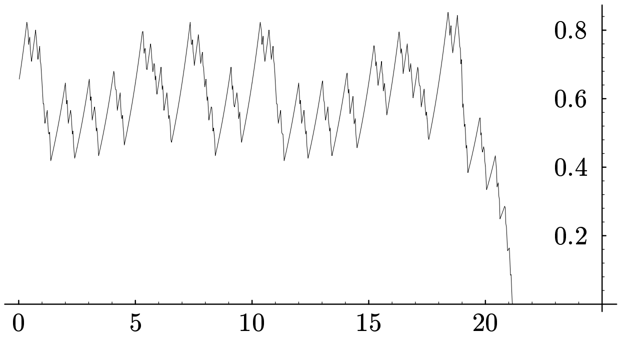

FIGURE 5. The graph of the function $h(x, s)$ of (4.1) versus s, for a randomly chosen x. The measure is the usual Cantor measure (that is, $S_1 x = \frac{1}{3}x$, $S_2 x = \frac{2}{3}x + \frac{1}{3}$, $p_1 = p_2 = \frac{1}{2}$), with $c = 3$.

fixed s, the density section measure ν_s is defined on the line by $\nu_s(A) = \mu\{x : h(x,s) \in A\}$ (the image of the measure μ under the mapping $x \to h(s,x)$), and the density diagram measure ν on the plane is $\nu = \nu_s\, ds$.

Our main goal in this section is to explain the periodicity and self-similarity apparent in Figure 6 and to explore what happens when we use more complicated measures μ. The periodicity is easy to explain: it arises from the fact that the contraction ratios are the same for the two transformations defining the Cantor set, and the nonzero separation of the two islands of the first generation.

Theorem 4.2. *Let $S_1, \dots, S_m$ be an i.f.s. of the special form $S_j x = \rho x + b_j$, and let μ be the self-similar measure satisfying*

$$\mu = \sum_{j=1}^{m} \frac{1}{m} \mu \circ S_j^{-1}. \tag{4.2}$$

Assume that the islands $S_j[0,1]$ of the first generation have a minimum separation $\varepsilon > 0$. Choose

$$\alpha = \frac{\log m}{\log \rho^{-1}}$$

and $c = \rho^{-1}$. Then the density diagram (with K equal to the attractor of the i.f.s.) is periodic of period 1 in the variable s, for $s \geq (\log \varepsilon / \log \rho) - 1$, and so are the measures ν_s and ν.

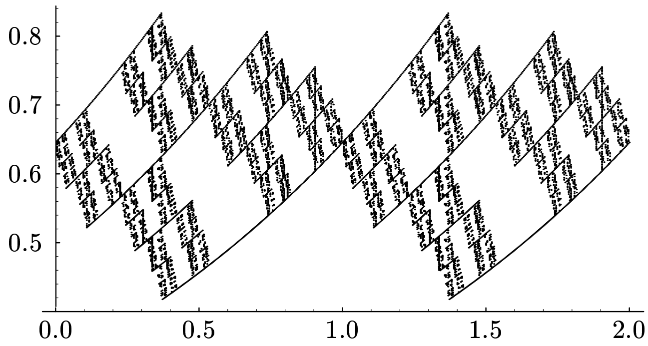

FIGURE 6. An overlay of 500 graphs of the type shown in Figure 5. This is the density diagram of the Cantor measure. The horizontal axis is restricted to $0 \leq s \leq 2$.

Proof. Each x in K must belong to one of the islands $S_j[0,1]$, say $x = S_j y$ with y in K. If $r \leq \varepsilon$ then $B_r(x)$ cannot intersect any of the other islands (except at an endpoint, which has μ measure zero), so $\mu(B_r(x)) = \frac{1}{m}\mu(B_{r/\rho}(y))$ by (4.2), since $B_{r/\rho}(y) = S_j^{-1}(B_r(x))$. This says $d_\alpha(\mu, x, r) = d_\alpha(\mu, y, r/\rho)$ because of our choice of α. Thus

$$h(x,s) = h(y, s-1) \tag{4.3}$$

if $r = \rho^s$. The condition $r \leq \varepsilon$ is the same as $s \geq \log\varepsilon/\log\rho$, in which case every point $(s,\, h(x,s))$ of the density diagram corresponds to a point $(s-1,\, h(y,\, s-1))$ one unit to the left. Conversely, every point $(s-1,\, h(y,\, s-1))$ arises in this fashion (just take $x = S_j y$ for any j), which proves the periodicity of the density diagram. The same argument shows that $h(x,s) \in A$ if and only if $h(y,\, s-1) \in A$, so $\nu_s = \nu_{s-1}$ is just the invariance of μ, $\mu(B) = \sum_{j=1}^{m} \mu(S_j B)$. $\square$

If the measure μ satisfies all the hypotheses except the minimum separation, so that some islands may touch (but are still assumed not to overlap), the density diagram will only have an asymptotic periodicity. We will not attempt to state this in a precise form.

For the Cantor measure, the periodicity begins at $s = 0$. To explain the apparent self-similarity in Figure 6 we restrict to $0 \leq s \leq 1$ and simplify, by what amounts to a horizontal and vertical rescaling, by returning to the parameter r and considering only the measure $\mu(B_r(x))$ rather than the density. Thus we are looking at the set

$$D = \{(r,\, \mu(B_r(x))) : \tfrac{1}{3} \leq r \leq 1,\, x \text{ in the Cantor set}\}.$$

This is shown in Figure 7. The mapping $(r,u) \to (\log_3 r^{-1}, u/(2r)^\alpha)$ is a homeomorphism of D onto the portion of the density diagram over $0 \leq s \leq 1$, so we can translate the properties of D back to properties of the density diagram.

The structure of D, which is contained in the rectangle $R = [\frac{1}{3}, 1] \times [\frac{1}{2}, 1]$, is given by four affine transformations T_1, T_2, T_3, T_4 of R to itself. We write (r,u) for the coordinates of D and $(r',u') = T_j(r,u)$. Then

$$T_1\begin{cases} r' = \frac{1}{3}r + \frac{2}{9}, \\ u' = \frac{1}{2}u + \frac{1}{4}, \end{cases} \qquad T_2\begin{cases} r' = \frac{1}{3}r + \frac{4}{9}, \\ u' = \frac{1}{2}u + \frac{1}{4}, \end{cases}$$
$$T_3\begin{cases} r' = \frac{1}{3}r + \frac{4}{9}, \\ u' = \frac{1}{2}u + \frac{1}{2}, \end{cases} \qquad T_4\begin{cases} r' = \frac{1}{3}r + \frac{2}{3}, \\ u' = \frac{1}{2}u + \frac{1}{2}. \end{cases}$$

We claim that $T_jD \subseteq D$. To see this, observe first that we can restrict $0 \le x \le \frac{1}{3}$ in computing D, because the contribution from the island $\frac{2}{3} \le x \le 1$ is the same. Then, since $x - r \le 0$, we can replace $\mu(B_r(x))$ by $\mu([0,\, x+r])$. Also, since $x + r \ge \frac{1}{3}$, we have

$$\mu([0,\, x+r]) = \tfrac{1}{2} + \mu([\tfrac{1}{3},\, x+r]),$$

so if $x + r \le \frac{2}{3}$ then $\mu([0,\, x+r]) = \frac{1}{2}$ while if $x + r \ge \frac{2}{3}$ then $\mu([0,\, x+r]) = \frac{1}{2} + \mu([0,\, x+r-\frac{2}{3}])$ (note that $x + r \le \frac{4}{3}$).

To each transformation T_j we associate a mapping $x \to x'$ of $[0, \frac{1}{3}]$ to itself, namely $x' = \frac{1}{3}x$ for T_2 and T_4 and $x' = \frac{1}{3}x + \frac{2}{9}$ for T_1 and T_3. Now for T_1 and T_2 we compute $x' + r' = \frac{1}{3}(x+r) + \frac{4}{9}$, which can be written $x' + r' - \frac{2}{3} = \frac{1}{3}(x + r - \frac{2}{3})$, so that $\mu([0,\, x'+r'-\frac{2}{3}]) = \frac{1}{2}\mu([0,\, x+r-\frac{2}{3}])$ if $x + r \ge \frac{2}{3}$. Thus, if (r,u) is a point of D arising from x in $0 \le x \le \frac{1}{3}$ with $x + r \ge \frac{2}{3}$, the point x' gives rise to the point (r',u') in D with $u' = \mu([0,\, x'+r']) = \frac{1}{2} + \mu([0,\, x'+r'-\frac{2}{3}]) = \frac{1}{4} + \frac{1}{2}(\frac{1}{2} + \mu([0,\, x+r-\frac{2}{3}])) = \frac{1}{2}u + \frac{1}{4}$. On the other hand, if $x + r \le \frac{2}{3}$ then $u = \frac{1}{2}$ and $x' + r' \le \frac{2}{3}$ also, so $u' = \frac{1}{2}$ and $u' = \frac{1}{2}u + \frac{1}{4}$ trivially. This shows

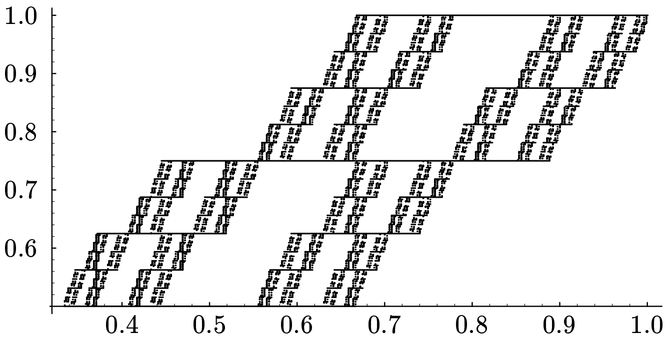

FIGURE 7. The transform D of the density diagram of Figure 6.

