Open access • Journal Article • DOI:10.1016/0167-2789(83)90125-2

The dimension of chaotic attractors — Source link

J. Doyne Farmer, Edward Ott, James A. Yorke

Institutions: Los Alamos National Laboratory, University of Maryland, College Park

Published on: 01 May 1983 - Physica D: Nonlinear Phenomena (Springer, New York, NY)

Topics: Effective dimension, Complex dimension, Hausdorff dimension, Dimension function and Correlation dimension

PORTIONS

OF THIS

DOCUMEN

ARE

ILLEGIBL

LA-UR-82-2856

CONF-820521--6

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

MASTER

LA-UR--82-2856

DE83 001931

TITLE: THE DIMENSION OF CHAOTIC ATTRACTORS

AUTHOR(S): J. Doyne Farmer, Edward Ott, and James A. Yorke

SUBMITTED TO: To appear in an issue of Physica D to be titled Order in Chaos, the proceedings of a conference held in Los Alamos, May 24-28, 1982.

DISCLAIMER

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

FORM NO 836 R4
ST NO 2629 5/81

September, 1982

The Dimension of Chaotic Attractors

J. Doyne Farmer
Center for Nonlinear Studies
Theoretical Division, MS B258
Los Alamos National Laboratory
Los Alamos, New Mexico 87545

Edward Ott
Laboratory of Plasma and Fusion Energy Studies
University of Maryland
College Park, Maryland

and

James A. Yorke
Institute for Physical Science and Technology
and
Department of Mathematics
University of Maryland
College Park, Maryland

Abstract

Dimension is perhaps the most basic property of an attractor. In this paper we discuss a variety of different definitions of dimension, compute their values for a typical example, and review previous work on the dimension of chaotic attractors. The relevant definitions of dimension are of two general types, those that depend only on metric properties, and those that depend on probabilistic properties (that is, they depend on the frequency with which a typical trajectory visits different regions of the attractor). Both our example and the previous work that we review support the conclusion that all of the probabilistic dimensions take on the same value, which we call the "dimension of the natural measure", and all of the metric dimensions take on a common value, which we call the "fractal dimension". Furthermore, the dimension of the natural measure is typically equal to the Lyapunov dimension, which is defined in terms of Lyapunov numbers, and thus is usually far easier to calculate than any other definition. Because it is computable and more physically relevant, we feel that the dimension of the natural measure is more important than the fractal dimension.

To appear in an issue of Physica D to be titled *Order in Chaos*, the proceedings of a conference held in Los Alamos, May 24-28, 1982.

Table of Contents

I. Introduction

It is the purpose of this paper to discuss and review questions relating to the dimension of chaotic attractors. Before doing so, however, we should first say what we mean by the word "attractor".

1.1. Attractors

In this paper we consider dynamical systems such as maps (discrete time, n)

$$x_{n+1} = F(x_n),$$

or ordinary differential equations (continuous time, t)

$$\frac{dx(t)}{dt} = G(x(t)).$$

where in both cases x is a vector. Thus given an initial value of x (at $n = 0$ for the map or $t = 0$ for the differential equations) an orbit is generated ($(x_1, x_2, \ldots, x_n, \ldots)$ for the map and x(t) for the differential equations). We shall be interested in attractors for such systems. Loosely speaking, an attractor is something that "attracts" initial conditions from a region around it once transients have died out. More precisely, an *attractor* is a compact set, A, with the property that there is a neighborhood of A such that for almost every[1] initial condition the limit set of the orbit as n or $t \to +\infty$ is A. Thus, almost every trajectory in this neighborhood of A passes arbitrarily close to every point of A. The *basin of attraction* of A is the closure of the set of initial conditions that approach A.

We are primarily interested in *chaotic* attractors. We give a definition of chaos in Sec. III, but the reader may also wish to see the reviews given in references [1-4].

1.2. Why Study Dimension?

The dimension of an attractor is clearly the first level of knowledge necessary to characterize its properties. Generally speaking, we may think of the dimension as giving, in some way, the amount of information necessary to specify the position of a point on the attractor to within a given accuracy (cf. Sec. II). The dimension is also a lower bound on the number of essential variables needed to model the dynamics.

For simple attractors, defining and determining the dimension is easy. For example, using any reasonable definition of dimension, a stationary time independent equilibrium (fixed point) has dimension zero, a stable periodic oscillation (limit cycle) has dimension one, and a doubly periodic attractor (2-torus) has dimension two. It is because their structure is very regular that the dimension these simple attractors takes on integer values.

Chaotic (strange) attractors, however, often have a structure that is not simple; they are often not manifolds, and frequently have a highly fractured character. For chaotic attractors, intuition based on properties of regular, smooth examples does not apply. The most useful notions of dimension take on values that are typically not integers.

To fully understand the properties of a chaotic attractor, one must take into account not only the attractor itself, but also the "distribution" or "density" of points on the attractor. This is more precisely discussed in terms of the *natural*

[1] The phrase "almost every" here signifies that the set of initial conditions in this neighborhood for which the corresponding limit set is not A can be covered by a set of arbitrarily small volume (i.e. has Lebesgue measure zero).

measure associated with a given attractor. The natural measure provides a notion of the relative probability of occurence of different regions of the attractor. Just as chaotic attractors can have very complicated properties, the natural measures of chaotic attractors often have complicated properties that make the relevant assignment of a dimension a nontrivial problem.

Precise definitions of such terms as "natural measure" follow, but we would first like to give an example in order to motivate the central questions we are addressing in this paper.