$T_jD \subseteq D$ for $j = 1, 2$. For $j = 3, 4$, the argument is similar, except now $x' + r' - \frac{2}{3} = \frac{1}{3}(x+r)$ and so $u' = \frac{1}{2} + \mu([0,\, x'+r'-\frac{2}{3}]) = \frac{1}{2} + \frac{1}{2}\mu([0,\, x+r]) = \frac{1}{2}u + \frac{1}{2}$, and we always have $x' + r' \ge \frac{2}{3}$.

Thus $\bigcup_{j=1}^{4} T_jD \subseteq D$, but this is not the whole story. There are two line segments $L_1 = [\frac{4}{9}, \frac{5}{9}] \times \frac{1}{2}$ and $L_2 = [\frac{7}{9}, \frac{8}{9}] \times 1$ that are contained in D but are not in $\bigcup_{j=1}^{4} T_jD$. In fact,

$$D = \bigcup_{j=1}^{4} T_jD \cup L_1 \cup L_2,$$

and this decomposition is essentially disjoint. Thus D is a relatively self-affine set, relative to L_1 and L_2 (relatively self-similar measures were discussed in [Strichartz 1993b]).

If we generate a self-similar measure from the same i.f.s. using different probabilities, the density diagram becomes more complicated. For a linear i.f.s. with contraction ratios r_j we take the dimension

$$\alpha = \frac{\sum p_j \log p_j}{\sum p_j \log r_j}. \tag{4.4}$$

Figure 8 (top) shows the result for $p_1 = 0.4$ and $p_2 = 0.6$.

Nevertheless, we can restore self-similarity for small values of s by the simple device of separating the contributions to the density diagram from each of the first generation islands. This is illustrated in the second and third panels of Figure 8, where in place of D we graph D_1 and D_2 defined in the same way except x is restricted to the portion of the Cantor set in $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$, respectively. (In generating these graphs we chose the points x randomly according to the measure μ.) Now $D = D_1 \cup D_2$, and each D_j is relatively self-affine. For simplicity we deal with D_1 only. The four affine transformations are now defined as follows:

$$T_1\begin{cases} r' = \frac{1}{3}r + \frac{2}{9}, \\ u' = .4u + .24, \end{cases} \qquad T_2\begin{cases} r' = \frac{1}{3}r + \frac{4}{9}, \\ u' = .4u + .24, \end{cases}$$
$$T_3\begin{cases} r' = \frac{1}{3}r + \frac{4}{9}, \\ u' = .6u + .4, \end{cases} \qquad T_4\begin{cases} r' = \frac{1}{3}r + \frac{2}{3}, \\ u' = .6u + .4. \end{cases}$$

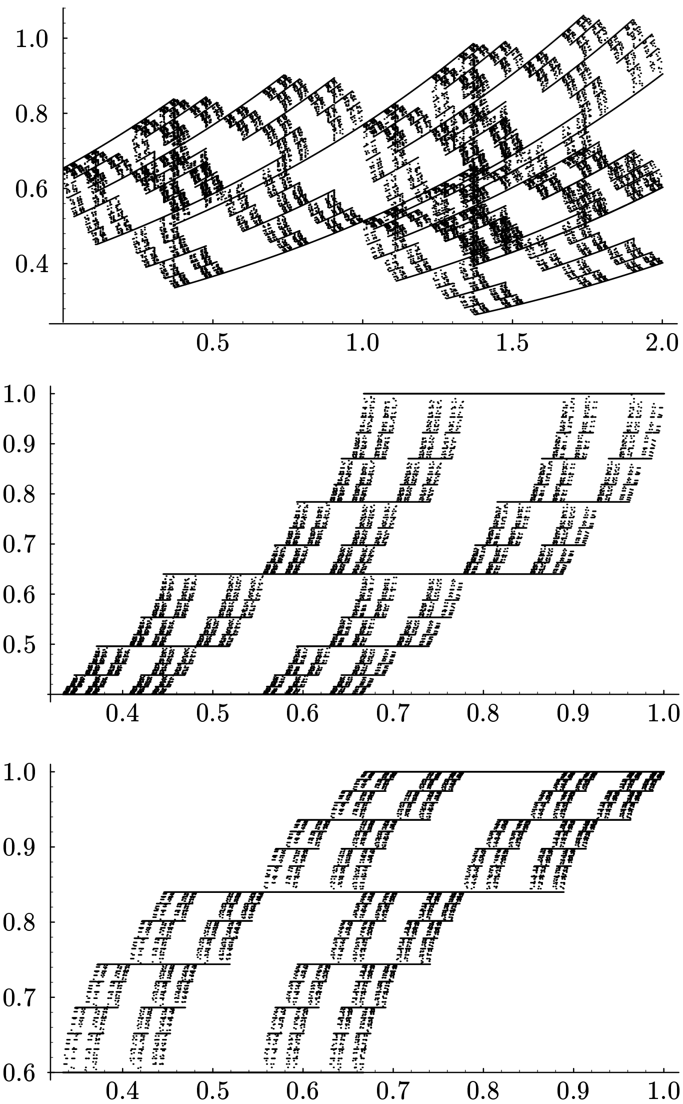

FIGURE 8. Top: The density diagram for the measure associated with the Cantor i.f.s. with weights $p_1 = .4$, $p_2 = .6$. Middle: The transform D_1 of the portion of the top diagram generated by points in the first island, $[0, \frac{1}{3}]$. Bottom: The transform D_2 of the portion of the top diagram generated by points in the second island, $[\frac{2}{3}, 1]$.

Notice that r' is defined as before, and we take x' as before. In this case we have $\mu(0,a) = \frac{2}{3}\mu(\frac{2}{3}, a+\frac{2}{3})$ for $0 < a \le \frac{1}{3}$, so that $\mu([0, x+r])$ equals

$$\begin{cases} .4 & \text{if } x+r \le \frac{2}{3}, \\ .4 + 1.5\mu([0,\, x+r-\frac{2}{3}]) & \text{if } x+r > \frac{2}{3}. \end{cases}$$

To show that $T_j D_1 \subseteq D_1$ for $j = 1, 2$, we have $x' + r' - \frac{2}{3} = \frac{1}{3}(x + r - \frac{2}{3})$ as before, so

$$\begin{aligned} \mu([0,\, x'+r']) &= .4 + 1.5\mu([0,\, x'+r'-\tfrac{2}{3}]) \\ &= .4 + .6\mu([0,\, x+r-\tfrac{2}{3}]) \\ &= .24 + .4\mu([0,\, x+r]), \end{aligned}$$

which gives $u' = .4u + .24$ as claimed. For $j = 3, 4$ we have $x' + r' - \frac{2}{3} = \frac{1}{3}(x+r)$, so

$$\begin{aligned} \mu([0,\, x'+r']) &= .4 + 1.5\mu([0,\, \tfrac{1}{3}(x+r)]) \\ &= .4 + .6\mu([0,\, x+r]), \end{aligned}$$

which gives $u' = .6u + .4$, as claimed. Finally, the two additional line segments are now $L_1 = [\frac{4}{9}, \frac{5}{9}] \times 0.4$ and $L_2 = [\frac{7}{9}, \frac{8}{9}] \times 1$.

Returning to the top diagram in Figure 8, we have now given an explanation of the left half (that is, $0 \le s \le 1$) as the superposition of two transforms of relatively self-affine sets. But the right half ($1 \le s \le 2$) is just a superposition of two affine images of the left half. In the notation of Theorem 4.2, we have

$$h(x, s) = \begin{cases} .4 \cdot 3^{\alpha} h(y,\, s-1) & \text{if } x = S_1 y, \\ .6 \cdot 3^{\alpha} h(y,\, s-1) & \text{if } x = S_2 y, \end{cases}$$

and the numbers $0.4 \cdot 3^{\alpha}$ and $0.6 \cdot 3^{\alpha}$ are distinct distortion factors in the vertical direction (in fact one is a contraction and one is an enlargement). Thus the image is "smudged" by this superposition, and as s increases the smudging gets progressively worse, until all detail is lost. Also, the density diagram is not bounded (or bounded away from zero) as $s \to \infty$. In Section 6 we will describe some averaging processes that will restore boundedness.

We consider next a more complicated example, the i.f.s. $S_1 x = \frac{1}{4}x$, $S_2 = \frac{1}{3}x + \frac{2}{3}$ with $p_1 = p_2 = \frac{1}{2}$ (this is the measure represented in Figure 1). The two affine maps have different contraction ratios, and this affects the density diagram, shown in Figure 9, and the graphs of D_1 and D_2 defined as above (using the islands $[0, \frac{1}{4}]$ and $[\frac{2}{3}, 1]$). At first glance it appears that D_1 and D_2 are similar to the analogous sets for the previous example, but

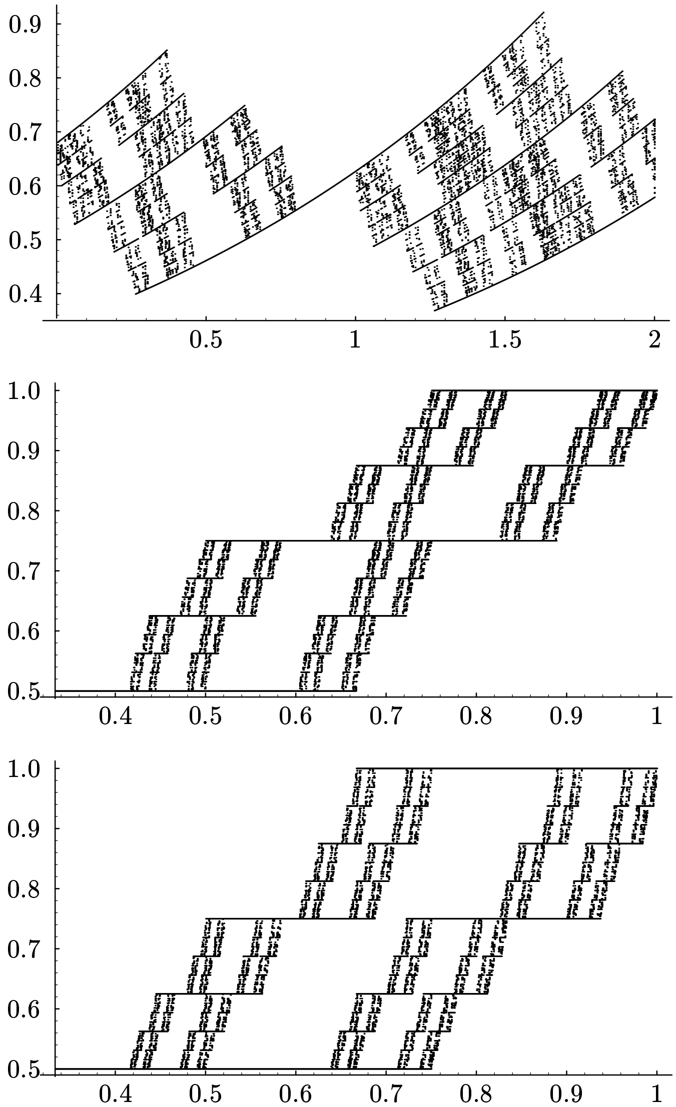

FIGURE 9. Top: The density diagram for the measure in Figure 1 on the scale $r = 3^{-s}$. Middle and bottom: D_1 and D_2 for the same density diagram.

this is not completely correct. Now there are only two transformations, the analogs of the previous T_2 and T_3, that map D_1 to a subset of D_1. They are

$$T_2 \begin{cases} r' = \frac{1}{4}r + \frac{1}{2} \\ u' = \frac{1}{2}u + \frac{1}{4} \end{cases} \quad \text{with} \quad x' = \tfrac{1}{4}x,$$

$$T_3 \begin{cases} r' = \frac{1}{3}r + \frac{1}{2} \\ u' = \frac{1}{2}u + \frac{1}{2} \end{cases} \quad \text{with} \quad x' = \tfrac{1}{3}x + \tfrac{1}{6}.$$

The images $T_2 D_1$ and $T_3 D_1$ are the lower and upper halves of the central part of D_1, and correspond to values of $\mu(B_r(x))$ as follows: T_2 : x lies in the island $S_1S_1[0,1]$ and $B_r(x)$ extends to the island $S_2S_1[0,1]$, and T_3 : x lies in the island $S_1S_2[0,1]$ and $B_r(x)$ extends to the island $S_2S_2[0,1]$.

The upper right and lower left portions of D_1 are visibly *not* affine images of the whole (note that the placement of the central "tower" is different). However, for each of these pieces we can find two affine mappings analogous to T_2 and T_3 above that map it into itself (the images again being the lower and upper halves of the central portions). For example, the upper right portion of D_1 is generated by the values of $\mu(B_r(x))$ where x lies in the island $S_1S_1[0,1]$ and $B_r(x)$ extends to the island $S_2S_2[0,1]$. Two affine images of this portion can be generated to cover the smaller portion generated by (a) x in $S_1S_1S_1[0,1]$ and $B_r(x)$ extending to $S_2S_2S_1[0,1]$, and (b) x in $S_1S_1S_2[0,1]$ and $B_r(x)$ extending to $S_2S_2S_2[0,1]$.

To give a complete description of D_1 as a relative self-affine set would require an infinite number of affine transformations, and a vector version of self-similarity as in [Mauldin and Williams 1988].

There is a further complication in this example when we try to explain the behavior of the density diagram as s increases. Note that we have chosen a scale of $r = 3^{-s}$, while the two contraction ratios are $\frac{1}{4}$ and $\frac{1}{3}$, so the base 3 is not in synch with the first ratio. This results in a horizontal "smudging" effect in addition to the vertical one noted in the previous example (the vertical smudging is present here as well because we did not choose the weights to produce the Hausdorff measure on the attractor). In the notation of Theorem 4.2, $h(x,s)$ equals

$$\begin{cases} \frac{1}{2}4^{\alpha}h\Big(y,\, s - \dfrac{\log 4}{\log 3}\Big) & \text{if } x = S_1 y, \\ \frac{1}{2}3^{\alpha}h(y,\, s-1) & \text{if } x = S_2 y, \end{cases}$$

so that we are superimposing dilated versions of the density diagram translated by periods 1 and $\log 4/\log 3$.

These are by no means the most complicated density diagrams that can arise in the linear i.f.s. case. For example, the lakes were chosen to be relatively large, so that only one endpoint of each interval $B_r(x)$ can lie in an island. Examples with smaller lakes or more than two islands present further difficulties that we are unable to analyze explicitly. Still, the density diagrams for small s, even for nonlinear i.f.s.'s and variable weights, present a qualitative picture that is analogous to what we have seen in the previous examples. In Figure 10 we show an example for the Hausdorff measure on the Julia set of the mapping $z^2 - 3.36$ (compare Figure 4, left), where $\alpha = .584332$ is the Hausdorff dimension.

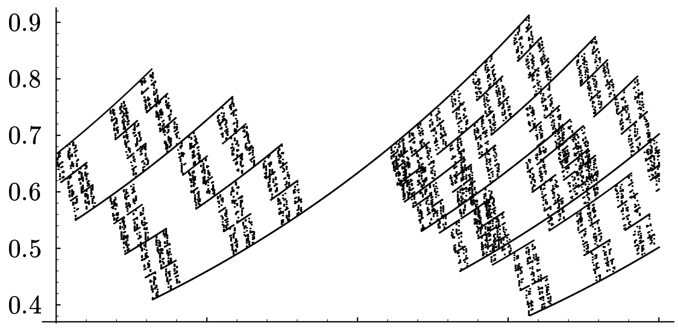

FIGURE 10. The density diagram for Hausdorff measure on the Julia set of the mapping $z^2 - 3.36$, on the scale $r = 3^{-s}$.

We conclude this section with a discussion of another approach to organizing the information in the density functions $h(x, s)$: rearrangements. Recall that if $f(x)$ is a nonnegative measurable function on a probability measure space, the *nonincreasing rearrangement* $f^*(t)$ defined on $[0, 1]$ is the unique nondecreasing right-continuous function equimeasurable with f (this means the measure of the set where f satisfies $f(x) > s$ is the same as for f^*) [Stein and Weiss 1971]. From an algorithmic point of view it is easy to compute f^*: just choose random sample points and sort the values of $f(x)$ in decreasing order. The idea is that f^* contains all the "size information" about f, and presents it in a consistent fashion on the standard measure space $[0, 1]$.

In our case we want to take the function $h(x, s)$ for x in K and $0 \le s \le N$ with the probability measure $d\mu \times ds/N$ and write $h_N^*(t)$ for the rearrangement, and let $h^*(t)$ be the limit as $N \to \infty$. Using the methods of [Graf 1993] it is easy to show that this limit exists for Hausdorff measure on the attractor of a linear i.f.s., and (using an ergodic theorem) that, for μ-amost every x, we get the same function h^* by rearranging $h(x, s)$ on $0 \le s \le N$ with respect to the probability measure ds/N, keeping x fixed. For the Cantor measure, or more generally the measures described in Theorem 4.1, the periodicity condition (4.3) means that we can obtain h^* simply by rearranging $h(x, s)$ for x in K and $s_0 \le s \le s_0 + 1$ for sufficiently large s_0. For other linear i.f.s.'s, it follows from [Graf 1993] that h^* is equal to the rearrangement of $h(x, s)$ over the set $\{(x, s) : s_1(x) \le s \le s_2(x)\}$ where $c^{-s_j(x)} = r_j(x)$ is defined to be the largest value of r such that $B_r(x)$ lies entirely in the j-th generation island containing x. (The measure $d\mu \times ds$ on this set has to be normalized to be a probability measure; also $r_j(x)$ may be zero on a set of measure zero). We have not implemented this construction because of its complexity.