Consider the following two dimensional map[2]:

$$\begin{aligned} x_{n+1} &= x_n + y_n + \delta \cos 2\pi y_n \quad &\mathit{mod}\ 1 \\ y_{n+1} &= x_n + 2y_n &\mathit{mod}\ 1 \end{aligned} \tag{1}$$

For small values of δ, Sinai [5] has shown that the attractor of this map is the entire square, and is thus of dimension 2. Therefore almost every initial condition generates a trajectory that eventually comes arbitrarily close to every point on the

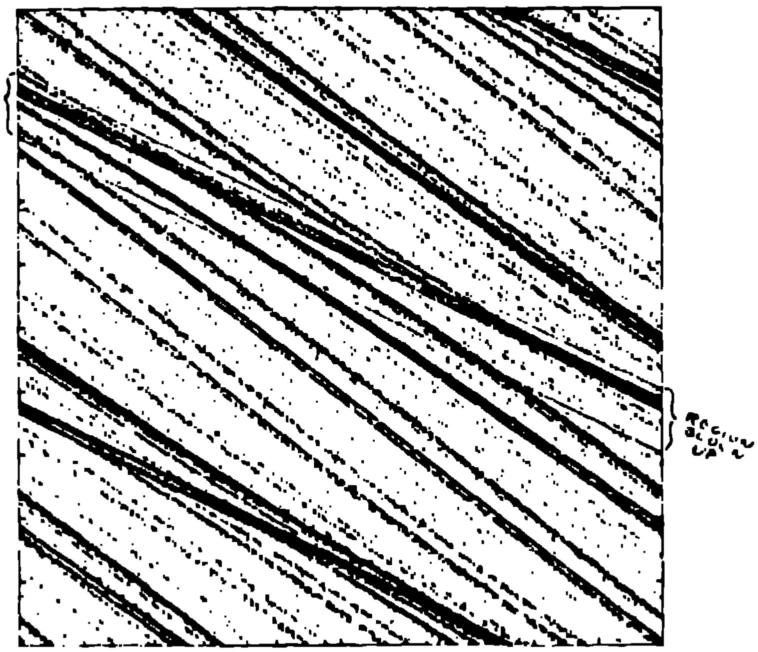

Figure 1

Successive iterates of the initial point $x_0 = 0.5$, $y_0 = 0.5$ using Eq. (1) with $\delta = 0.1$. 80,000 points are shown. Almost any initial condition gives a qualitatively similar plot, the location of the individual points of course changes, but the location of the dark bands does not. The density of these points is described by the ***natural measure*** of this attractor. (For example, the outlined parallelogram (which is blown up in Fig 2) contains approximately 27% of the points of a typical trajectory, and thus can be said to have a natural measure of approximately 0.27.)

---

[2]Mod 1 means that the values of x and y are truncated to be less than or equal to one and their integer part are discarded, so that the map is defined on the unit square.

square. However, consider the typical trajectory shown in Fig. 1. Certain regions are visited far more often than others. The natural measure of a given region is proportional to the frequency with which it is visited (see Sec. 2.2.2), in this case the natural measure is highly concentrated in diagonal bands whose density of points is much greater than the average.[8]. Furthermore, as shown in Fig. 2, if a microscope is used to magnify a small piece of the attractor, the same sort of structure is still seen.

For this map we do not know if the value of δ chosen to construct Fig. 1 is small enough to insure that the dimension of the attractor is two. For practical purposes, though, this may be irrelevant. Even if a trajectory eventually comes arbitrarily close to any given point, the amount of time required for this to happen may be enormous. In order to assign a relevant dimension that will characterize the trajectories on the attractor, the natural measure must be taken into account. For this example the dimension that characterizes properties of the natural measure is between one and two.

These considerations are not as esoteric as they might seem. We are ultimately not as interested in whether the dimension of a given attractor is 3.1 or 3.2 as we are in whether it is on the order of three or on the order of thirty. As we shall see, a proper understanding of *probabilistic* notions of dimension leads to an

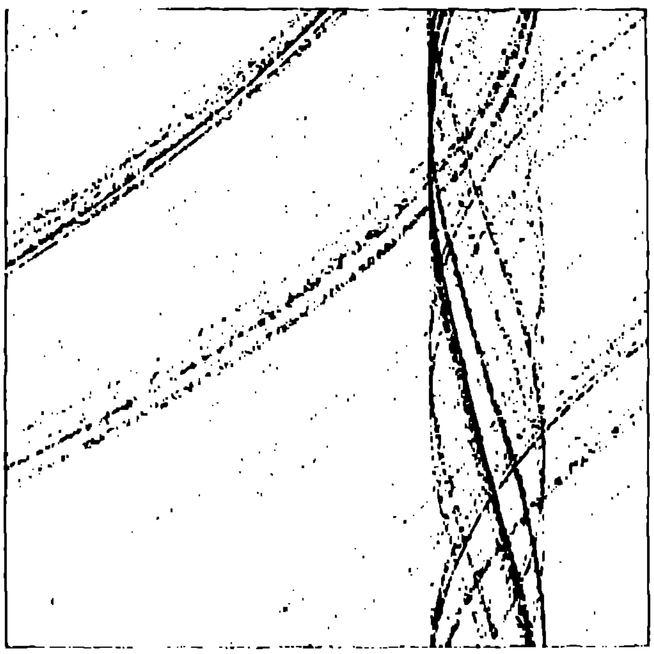

Figure 2

A blow-up of the strip marked in Fig. 1. Expanding it horizontally, What appears to be a single band in Fig. 1 is now seen to be a collection of bands.

[8]In fact, for small values of δ, Sinai [5] has shown that for any $\varepsilon > 0$, there exists a collection of tiny squares whose total area is less than ε, and such that almost every trajectory spends $1 - \varepsilon$ of the time inside this collection of squares. These squares cover what is called the *core* of the attractor. (See Sec. VII.)

efficient method of computing the dimension of chaotic attractors, that provides the best known method of answering such questions.

The main points of this paper can be summarized as follows:

1. Although there are a variety of different definitions of dimension, the relevant definitions are of two types, those which only depend on metric properties, and those which depend on metric and probabilistic properties (i.e., they involve the natural measure of the attractor).
2. Current evidence supports the conclusion that all of the metric dimensions typically take on the same value, and all of the probabilistic dimensions also typically take on the same value.
3. Current evidence supports a conjectured relationship whereby the dimension of the natural measure can be found from a knowledge of the stability properties of an orbit on the attractor (i.e., knowledge of the Lyapunov numbers).
4. For typical chaotic attractors we conjecture that the distribution of frequencies with which an orbit visits different regions of the attractor is, in a certain sense, log-normal (Sec. V).

Points 1-3 are summarized in Table 1. The first two entries in the table are metric dimensions, while the next five are probabilistic dimensions. Under the hypothesis that all the metric dimensions yield the same value (Point 2), we call this value the ***fractal dimension*** and denote it d_F. Similarly, if all the probabilistic dimensions yield the same value, we call this value the ***dimension of the natural measure***, and denote it d_μ. Although in special cases d_F equals d_μ, typically $d_F > d_\mu$. Finally, the last entry in Table 1, the Lyapunov dimension, is by definition the predicted value of d_μ obtained from the Lyapunov numbers (cf. Point 3). The Lyapunov dimension is in a different category than the other dimensions listed, since it is defined in terms of dynamical properties of an attractor, rather than

Name of Dimension	Symbol	Generic Name	Symbol
capacity Hausdorff dimension	d_C d_H	fractal dimension	d_F
information dimension ϑ-capacity ϑ-Hausdorff dimension pointwise dimension Hausdorff dimension of the core	d_I $d_C(\vartheta)$ $d_H(\vartheta)$ d_p $d_H(core)$	dimension of the natural measure	d_μ
Lyapunov dimension	d_L		

Table 1.

Current evidence indicates that typically the two dimensions in the first box take on the same value, called the fractal dimension, while the five dimensions in the second box take on another typically smaller value, called the dimension of the measure.

metric and natural measure properties.

1.3. Outline

This paper is organized as follows: In Sec. II we give several definitions of dimension. Sec. III reviews conjectures relating Lyapunov numbers to dimension. These conjectures are particularly important because the Lyapunov numbers provide the only known efficient method to compute dimension. In sections IV, V, VI, and VII, we compute all the dimensions discussed here for an explicitly soluable example, the generalized Baker's transformation. In addition, based on this example, in Sec. V we propose a new conjecture concerning the frequency with which different values of the probability occur. Section VII gives a discussion of the "core" of attractors, and Sec. VIII gives another example supporting the connection between Lyapunov numbers and dimension (an attractor which is topologically a torus but is nowhere differentiable). Section IX reviews relevant results from numerical computations of the dimension of chaotic attractors. Concluding remarks are given in Sec. X.

In general terms, this paper has two functions. One is to present a review of the current status of research on the dimension of chaotic attractors. The other purpose is to present new results (Secs. IV-VI).

II. Definitions of Dimension

In this section we define and discuss six different concepts of dimension. The first two of these, the capacity and the Hausdorff dimension, require only a metric (i.e., a concept of distance) for their definition, and consequently we refer to them as "metric dimensions". The other dimensions we will discuss in this section are the information dimension, the ϑ-capacity, the ϑ-Hausdorff dimension, and the pointwise dimension. These dimensions require both a metric and a probability measure for their definition, and hence we will refer to them as "probabilistic dimensions".

In this paper we compute the values of these dimensions for an example that we believe is general enough to be "typical" of chaotic attractors, at least regarding the question of dimension. We find that the metric dimensions take on a common value. Whenever this is the case, we will refer to this common value d_F as the ***fractal dimension***[4]. For our example we also find that the probabilistic dimensions take on a common value d_μ, which we will refer to as the *dimension of the natural measure*. As we summarize in Conjecture 1, we feel that this equality is a general property, true for typical cases.

> *Conjecture 1.* For a typical chaotic attractor the capacity and Hausdorff dimensions have a common value d_F, and the information dimension, ϑ-capacity, ϑ-Hausdorff dimension, and pointwise dimensions have a common value d_μ, i.e.
>
> $$d_C = d_H = d_F$$
>
> and
>
> $$d_I = d_C(\vartheta) = d_H(\vartheta) = d_P = d_\mu$$

---

[4]The term ***fractal*** was originally coined by Mandelbrot [6]. However, he uses "fractal dimension" as a synonym for Hausdorff dimension. We should also mention that in some of our previous papers on this subject [7-10], we used the term "fractal dimension" as a synonym for capacity, rather than our current usage as described in the text.

Note: For the case of diffeomorphisms in two dimensions, L.S. Young has rigorously proven that information dimension, pointwise dimension, and the Hausdorff dimension of the core (see Sec. VII) all take on the same value [11].

In addition to the dimensions defined in this section, we will also discuss three others[5]: the Lyapunov dimension, the capacity of the core, and the Hausdorff dimension of the core. Lyapunov dimension is discussed in Sec. III, and the latter two dimensions are discussed in Sec. VII. For our example the Lyapunov dimension and Hausdorff dimension of the core are equal to d_μ, while the capacity of the core is equal to d_F.

2.1. Metric Dimensions

We begin by discussing two concepts of dimension which apply to sets in spaces on which a concept of distance, i.e., a metric is defined. In particular we begin by discussing the capacity and the Hausdorff dimension.