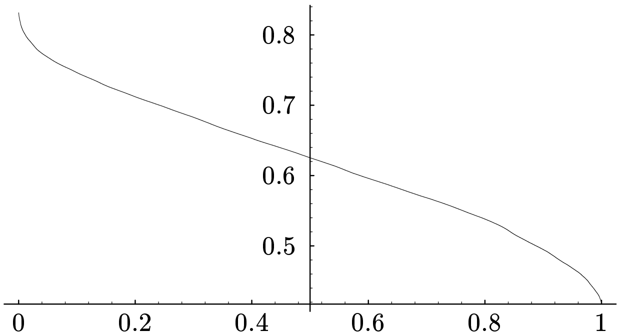

FIGURE 11. The rearranged density function h^* for the Cantor measure, computed by choosing 1000 random points x and sampling $h(x, s)$ at 1000 regularly spaced values in $0 \le s \le 1$.

In Figure 11 we show h^* for the Cantor measure, obtained by computing h_1^*. In Figure 12 we show

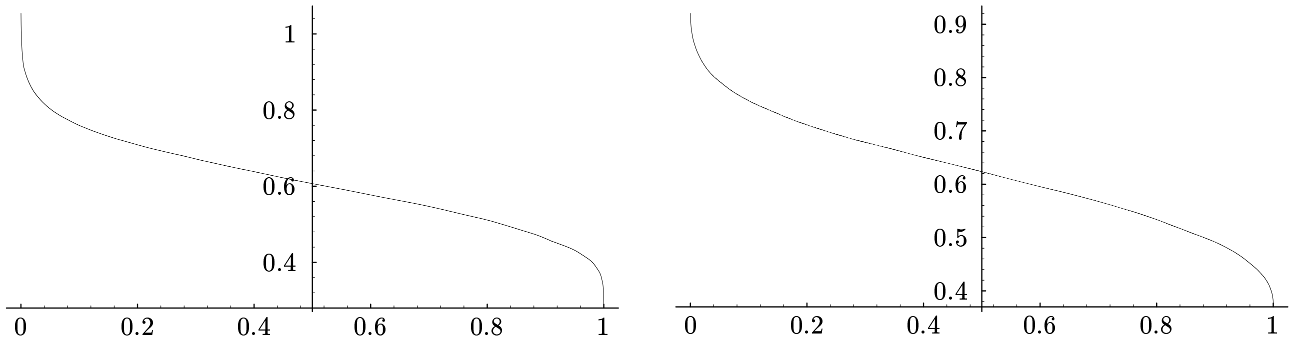

FIGURE 12. The approximation h_5^* to h^* for two measures. Left: the measure in Figure 1, using α given by (4.4) and $c = 3$. Right: the Hausdorff measure on the Julia set of the mapping $z^2 - 3.36$, with $\alpha = .584332$ and $c = 3$. In both cases the figure was computed by choosing 1000 random points x and sampling s at 1000 regularly spaced values in $0 \le s \le 5$.

h_5^* as an approximation to h^* for two measures. On the left we use the measure of Figure 1. Since this is not the Hausdorff measure for the attractor, the previous discussion does not apply, and in fact we know that the limit h^*, if it exists, must be unbounded. On the right we do the same for the Hausdorff measure on the Julia set of the mapping $z^2 - 3.36$. In this case we expect h^* to be bounded, but the existence of the limit has not yet been established.

Problem 4.3. Under what conditions does the limit h^* exist? When can we assert that, for μ almost every x, the rearrangments of $h(x,s)$ on $0 \le s \le N$ converge to h^*?

We would also like to raise the inverse question:

Problem 4.4. Can you recover the measure μ from the rearranged density function h^*? If not, how much information concerning μ is contained in the function h^*?

5. COMPUTATION OF HAUSDORFF MEASURE

If E is any set of finite nonzero Hausdorff measure $\mathcal{H}_\alpha$ in dimension α, the upper density

$$\limsup_{|I|\to 0} \frac{\mathcal{H}_\alpha(E \cap I)}{|I|^\alpha}$$

for x contained in I, is one for almost every x in E (with respect to $\mathcal{H}_\alpha$) [Falconer 1985]. We want to turn this observation around in order to compute $\mathcal{H}_\alpha$ restricted to E exactly, in situations where we know how to compute $\mathcal{H}_\alpha$ up to a constant multiple.

Consider first the case of a linear i.f.s. with no overlaps. If $S_j x = \rho_j x + b_j$ are the mappings, the choice of natural weights $p_j = \rho_j^\alpha$ (with α chosen so that $\sum_{j=1}^m p_j = 1$) leads to a self-similar probability measure μ that must be a multiple of $\mathcal{H}_\alpha$ restricted to the attractor K:

$$c\mu = \mathcal{H}_\alpha\big|_K. \tag{5.1}$$

Because of the self-similarity, it is easy to see that the upper density can be given by a simple supremum, namely

$$\sup_I \frac{\mathcal{H}_\alpha(I \cap K)}{|I|^\alpha} = 1,$$

where the supremum is taken over all intervals I in $[0,1]$. In fact it is also easy to see that the intervals must have both endpoints in K, for if not we could contract the length of the interval without changing the measure. Thus we find the constant c in (5.1) via

$$c^{-1} = \sup_I \frac{\mu(I)}{|I|^\alpha}. \tag{5.2}$$

It is not difficult to design efficient algorithms to search through enough plausible choices for I to compute this supremum, and to produce a conjectural interval that attains the supremum. The point is that if $\tilde{I}$ is such an interval, then so is $S_j\tilde{I}$ for any mapping in the i.f.s., or, more generally, $S_J\tilde{I}$, $J = (j_1, \dots, j_n)$, for any composition of mappings. (Here $S_J = S_{j_1}S_{j_2}\cdots S_{j_n}$.) We can also expand the interval $\tilde{I}$ by applying S_j^{-1} if $\tilde{I}$ is contained in the first generation island $S_j[0,1]$, and still maintain the supremum in (5.2). In displaying the interval $\tilde{I}$, we always repeat this expansion until $\tilde{I}$ is not contained in any first generation islands. This gives us a "largest" example.

Lemma 5.1. *Suppose that the islands of the first generation are all separated. Then there exists an interval $\tilde{I}$ that achieves the supremum in* (5.2).

Proof. Let $\{I_j\}$ be a sequence of intervals such that

$$\frac{\mu(I_j)}{|I_j|^\alpha} \tag{5.3}$$

approaches the supremum as $j \to \infty$. By repeated use of the expansion argument we may replace I_j with another interval that is not contained in any island of the first generation, without changing the ratio (5.3). Having done this, we have a positive lower bound for the length of all I_j by the smallest length of the first generation lakes. Then by a compactness argument we can pass to a subsequence that converges to a nondegenerate interval $\tilde{I}$. Since we know the measure has no atoms we obtain

$$\frac{\mu(\tilde{I})}{|\tilde{I}|^\alpha} = \lim_{j\to\infty} \frac{\mu(I_j)}{|I_j|^\alpha},$$

so the supremum is attained. □

Lemma 5.2. *Let I be an interval. Then given any $\varepsilon > 0$ there exists a set E_ε with $\mu(E_\varepsilon) \le \varepsilon$ and a finite disjoint covering of $K \setminus E_\varepsilon$ by images of I under iterated similarities S_J.*

Proof. Let $\mathcal{F}_N$ denote the field of sets generated by the islands and lakes of the N-th generation. Suppose first that I belongs to $\mathcal{F}_N$ for some N. We will construct inductively a disjoint sequence $S_{J_1}I$, $S_{J_2}I$, ... whose union fills up K, except for a set of measure zero. The sequence begins with $S_1I, S_2I, \dots, S_mI$, which are disjoint and satisfy

$$\mu(S_1I \cup \cdots \cup S_mI) = \lambda\mu(K)$$

for some positive λ, so

$$\mu\Bigl(K \setminus \bigcup_{j=1}^{m} S_{J_j}I\Bigr) = (1-\lambda)\mu(K).$$

Now $\bigcup_{j=1}^m S_{J_j}I$ belongs to $\mathcal{F}_{N+1}$, so its complement can be written $\bigcup_{i=1}^M S_{J_i'}K$ for some collection of multi-indices J_i' of length $N+1$. We continue our sequence with $S_jS_{J_i'}I$ letting j vary over $[1,m]$ and i vary over $[1,M]$. The sequence remains disjoint, but now the complement has measure $(1-\lambda)^2\mu(K)$ and belongs to $\mathcal{F}_{2N+2}$. We can continue this process indefinitely, each time obtaining a disjoint sequence of images of I whose complement has measure $(1-\lambda)^n\mu(K)$.

Finally, if I does not belong to any $\mathcal{F}_N$, we can enlarge I slightly to $I' \supseteq I$, so that I' belongs to some $\mathcal{F}_N$ and $\mu(I' \setminus I) \le \varepsilon\mu(I)$. Then we apply the above construction to I' and shrink back to I. The shrinking preserves disjointness, and we can make the measure of the complement as small as we want by taking ε small enough. □

If we can find an interval $\tilde{I}$ that achieves the supremum in (5.2), Lemma 5.2 effectively tells us how to cover K so as to calculate the Hausdorff measure exactly.