2.1.1. Capacity

The capacity of a set was originally defined by Kolmogorov [13]. It is given by

$$d_C = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log \left(\frac{1}{\varepsilon}\right)} \tag{2}$$

where, if the set in question is a bounded subset of a p-dimensional Euclidean space $\mathbf{R}^p$, then $N(\varepsilon)$ is the minimum number of p-dimensional cubes of side ε needed to cover the set. For a point, a line, and an area, $N(\varepsilon) \sim 1$, $N(\varepsilon) \sim \varepsilon^{-1}$, and $N(\varepsilon) \sim \varepsilon^{-2}$, and Eq. (2) yields $d_C = 0$, 1, and 2, as expected. However, for more general sets (dubbed *fractals* by Mandelbrot), d_C can be noninteger. For example, consider the Cantor set obtained by the limiting process of deleting middle thirds, as illustrated in Fig. 3. If we choose $\varepsilon = (1/3)^m$, then $N = 2^m$, and Eq (2) yields

$$d_C = \frac{\log 2}{\log 3} = 0.630\ldots$$

If one is content to know where the set lies to within an accuracy ε, then to specify the location of the set, we need only specify the position of the $N(\varepsilon)$ cubes covering the set. Equation (2) implies that for small ε, $\log N(\varepsilon) \approx d_C \log \left(\frac{1}{\varepsilon}\right)$. Hence, the dimension tells us how much information is necessary to specify the location of the set to within a given accuracy. If the set has a very fine scaled structure (typical of chaotic attractors), then it may be advantageous to introduce some coarse-graining into the description of the set. In this case, ε may be thought of as specifying the degree of coarse-graining.

2.1.2. Hausdorff Dimension

The capacity may be viewed as a simplified version of the Hausdorff dimension, originally introduced by Hausdorff in 1919 [14]. (We have reversed historical order and defined capacity before Hausdorff dimension because the definition of Hausdorff dimension is more involved.) We believe that for attractors these two dimensions are generally equal. While it is possible to construct simple examples of sets

---

[5]Note that in this paper we will not discuss the concept of *topological dimension*, since its application to chaotic dynamics is not clear. Its value is an integer and it is generally equal to neither d_F nor d_μ. For discussions of topological dimension, we refer the reader to Hurwicz and Wallman [12].

[6]Sets can be constructed for which the limit of Eq. (2) does not exist. We would then say that the capacity is not defined.

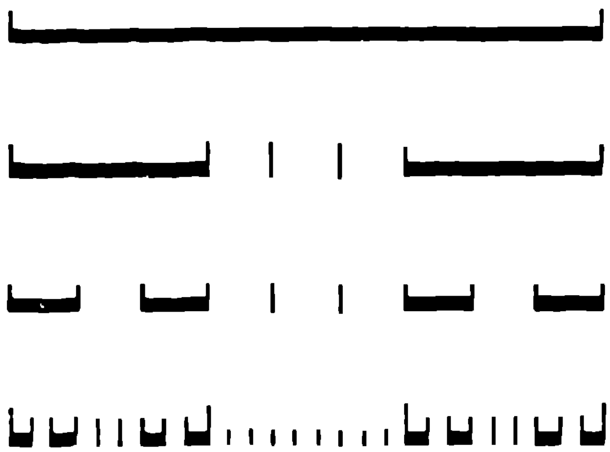

Figure 3

The first few steps in the construction of the classic example of a Cantor set.

where the Hausdorff dimension and the capacity are unequal[7], these do not seem to apply to attractors. (Although they may apply to the core of attractors. See Sec. VII.)

To define the Hausdorff dimension of a set lying in a p dimensional Euclidean space, consider a covering of it with p-dimensional cubes of variable edge length ε_i. Define the quantity $l_d(\varepsilon)$ by

$$l_d(\varepsilon) = \inf \sum_i \varepsilon_i^d,$$

where the infimum (i.e. minimum) extends over all possible coverings subject to the constraint that $\varepsilon_i \le \varepsilon$. Now let

$$l_d = \lim_{\varepsilon \to 0} l_d(\varepsilon).$$

Hausdorff showed that there exists a critical value of d above which $l_d = 0$ and below which $l_d = \infty$. This critical value, $d = d_H$, is the Hausdorff dimension. (Precisely at $d = d_H$, l_d may be either 0, ∞, or a positive finite number.) This concept of dimension will be used in Secs. IV, VI, and VII. It is easy to see that $d_C \ge d_H$[8].

[7]For example, for the set of numbers $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\ldots$, the Hausdorff dimension is zero while (1) yields $d_C = \frac{1}{2}$.

[8]To show that $d_C \ge d_H$, consider a covering consisting of cubes of equal side $\varepsilon_i = \varepsilon$. Then due to the infimum in the definition of $l_d(\varepsilon)$, we see that $\bar{l}_d(\varepsilon) = \sum_i \varepsilon^d = N(\varepsilon)\varepsilon^d$ satisfies $\bar{l}_d(\varepsilon) \ge l_d(\varepsilon)$.

2.2. Probabilistic Dimensions

2.2.1. The Natural Measure on an Attractor

Note that, in computing d_C from Eq. (2), all cubes used in covering the attractor are equally important even though the frequencies with which an orbit on the attractor visits these cubes may be very different. In order to take the frequency with which each cube is visited into account, we need to consider not only the attractor itself, but the relative frequency with which a typical orbit visits different regions of the attractor as well. We can say that some regions of the attractor are more probable than others, or alternatively we may speak of a measure on the attractor[9]. We define the natural measure of an attractor as follows: For each cube C and initial condition x in the basin of attraction, define μ(x,C) as the fraction of time that the trajectory originating from x spends in C.[10] If almost every such x gives the same value of μ(x,C), we denote this value μ(C) and call μ the *natural measure* of the attractor[15]. The natural measure gives the relative probability of different regions of the attractor as obtained from time averages, and therefore is the "natural" measure to consider. We will assume throughout that any attractor we consider has a natural measure, at least whenever C is one of the cubes we are using to cover the attractor.

The four definitions discussed in the remainder of this section are defined for attractors with a metric and a natural measure defined on them.

2.2.2. Information Dimension

The *information dimension*, d_I, is a generalization of the capacity that takes into account the relative probability of the cubes used to cover the set. This dimension was originally introduced by Balatoni and Renyi [16].

The information dimension is given by

$$d_I = \lim_{\varepsilon \to 0} \frac{I(\varepsilon)}{\log\left(\frac{1}{\varepsilon}\right)}, \tag{3}$$

where

$$I(\varepsilon) = \sum_{i=1}^{N(\varepsilon)} P_i \log \frac{1}{P_i},$$

and P_i is the probability contained within the i^{th} cube. Letting the i^{th} cube of side ε be C_i, $P_i = \mu(C_i)$. Note that if all cubes have equal probability then $I(\varepsilon) = \log N(\varepsilon)$, and hence $d_C = d_I$. However, for unequal probabilities $I(\varepsilon) < \log N(\varepsilon)$. Thus, in general, $d_C > d_I$.

In information theory the quantity I(ε) defined in Eq (3) has a specific meaning [17]. Namely, it is the amount of information necessary to specify the state of the system to within an accuracy ε, or equivalently, it is the information obtained in making a measurement that is uncertain by an amount ε. Since for small ε, $I(\varepsilon) \approx d_I \log \frac{1}{\varepsilon}$, we may view d_I as telling how fast the information necessary to specify a point on the attractor increases as ε decreases. (For a more extensive discussion of the physical meaning of the information dimension, see Refs 7 and

Thus taking the limit $\varepsilon \to 0$ and making use of Eq. (2) we see that $d_H > d_C$.

[9]Although there are many measures possible for a given attractor, we are only interested in one of them, the natural measure.

[10]$\mu(x,C) = \lim_{T\to\infty} \mu_T(x,C)$, where $\mu_T(x,C)$ is the fraction of time spent in C up to some finite time T.

8.)

2.2.3. ϑ-Capacity

Another definition of dimension which we shall be interested in is what we will call the ϑ-capacity, $d_C(\vartheta)$. Essentially, this quantity is the capacity of that part of the attractor of highest probability,

$$d_C(\vartheta) = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon;\vartheta)}{\log \frac{1}{\varepsilon}} \tag{4}$$

where $N(\varepsilon;\vartheta)$ is the minimum number of cubes of side ε needed to cover at least a fraction ϑ of the natural measure of the attractor. In other words, the cubes must be chosen so that their combined natural measure is at least ϑ. Thus $d_C(1) = d_C$. For the examples we study here, we find that for any value of $\vartheta < 1$, the ϑ-capacity is independent of ϑ, but that $d_C(\vartheta)$ for $\vartheta < 1$ may differ from its value at $\vartheta = 1$. In particular $d_C(\vartheta) = d_\mu$ for $\vartheta < 1$ and $d_C(\vartheta) = d_C$ for $\vartheta = 1$. ϑ—capacity was originally defined by Frederickson et al. [7]. Similar quantities have also been defined by Ledrappier [18], and Mandelbrot [19].

2.2.4. ϑ-Hausdorff Dimension

In analogy with the relationship between capacity (a metric dimension) and ϑ—capacity (a probability dimension), we introduce here a probability dimension based on the Hausdorff dimension. We call this new dimension the ϑ-Hausdorff dimension and denote it $d_H(\vartheta)$. To define the ϑ-Hausdorff dimension, modify the definition of Hausdorff dimension as follows: Define $l_d(\varepsilon,\vartheta)$ by

$$l_d(\varepsilon,\vartheta) = \inf \sum_i \varepsilon_i^d,$$

where now the infimum extends over all possible $\varepsilon_i < \varepsilon$ which cover a fraction ϑ of the total probability of the set. We define $d_H(\vartheta)$ as that value of d below which $l_d(\vartheta) = \infty$ and above which $l_d(\vartheta) = 0$, where $l_d(\vartheta) = \lim_{\varepsilon \to 0} l_d(\varepsilon,\vartheta)$. This concept of dimension will be used in Sec. VI.

2.2.5. Pointwise Dimension

Roughly speaking, the pointwise dimension d_p is the exponent with which the total probability contained in a ball decreases as the radius of the ball decreases. To make this notion more precise, let μ denote the natural probability measure on the attractor, and let $B_\varepsilon(x)$ denote a ball of radius ε centered about a point x on the attractor. Roughly speaking, $\mu(B_\varepsilon(x)) \sim \varepsilon^{d_p}$. More precisely, Young [11] defines this dimension as

$$d_p(x) = \lim_{\varepsilon \to 0} \frac{\log \mu(B_\varepsilon(x))}{\log \varepsilon}. \tag{5}$$

If $d_p(x)$ is independent of x for almost all x with respect to the measure μ,[11] we call $d_p(x) = d_p$ the *pointwise* dimension. Similar definitions of dimension have also been given by Takens [20] and Janssen and Tjon [21].

2.3. Using a Grid of Cubes to Compute Dimension

Some of the definitions we have used, such as the capacity, allow any location or orientation of the cubes used to cover the attractor. In a numerical

[11] By "almost all x with respect to the measure μ" we mean that the set of x which does not satisfy this is a set of μ measure zero.

experiment, however, it is much more convenient to select the cubes used to cover the attractor out of a fixed grid, as shown in Fig. 4. For these dimensions (d_C, d_I, and $d_C(\vartheta)$) it can be shown that selecting from a fixed grid of cubes gives the same value of the dimension as an optimal collection of cubes. For example, for the case of an attractor in a two dimensional space, using a fixed grid to compute $N(\varepsilon)$ in Eq. (2) results in an increase of at most a factor of four in $N(\varepsilon)$, which has no effect on the value of the dimension. Note that this is *not* true for the Hausdorff dimension, which requires a more general cover.