We examined in detail the case $r_1 = r_2 = r_3 = .2$ with first and third islands anchored at the ends of the unit interval, and the middle island floating in between, so $S_1x = .2x$, $S_2x = .2x + b_2$ and $S_3x = .2x + .8$, with $.2 \le b_2 \le .4$ (the range $.4 \le b_2 \le .6$ yields equivalent attractors under reflection). For $b_2 = .4$ the islands are symmetrically arranged and it is known that the supremum in (5.2) is one, attained by $I = [0,1]$. See [Falconer 1985] for a proof, which will also work if b_2 is sufficiently close to .4. We experimentally estimated the supremum for values of b_2 incrementing by .02

from .2 to .4. We used an ad hoc procedure to sift through a reasonable selection of intervals from $\mathcal{F}_{10}$. After finding an interval that achieved the maximum, we expanded it until we obtained an interval not contained in any first generation islands. Altogether we found only four different island configurations, the most complicated belonging to $\mathcal{F}_3$. These configurations are illustrated in Figure 13 (left) for typical values of b_2. To confirm these estimates and to locate the transition points from one configuration to the next, we present in Figure 13 (right) the graph of the exact value of $\mu(I)/|I|^\alpha$ for each of the four configurations as a function of b_2. In addition we have graphed two other configurations, one from $\mathcal{F}_4$ and one that does not belong to any $\mathcal{F}_N$, that seemed like plausible candidates for achieving the supremum; both fail decisively to be in contention.

In the summer of 1994 (after the first version of this paper was completed), the first author and Elizabeth Ayer applied the experimental method just described in a more systematic fashion. The results obtained [Ayer and Strichartz 1995] show that it is not always true that the supremum is achieved by an interval in $\mathcal{F}_N$ for some N, and that the exact Hausdorff measure is a discontinuous function of the parameters of the linear i.f.s. Also, after this work was completed, we became aware of earlier work [Marion 1986; 1987] that deals with the same problem using a similar approach, and contains results equivalent to Lemmas 5.1 and 5.2.

Next we consider the attractor of a nonoverlapping i.f.s. with nonlinear mappings. In Section 3 we described an algorithm for computing an approximation to a multiple of Hausdorff measure, normalized to be a probability measure. We can apply essentially the same procedure as in the linear case to compute approximately the correct normalization. In this case we cannot replace the lim sup

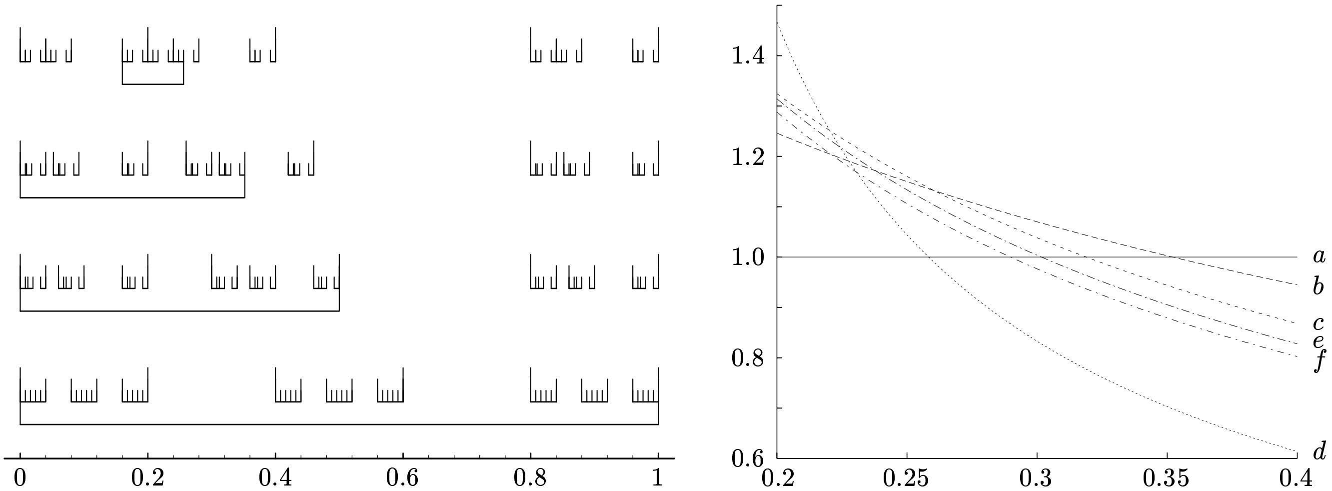

FIGURE 13. Left: Configurations of intervals of maximum density for four typical "floating island" examples. Vertical ticks above each baseline mark the endpoints of islands in $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ (largest lines for $\mathcal{F}_1$, shortest for $\mathcal{F}_3$). A bracket below the baseline indicates the interval of maximum density. Example 1 (bottom) has $b_2 = .4$, so the floating island is in the middle and the entire interval is the interval of maximum density. Example 4 (top) has $b_2 = .2$, so there is no lake between the first two islands in $\mathcal{F}_1$. Right: Plots of $\mu(I)/|I|^\alpha$ as a function of b_2 for six configurations, the first four being the ones shown on the left. Curve a corresponds to $[0,1]$, curve b to $S_1[0,1]$ through $S_2[0,1]$, curve c to $S_1[0,1]$ through $S_1 \circ S_2[0,1]$, curve d to $S_1 \circ S_3[0,1]$ through $S_2 \circ S_2 \circ S_2[0,1]$, curve e to $S_0[0,1]$ through $S_2 \circ S_2 \circ S_2 \circ S_2[0,1]$, and curve f to $S_0[0,1]$ through $\lim S_2 \circ \cdots \circ S_2[0,1]$. Note that e and f are always below c, but sometimes above a, b and d.

with a sup. However, it is plausible to approximate the lim sup with the supremum in (5.2) where the size of the intervals $|I|$ is restricted to be small, but is allowed to vary considerably. A graph of the function

$$\varphi(r) = \sup\left\{\frac{\mu(I)}{|I|^\alpha} : |I| = r\right\}$$

for some examples (Figure 14) shows a similar pattern for both linear and nonlinear examples: local maximum peaks occur regularly, and the local maxima appear to converge rapidly to what is presumably the lim sup.

The procedure we used was based on interval partition data $(\mathcal{J}, \nu)$ generated without any consolidation, so the intervals of $\mathcal{J}$ are exactly the islands of the last generation. We chose 1000 of these islands at random and maximized $\mu(I)/|I|^\alpha$ over all intervals starting at the left at one of the chosen islands and extending to the right between 100 and 900 islands. The choice of a minimum length of 100 islands for the intervals was made to overcome the inevitable lack of accuracy in the computation of the measure on the smallest scale. By allowing a range of 100 to 900 islands we are sure to include several local maximum peaks of $\varphi(r)$.

We present the results of this procedure on the Julia sets for $z^2 - a$ in the range $2 \le a \le b$ in Table 1. Note that $2b$ is the diameter of the Julia set. Since we performed the procedure on the scaled version (affinely transformed to have diameter one), the reciprocal of the maximum density gives the relative Hausdorff measure $\mu(K)/|K|^\alpha$. It is intriguing that this function appears to show a discontinuity at $a = 2$. This possibility deserves further investigation, but this will require a more accurate procedure. Note that the region around $a = 2$ involves large derivatives of the functions

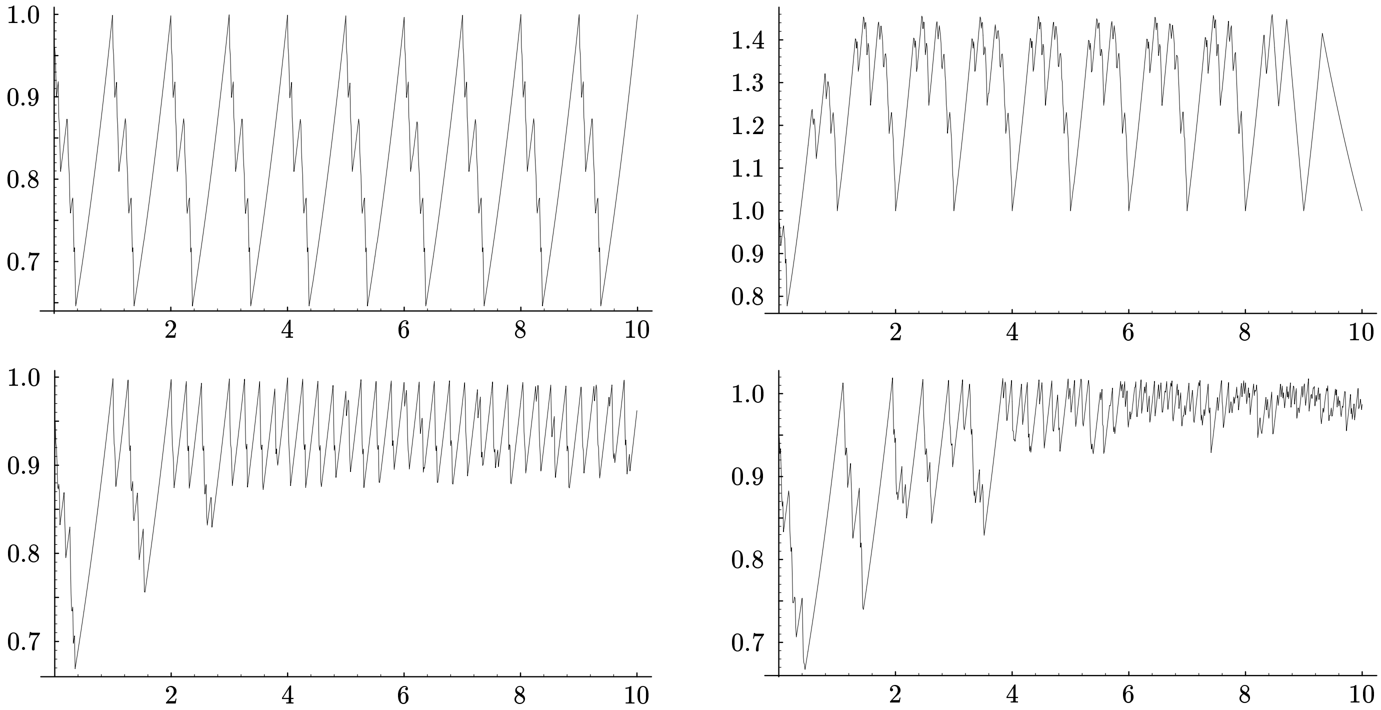

FIGURE 14. Graphs of $\sup\{\mu(I)/|I|^\alpha : |I| = r\}$ on a logarithmic scale. Top left: Cantor measure. The periodicity and the value 1 for the maximum are expected. Top right: "floating island" example with $b_2 = .2$. The asymptotic periodicity is expected. The graph beyond $s = 8$ is degraded by computational error. Bottom left: measure in Figure 1. Because the contraction ratios $\frac{1}{4}$ and $\frac{1}{3}$ are distinct, no periodicity is expected. Bottom right: Hausdorff measure on the Julia set in Figure 4.

$\pm\sqrt{x+a}$ (the derivative becomes unbounded at $a = 2$), so it may be very difficult to obtain the desired accuracy. The computations we did for $a = 2$ are somewhat reassuring. The error in the dimension was 2×10^{-6}, and the error in the relative Hausdorff measure was 10^{-4}, which is small compared to the apparent size of the jump discontinuity (3×10^{-2}) of the relative Hausdorff measure.

a	b	α	max dens	$\mu(K)/\|K\|^\alpha$	$\mu(K)$
2.00	2.0	0.999998	1.000104	0.999896	3.999570
2.31	2.1	0.766235	1.037192	0.964142	2.895318
2.64	2.2	0.681853	1.030990	0.969942	2.663703
2.99	2.3	0.625882	1.025456	0.975176	2.534492
3.36	2.4	0.584332	1.021211	0.979230	2.448814
3.75	2.5	0.551618	1.018020	0.982299	2.386757
4.16	2.6	0.524880	1.015831	0.984415	2.338806
4.59	2.7	0.502441	1.013995	0.986198	2.301171
5.04	2.8	0.483233	1.012437	0.987716	2.270812
5.51	2.9	0.466536	1.011124	0.988998	2.245754
6.00	3.0	0.451837	1.009989	0.990110	2.224749

TABLE 1. Computed values of Hausdorff dimension α, maximum density, relative Hausdorff measure, and Hausdorff measure, for the Julia set of $z^2 - a$ on the interval $[-b, b]$. The relative Hausdorff measure is the reciprocal of the maximum density, and the Hausdorff measure is the relative Hausdorff measure multiplied by $(2b)^\alpha$.

6. AVERAGE DIMENSIONS AND DENSITIES

The *entropy dimension* of a measure μ, if it exists, is the limit

$$\lim_{r\to 0} \frac{\int \log \mu(B_r(x))\, d\mu(x)}{\log r}.$$

It is shown in [Young 1992] that if the measure satisfies

$$\lim_{r\to 0} \frac{\log \mu(B_r(x))}{\log r} = \alpha \quad \text{for } \mu\text{-a.e. } x, \tag{6.1}$$

the entropy dimension exists and equals α. The associated density question is the behavior of

$$\varphi(r) = \int \log \frac{\mu(B_r(x))}{(2r)^\alpha}\, d\mu(x) \tag{6.2}$$

as $r \to 0$. If the limit of $\varphi(r)$ exists, we call its exponential the *entropy density*, but usually we need to consider *upper* and *lower entropy densities* defined by the lim sup and lim inf. Because (6.2) involves averaging, we can expect to see better behavior than for the pointwise ratio $\mu(B_r(x))/(2r)^\alpha$.

For the measures we are considering it is usually easy to compute α and study the behavior of (6.2) experimentally. (To approximate the integral by a Cauchy sum we partition the interval $[0,1]$ by taking the inverse image under the ramp function of a uniform partition, and evaluate the integrand at a random point in each subinterval.) For a nonoverlapping linear i.f.s. with constant weights p_j, the value of α in (6.1) is just

$$\alpha = \frac{\sum p_j \log p_j}{\sum p_j \log r_j}. \tag{6.3}$$

For the general (nonoverlapping) case the presumed value of α is

$$\alpha = \frac{\sum \int p_j \log p_j \, d\mu}{\sum \int p_j \log |S_j'|\, d\mu}. \tag{6.4}$$

For a proof in the case of constant weights (and contractivity) see [Strichartz 1993b]. For μ equal to a multiple of Hausdorff measure on K, the entropy dimension equals the Hausdorff dimension in essentially all the cases we are considering.

In Figure 15 we show the graphs of $\varphi(r)$ as a function of $\log r$ for some typical cases. The first three examples can be easily explained.

Theorem 6.1. *Assume the i.f.s. is linear and separated (there is a positive distance between islands $S_j[0,1]$) and the weights p_j are constant. Then for φ given by* (6.2) *and α given by* (6.3), *we have*

$$\lim_{r\to 0}(\varphi(r) - \tilde\varphi(r)) = 0,$$

where

(a) *$\tilde\varphi(r)$ is a continuous, positive multiplicatively periodic function (that is, $\tilde\varphi(\lambda r) = \tilde\varphi(r)$) if the contraction ratios satisfy $r_j = \lambda^{k_j}$ for positive integers k_j and some λ, and*

(b) *$\tilde\varphi(r)$ is a positive constant otherwise.*

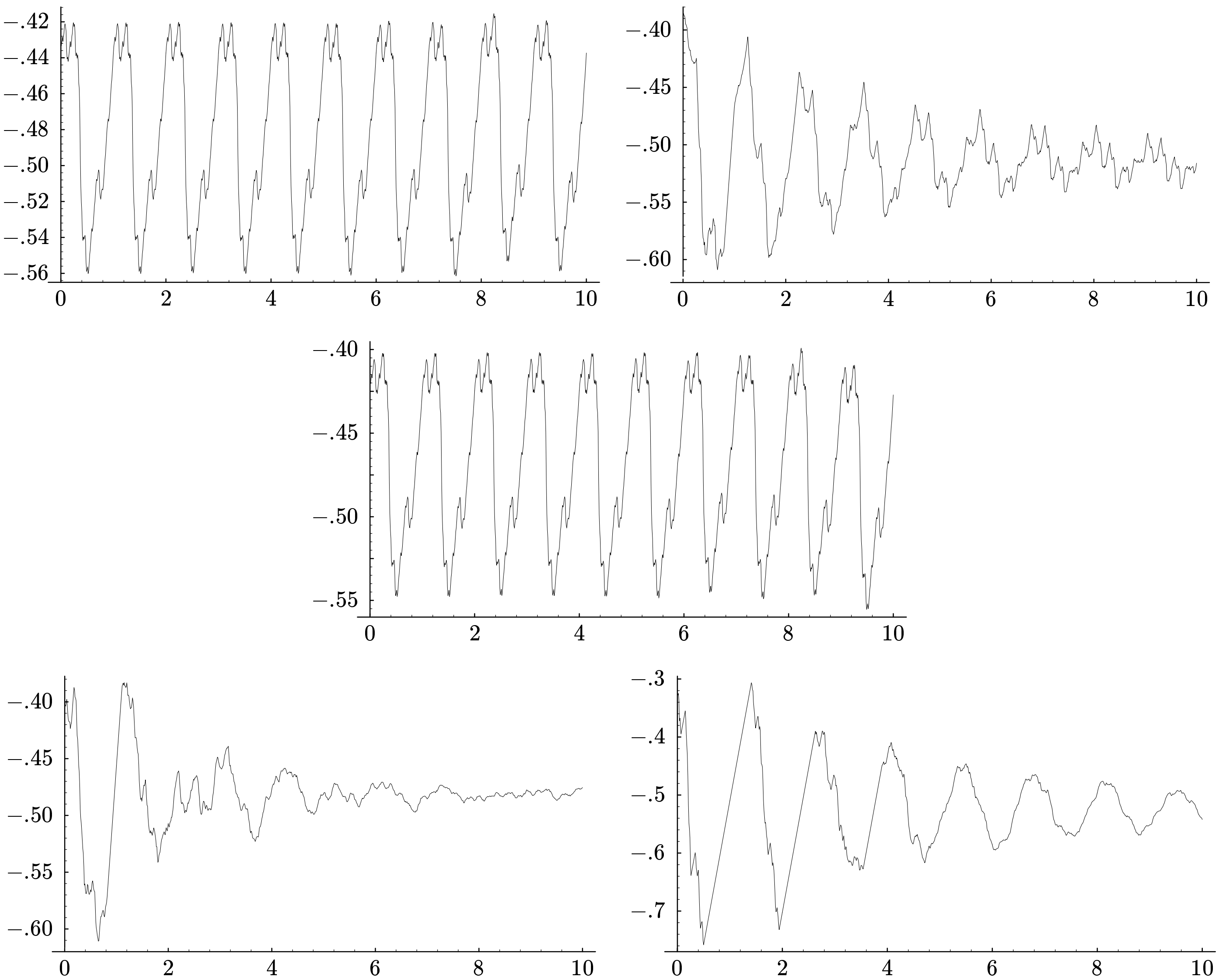

FIGURE 15. Graphs of $\varphi(3^{-s})$ versus s. Top left: Cantor measure; computational error begins to degrade the results around $s = 8$. Top right: measure of Figure 1; convergence to a constant is slow. Middle: measure associated to the Cantor i.f.s. $S_1x = \frac{1}{3}x$, $S_2x = \frac{1}{3}x + \frac{2}{3}$ with weights $p_1 = .4$, $p_2 = .6$; computational error becomes noticeable around $s = 6$. Bottom left: Hausdorff measure on the Julia set of $z^2 - 3.36$. Bottom right: Hausdorff measure on the Julia set of $z^2 - 6$.