In principle, the definitions of dimension given in this section and the use of a fixed grid provide specific prescriptions for obtaining capacity, information dimension, and ϑ-capacity. To find approximate values for these dimensions, one can generate an orbit on the attractor using a computer, and then divide the space containing the orbit into cubes of side ε in order to estimate the numbers $N(\varepsilon)$, $I(\varepsilon)$, or $N(\varepsilon;\vartheta)$. By examining how $N(\varepsilon)$, $I(\varepsilon)$, and $N(\varepsilon;\vartheta)$ vary as ε is decreased the value of these dimensions can be estimated.

As discussed in Sec. IX, however, in practice the agenda described above for computing dimension may be difficult, costly, or impossible. Thus it is of interest to consider other means of obtaining the dimension of chaotic attractors. The next section deals with this question. In particular, we discuss a conjecture that the dimension of chaotic attractors can be determined directly from the dynamics in terms of Lyapunov numbers

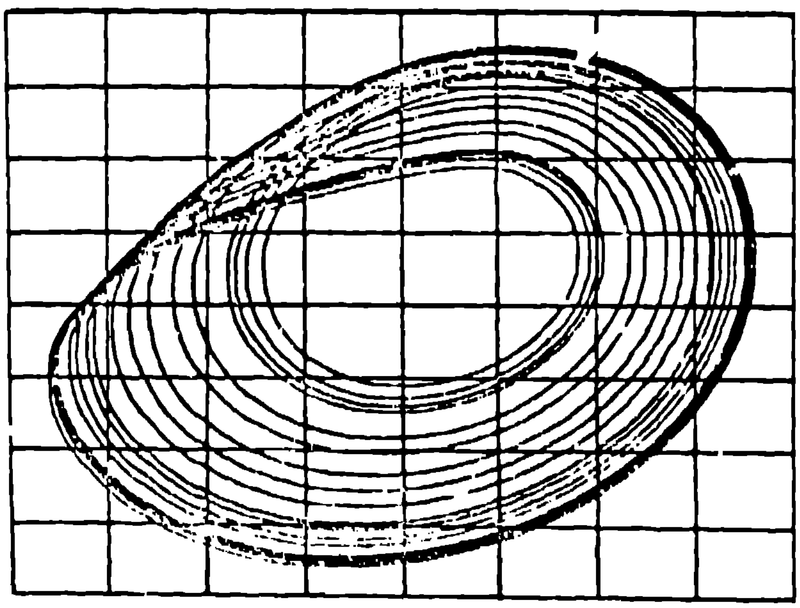

Figure 4

The region of phase space containing an attractor can be divided with a fixed grid of cubes, which can be used to compute capacity, information dimension, or ϑ-capacity.

III. Lyapunov Numbers and Lyapunov Dimension

The Lyapunov numbers quantify the average stability properties of an orbit on an attractor. For a fixed point attractor of a mapping, the Lyapunov numbers are simply the absolute values of the eigenvalues of the Jacobian matrix evaluated at the fixed point. The Lyapunov numbers generalize this notion for more complicated attractors. As we shall see, for a typical attractor there is a connection between average stability properties and dimension. The possiblity of such a connection was first pointed out by Kaplan and Yorke [22] and later by Mori [23].

3.1. Definition of Lyapunov Numbers

For expository purposes, for most of this paper we shall consider p-dimensional maps,

$$x_{n+1} = F(x_n),$$

where x is a p-dimensional vector. We emphasize, however, that similar considerations to those below apply to flows (e.g., systems of differential equations), including infinite dimensional systems such as partial differential equations. To define the Lyapunov numbers, let $J_n = [J(x_n)J(x_{n-1})\ldots J(x_1)]$ where J(x) is the Jacobian matrix of the map, $J(x) = \frac{\partial F}{\partial x}$, and let $j_1(n) > j_2(n) \geq \cdots \geq j_p(n)$ be the magnitudes of the eigenvalues of J_n. The Lyapunov numbers are

$$\lambda_i = \lim_{n\to\infty} [j_i(n)]^{\frac{1}{n}}, \qquad i = 1, 2, \ldots, p, \tag{6}$$

where the positive real n^{th} root is taken. The Lyapunov numbers generally depend on the choice of the initial condition x_1. The Lyapunov numbers were originally defined by Oseledec [24]. We have the convention

$$\lambda_1 > \lambda_2 > \ldots > \lambda_p$$

For a two dimensional map, for example, λ_1 and λ_2 are the average principal stretching factors of an infinitesimal circular area (cf. Fig. 5.). For a chaotic

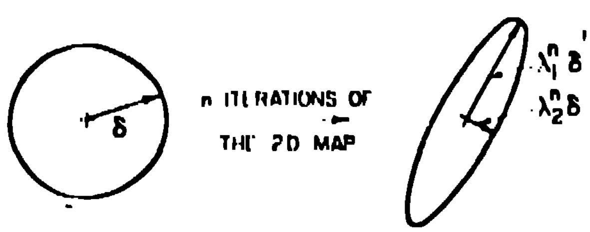

Figure 5

n iterations of a two dimensional map transform a sufficiently small circle of radius δ approximately into an ellipse with major and minor radii $(\lambda_1)^n \delta$ and $(\lambda_2^n)\delta$, where λ_1 and λ_2 are the Lyapunov numbers.

attractor on the average nearby points initially diverge at an exponential rate, and hence at least one of the Lyapunov numbers is greater than one. This makes quantitative the notion of "sensitive dependence on initial conditions". We will take $\lambda_1 > 1$ as our definition of *chaos*. (Note that many authors refer to *Lyapunov exponents* rather than Lyapunov numbers. The Lyapunov exponents are simply the logarithms of the Lyapunov numbers.)

In this paper we assume that *almost every* initial condition in the basin of any attractor that we consider has the same Lyapunov numbers. Thus, the spectrum of Lyapunov numbers may be considered to be a property of an attractor. This assumption is supported by numerical experiments [25]. Exceptional trajectories, such as unstable fixed points on the attractor, typically do not sample the whole attractor and thus typically have Lyapunov numbers that are different from those of the attractor. Those points in the basin of attraction that have different Lyapunov numbers or for which Lyapunov numbers do not exist are here assumed to be of measure zero. (In other words, they may be covered by a collection of cubes of varying size having arbitrarily small total volume).