In both cases the upper entropy density is finite and the lower entropy density is positive, *and in case* (b) *the entropy density exists*.

Proof. We will show that φ satisfes

$$\varphi(r) = \sum p_j \varphi\Big(\frac{r}{r_j}\Big) \tag{6.5}$$

for all sufficiently small r. The conclusion then follows from the renewal theorem (see [Lau 1992] for a similar application of this method).

We will prove (6.5) for all r less than the minimum separation between first generation islands, so that $B_r(x) \cap K$ lies entirely inside S_jK, for each x in S_jK. Thus

$$\mu(B_r(x)) = p_j \mu(B_{r/r_j}(S_j^{-1}x))$$

and

$$\int_{S_j K} \log \mu(B_r(x))\, d\mu(x) = p_j \log p_j + p_j \int_K \log \mu(B_{r/r_j}(y))\, d\mu(y)$$

after a change of variable. Summing on j we get

$$\varphi(r) + \alpha \log 2r = \sum p_j \log p_j + \sum p_j \Bigl(\varphi\Bigl(\frac{r}{r_j}\Bigr) + \alpha \log \frac{2r}{r_j}\Bigr)$$

which yields (6.5) upon simplification. □

It seems likely that the result remains valid if we drop the separation hypothesis, and assume only that the i.f.s. is nonoverlapping, since the application of the renewal theorem allows an error term in (6.5). The behavior seen in the last two panels of Figure 15 for Hausdorff measure on Julia sets suggests that conclusion (b) of the theorem may well hold for many nonlinear i.f.s.'s and variable weights.

Problem 6.2. Under what conditions does (6.4) give the correct value for the entropy dimension? When does conclusion (b) of Theorem 6.1 hold?

A related notion is the L^2 dimension, defined by

$$\lim_{r\to 0} \frac{\log \int \mu(B_r(x))\, d\mu(x)}{\log r}$$

if the limit exists. See [Strichartz 1993b] for several equivalent definitions. The existence of the limit is proved in [Lau 1992] for linear i.f.s.'s and constant weights, with the value given by the unique solution β of the equation

$$\sum p_j^2 r_j^{-\beta} = 1. \qquad (6.6)$$

The existence of the limit for nonlinear i.f.s.'s and constant weights is proved in [Strichartz 1993b], but there is no effectively computable formula analogous to (6.6) for the value of the limit. The L^2 dimension is related to the asymptotic behavior of the Fourier transform of the measure (for these ideas see the above papers and also [Lau and Wang 1993; Strichartz 1990a; b; 1991; 1993a; 1994; Janardhan et al. 1992].

Once the L^2 dimension β is determined, we can form the function

$$\psi(r) = \int \mu(B_r(x))/(2r)^\beta d\mu(x)$$

and define the upper and lower L^2 densities as the lim sup and lim inf of ψ as $r \to 0$. Actually, slightly different definitions are used in [Lau 1992] and [Strichartz 1993b], which yield equivalent but not identical density values. It is easy to modify the proof of Theorem 6.1 to obtain the analogous result for $\psi(r)$; in place of (6.5) we have

$$\psi(r) = \sum p_j^2 r_j^{-\beta} \psi(r/r_j)$$

for small r. The analogous result for the related definition of L^2 density is proved in [Lau 1992] under weaker hypotheses (nonoverlapping rather than separated i.f.s.).

In Figure 16 we show the graphs of ψ on a log-log scale for the same examples as before. For the linear i.f.s. cases we used (6.6) to determine β, while for the Hausdorff measure on Julia sets we used the Hausdorff dimension.

The density results for the Julia sets suggest that something more is going on. In Figure 17 we give the plot of the maximum and minimum values of

$$\mu(B_r(x))/(2r)^\alpha$$

as x varies over K as a function of $\log r$ for two of these Julia sets. This suggests that there are uniform upper and lower bounds,

$$0 < c_1 \le \mu(B_r(x))/(2r)^\alpha \le c_2 < \infty \qquad (6.7)$$

for all x in K and $r \le 1$. Such estimates would immediately imply the equality of all dimensions and the boundedness (and boundedness away from zero) of all densities.

Problem 6.3. Under what conditions does the Hausdorff measure on an attractor K of an i.f.s. satisfy (6.7)?

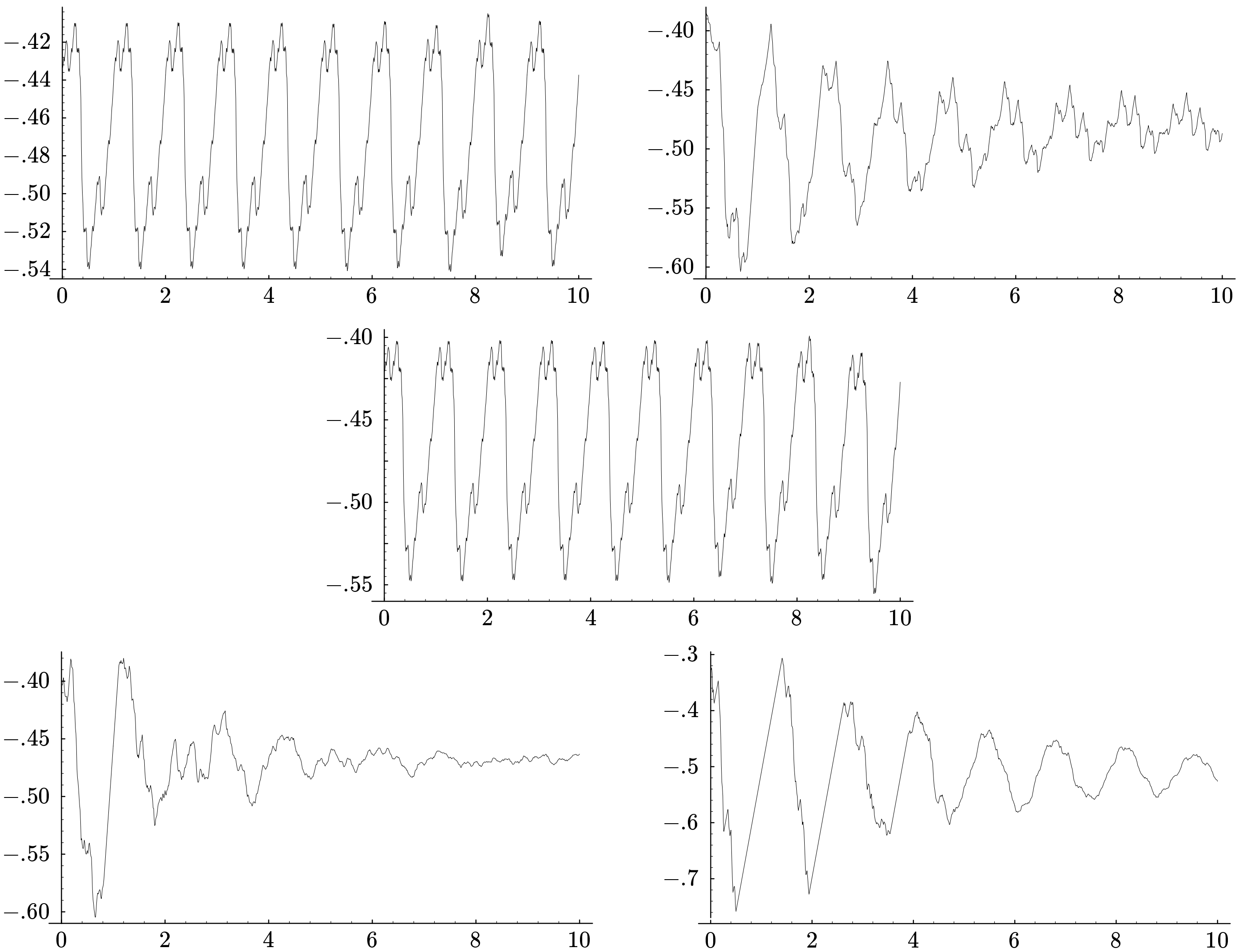

FIGURE 16. Graphs of $\log \psi(3^{-s})$ for the same measures as in Figure 15.