3.2. Definition of Lyapunov Dimension

The following discussing contains a heuristic argument that motivates a connection between Lyapunov numbers and dimension. Consider a two dimensional map. Suppose we wish to compute the capacity of a chaotic attractor, for which $\lambda_1 > 1 > \lambda_2$. Cover the attractor with $N(\varepsilon)$ squares of side ε. Now, iterate the map q times. For q fixed and ε small enough, the action of the mapping is roughly linear over the square, and each square will be stretched into a long thin parallelogram. From the definition of the Lyapunov numbers, the average length of these parallelograms is $(\lambda_1)^q \varepsilon$, and the average width is $(\lambda_2)^q \varepsilon$. Now, suppose we had used a finer cover of squares of side $(\lambda_2)^q \varepsilon$. (See Fig. 6.) To cover each parallelogram takes about $(\frac{\lambda_1}{\lambda_2})^q$ smaller squares. Thus, *if it is supposed* that all squares on the attractor behave in this typical way, then one is lead to the estimate

$$N(\lambda_2^q \varepsilon) \approx (\frac{\lambda_1}{\lambda_2})^q N(\varepsilon). \qquad (7)$$

Motivated by Eq. (2), assume $N(\varepsilon) \approx k(\frac{1}{\varepsilon})^{d_C}$, and substitute into both side of Eq. (7). This gives

$$k(\frac{1}{\lambda_2^q \varepsilon})^{d_C} \approx k(\frac{\lambda_1}{\lambda_2})^q (\frac{1}{\varepsilon})^{d_C}.$$

Collecting terms, taking logarithms, and solving for d_C gives

$$d_C = 1 + \frac{\log \lambda_1}{\log \frac{1}{\lambda_2}}.$$

We will see that this expression is often meaningful even when this heuristic derivation is invalid, so we will call it the *Lyapunov dimension* d_L.

$$d_L = 1 + \frac{\log \lambda_1}{\log \frac{1}{\lambda_2}} \qquad (8)$$

Generalization of the above heuristic argument to p dimensional maps gives (c.f. Ref. [7])

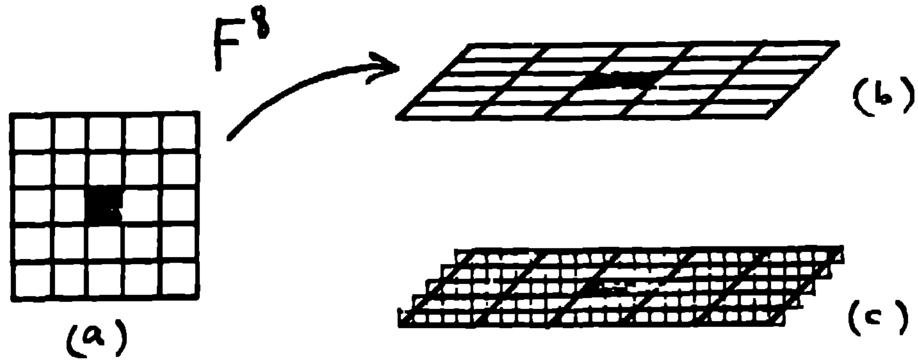

Figure 6

A schematic illustration of the heuristic argument for the Lyapunov dimension. The image of each small square in (a) is approximately a parallelogram which has been stretched horizontally by a factor of λ_1^8 and contracted vertically by a factor λ_2^8. The images in (b) thus require a smaller cover of squares as shown in (c).

$$d_L = k + \frac{\log(\lambda_1\lambda_2\ldots\lambda_k)}{\log\frac{1}{\lambda_{k+1}}}, \tag{9}$$

where k is the largest value for which $\lambda_1\lambda_2\ldots\lambda_k > 1$. If $\lambda_1 < 1$, define $d_L = 0$, if $\lambda_1\lambda_2\ldots\lambda_p > 1$, define $d_L = p$. We shall refer to d_L as the *Lyapunov dimension*. This quantity was originally defined by Kaplan and Yorke [22], who originally gave it as a lower bound on the fractal dimension.

From the above argument one might be tempted to guess that $d_C = d_L$. The Lyapunov numbers are *average* quantities, however, and to compute an average, each cube must be weighted according to its probability. The capacity does not distinguish between probable and improbable cubes. To understand how some cubes might have vastly different probabilities than others, consider an atypical square of a two dimensional map. If the area of the images of this square decreases half as fast as the average for k iterations, then its k^{th} image will be 2^k times larger than the image of a typical square, and the number of squares needed to cover it will be 2^k times greater than the typical value. In fact, as will be evident from considerations of explicit examples (cf. Sec. V), it is commonly the case that the vast majority of cubes needed to cover the attractor are atypical, and do not represent the properties of time averages. By this we mean that all the atypical cubes taken together contain an extremely small fraction of the total probability on the attractor yet account for almost all of N(ε). Furthermore, this tendency increases as ε decreases. The behavior of the atypical cubes under iteration is in general not described by the Lyapunov numbers. It is clear, then, that in order for this estimate to be valid, we must consider only the more probable cubes, i.e., the estimate should be in terms of the dimension of the natural measure rather than the capacity. Assuming the equality of probabilistic dimensions (Conjecture 1), we are led to the following conjecture:

> *Conjecture 2:* For a typical[12] attractor $d_\mu = d_L$.

In the following six sections we present evidence supporting this conjecture. Also, L.S. Young has proved some rigorous results along these lines, which are reviewed in the next subsection.