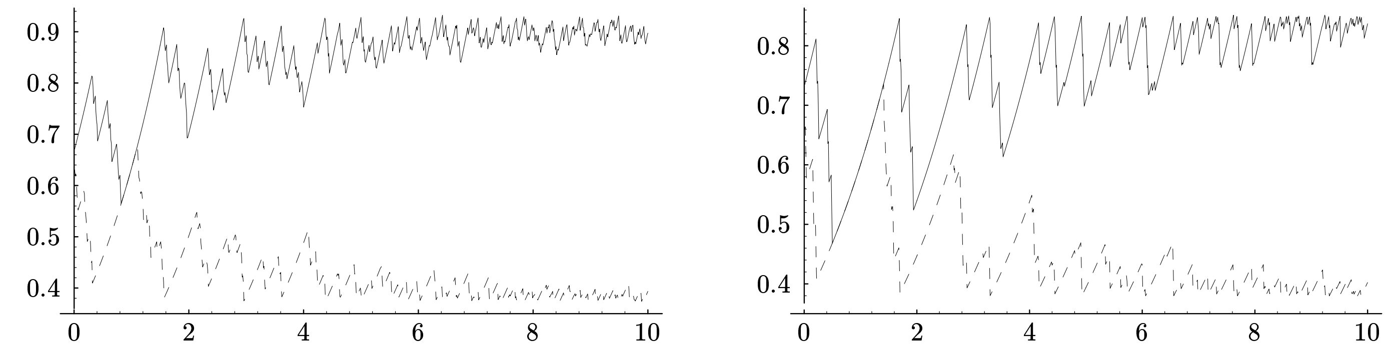

FIGURE 17. Graphs of the maximum and minimum values of $\mu(B_r(x))/(2r)^\alpha$ (as a function of s for $r = 3^{-s}$) as x varies, for the same Julia set measures illustrated in the bottom row of Figures 15 and 16.

ACKNOWLEDGEMENTS

We are very grateful to Amalia Brooks of Oakdale College who assisted in carrying out the experiments reported here. We also thank Larry Brown, Rick Durrett, John Hubbard and John Smillie for helpful comments, and Rolf Riedi and one of the referees for references.

REFERENCES

[Aaronson et al. 1993] J. Aaronson, M. Denker and M. Urbanski, "Ergodic theory for Markov fibred systems and parabolic rational maps", *Trans. Amer. Math. Soc.* **337** (1993), 495–548.

[Ayer and Strichartz 1995] E. Ayer and R. S. Strichartz, "Exact Hausdorff measure and intervals of maximum density for Cantor sets", preprint, Cornell University, 1995.

[Bandt 1989] C. Bandt, "Self-similar sets 3. Constructions with sofic systems", *Monatsch. Math.* **108** (1989), 89–102.

[Bandt 1992] C. Bandt, "The tangential distribution for self-similar measures", talk presented at the Conference on real analysis and number theory, Capri, 1992. Author can be reached at bandt@math-inf.uni-greifswald.d400.de.

[Barnsley 1988] M. Barnsley, *Fractals Everywhere*, Academic Press, Boston, 1988.

[Barnsley and Elton 1988] M. F. Barnsley and J. H. Elton, "A new class of Markov processes for image encoding", *Adv. Appl. Prob.* **20** (1988), 14–32.

[Barnsley et al. 1989] M. F. Barnsley, J. H. Elton and D. P. Hardin, "Recurrent iterated function systems", *Constr. Approx.* **5** (1989), 3–331.

[Bedford and Fisher 1992] T. Bedford and A. M. Fisher, "Analogues of the Lebesgue density theorem for fractal sets", of reals and integers *Proc. London Math. Soc.* (3) **64** (1992), 95–124.

[Blanchard 1984] P. Blanchard, "Complex analytic dynamics on the Riemann sphere", *Bull. Amer. Math. Soc.* **11** (1984), 85–141.

[Bowen 1975] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Math. **470**, Springer, Berlin, 1975.

[Cawley and Mauldin 1992] R. Cawley and R. D. Mauldin, "Multifractal decompositions of Moran fractals", *Advances in Math.* **92** (1992), 196–236.

[Dudley 1966] R. M. Dudley, "Convergence of Baire measures", *Studia Math.* **27** (1966), 251–268.

[Edgar and Mauldin 1992] G. A. Edgar and R. D. Mauldin, "Multifractal decompositions of diagraph recursive fractals", *Proc. London Math.* **65** (1992), 604–628.

[Falconer 1985] K. J. Falconer, *The geometry of fractal sets*, Cambridge University Press, Cambridge, 1985.

[Falconer 1990] K. J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, Wiley, New York, 1990.

[Falconer 1992] K. J. Falconer, "Wavelet transforms and order-two densities of fractals", *J. Stat. Phys.* **67** (1992), 781–793.

[Graf 1993] S. Graf, "On Bandt's tangential distribution of self-similar measures", preprint, University of Passau, Germany, 1993.

[Hofbauer 1992] F. Hofbauer, "Hausdorff and conformal measures for expanding piecewise", monotonic maps of the interval *Studia Math.* **103** (1992), 191–206.

[Hutchinson 1981] J. E. Hutchinson, "Fractals and self-similarity", *Indiana U. Math. J.* **30** (1981), 713–747.

[Janardhan et al. 1992] P. Janardhan, D. Rosenblum and R. S. Strichartz, "Numerical experiments in Fourier asymptotics of Cantor", measures and wavelets *Experimental Math.* **1** (1992), 249–273.

[Lau 1992] K.–S. Lau, "Fractal measures and mean p-variations", *J. Functional Anal.* **108** (1992), 427–457.

[Lau 1993] K.–S. Lau, "Dimension of a family of singular Bernoulli convolutions", *J. Functional Anal.* **116** (1993), 335–358.

[Lau and Wang 1993] K.–S. Lau and J. Wang, "Mean quadratic variations and Fourier asymptotics of self-similar measures", *Monat. Math.* **115** (1993), 99–132.

[Marion 1986] J. Marion, "Mesure de Hausdorff d'un fractal à similitude interne", *Ann. sc. math. Québec* **10** (1986), 51–84.

[Marion 1987] J. Marion, "Mesures de Hausdorff d'ensembles fractals", *Ann. sc. math. Québec* **11** (1987), 111–132.

[Patzschke and Zähle 1993] N. Patzschke and M. Zähle, "Fractal differentiation in the self-affine case III. The density of the Cantor set", *Proc. Amer. Math. Soc.* **117** (1993), 137–144.

[Rachev 1991] S. T. Rachev, *Probability metrics and the stability of stochastic models*, Wiley, New York, 1991.

[Ruelle 1983] D. Ruelle, "Bowen's formula for the Hausdorff dimension of self-similar sets", pp. 351–358 in *Scaling and self-similarity in physics: renormalization in statistical mechanics and dynamics* (edited by J. Frohlich), Progress in Physics **7**, Birkhäuser, Boston, 1983.

[Stein and Weiss 1971] E. M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces*, Princeton University Press, Princeton, 1971.

[Strichartz 1990a] R. S. Strichartz, "Fourier asymptotics of fractal measures", *J. Functional Anal.* **89** (1990), 154–187.

[Strichartz 1990b] R. S. Strichartz, "Self-similar measures and their Fourier transforms I", *Indiana U. Math. J.* **39** (1990), 797–817.

[Strichartz 1991] R. S. Strichartz, "Spectral asymptotics of fractal measures on Riemannian manifolds", *J. Functional Anal.* **102** (1991), 176–205.

[Strichartz 1993a] R. S. Strichartz, " Self-similar measures and their Fourier transforms II", *Trans. Amer. Math. Soc.* **336** (1993), 335–361.

[Strichartz 1993b] R. S. Strichartz, "Self-similar measures and their Fourier transforms III", *Indiana U. Math. J.* **42** (1993), 367–411.

[Strichartz 1994] R. S. Strichartz, "Self-similarity in harmonic analysis", *J. Fourier Anal. Appl.* **1** (1994), 1–37.

[Sullivan 1983] D. Sullivan, "Conformal dynamical systems", pp. 725–752 in *Geometric Dynamics*, Rio de Janeiro, 1981 (edited by J. Palis), Lecture Notes in Math. **1007**, Springer, Berlin, 1983.

[Tsujii 1991] Y. Tsujii, "Markov self-similar sets", *Hiroshima Math. J.* **21** (1991), 491–519.

[Young 1982] L.-S. Young, "Dimension, entropy and Lyapunov exponents", *Ergod. Th. Dynam. Sys.* **2** (1982), 109–124.

ELECTRONIC AVAILABILITY

Programs implementing the algorithms described in this article are available upon request from the third author.

Robert S. Strichartz, Mathematics Department, White Hall, Cornell University, Ithaca, NY 14853 (math@math.cornell.edu)

Arthur Taylor, Mathematics Department, Brown University, Providence, RI 02912 (ataylor@math.cornell.edu)

Tong Zhang, Computer Science Department, Stanford University, Stanford, CA 94305 (tzhang@cs.stanford.edu)

Received December 28, 1993; accepted in revised form May 25, 1995