In the special case that *every* initial condition on the attractor generates the same Lyapunov numbers, we will say that the attractor has *absolute* Lyapunov numbers. In this case it is not necessary to distinguish probable from improbable cubes, and the above conjecture can be made in terms of the fractal dimension rather than the dimension of the natural measure. We call this Conjecture 3.

> *Conjecture 3:* If $\lambda_1, \lambda_2, \ldots \lambda_p$ generated from every (not just almost every) initial condition take on the same value, and if $\lambda_1 > 1$, then for a typical[12] attractor of this type $d_F = d_L = d_\mu$.

The requirement of Conjecture 3 that every initial condition on the attractor generate the same Lyapunov numbers is very restrictive and only holds for special cases. For example, it holds if the Jacobian matrix of the map is independent of x. In more general cases, the requirement of Conjecture 3 would be expected to fail because of the existence of unstable fixed and periodic points on the attractor. For example, if x_1 is chosen to be precisely on an unstable fixed point, the Lyapunov numbers generated will simply be the eigenvalues of $J(x_1)$. These will typically be different from those generated by a chaotic orbit on the attractor. Examples for which Conjecture 3 is valid will be special cases of the more general example presented in the following section. In addition, an example for which Conjecture 3 can be proven to hold is given in Sec. VIII.

3.3. Review of Rigorous Results Concerning Lyapunov Dimension

In addition to the analytic and numerical evidence we will give for conjectures 1-3 in the remainder of this paper, there are several rigorous results supporting these statements which are reviewed in this section. For example Ledrappier [8] has proven an inequality that is somewhat similar to Conjecture 2. In particular, he defines a dimension that we will call d_{led}, which is the ϑ-capacity in the limit as ϑ goes to one, i.e.

$$d_{led} = \lim_{\vartheta \to 1} d_C(\vartheta).$$

For C^2 diffeomorphisms[13] he has shown that

---

[12] The reason for the word "*typical*" is that there exist examples of maps that do not satisfy $d_\mu = d_L$. These maps are exceptional, however, in that arbitrarily small perturbations of them restore the conjectured equality of d_μ and d_L. An example of such an atypical case is where a point x_o is attracting and yet has $\lambda_1 = 1$ (i.e., the Jacobian matrix $\frac{\partial F}{\partial x}$ has an eigenvalue +1 at x_o). The attraction here is due to higher order terms. The attractor is a point and so has dimension zero, yet $d_L \geq 1$. Small perturbations, however, will destroy this delicate balance. For example, the one dimensional map $x_{i+1} = F(x_i) = ax_i - x_i^3$ has a fixed point at $x = 0$ with $\lambda_1 = 1$ for $a = 1$ yet $x = 0$ is attracting. This situation is changed, however, as soon as $a \neq 1$. When $|a| < 1$, $d_L = 0$, and when $|a| > 1$, $x = 0$ is no longer attracting.

[13] A C^2 diffeomorphism is a twice differentiable invertible mapping whose Jacobian has non zero determinant everywhere.

$d_L \geq d_{Led}$

The proof is a rigorous version of the heuristic argument that we have given (Fig. 6). Also, Douady and Oesterle [26] have proven that an upper bound for the fractal dimension can be obtained yielding an expression like Eq. (8), where the numbers they use are basically upper bounds for the Lyapunov numbers.

L.S. Young [11] has proven several results that strongly support conjectures 1 and 2. Particularly relevant are the following two theorems[14].

1. *If d_p exists then*

$$d_p = d_I = d_H(core) = d_{Led}. \tag{10}$$

2. *For two dimensional C^2 diffeomorphisms with $\lambda_1 > 1 > \lambda_2$, d_p exists, and*

$$d_p = \frac{h_\mu}{\log \lambda_1}\left(1 + \frac{\log \lambda_1}{\log \frac{1}{\lambda_2}}\right) \tag{11}$$

(See Sec. VII for a definition of $d_H(core)$.) h_μ denotes the Kolmogorov entropy[15] of the attractor taken with respect to the measure μ, and λ_1 and λ_2 are the Lyapunov numbers with respect to μ. (More precisely, almost every initial condition x with respect to μ give λ_1 and λ_2 as the Lyapunov numbers.)

For Axiom-A attractors Bowen and Ruelle [15] have shown that there is a natural measure such that such h_μ with respect to this measure is the sum of the one. For attractors with only one Lyapunov number greater than one, this implies that $h_\mu = \log \lambda_1$. Thus, for Axiom-A attractors of two dimensional maps, Eqs. (9–11) yield $d_\mu = d_L$. Therefore Young has shown that Conjecture 2 holds for this case. (It has been conjectured that the relationship between h_μ and the positive λ_i holds for non Axiom-A attractors that have a natural measure.) This result for the case of Axiom-A attractors of two dimensional maps has also been obtained independently by Pelikan [30].

IV. Generalized Baker's Transformation: Scaling

4.1. Definition of Generalized Baker's Transformation

In this section we define the example which we will study in detail in this and the following four sections. Although we feel that this example is general enough to be typical of low dimensional chaotic attractors (at least concerning its dimensional properties), it is also simple enough that all of the dimensions discussed in this paper can be analytically calculated[16]. Thus, for this example, we shall be able to verify Conjectures 1-3 in a case where generally $d_p \neq d_\mu$. As we shall show in Sec. V, another nice property of this map is that it allows us to investigate certain properties of the natural probablity distribution in detail.

[14] For these results Young does not require the existence of a natural measure, but rather assumes simply the existence of some invariant measure μ. In this case the Lyapunov numbers are those obtained when starting at almost every initial point with respect to μ.

[15] The Kolmogorov entropy, originally defined by Shannon [17] and applied to dynamical systems by Kolmogorov [27] and Sinai [28], puts a quantitative value on the average amount of new information obtained from a sequence of measurements. See [9] or [29] for physically motivated reviews. Note that this is also called metric entropy. The name *metric* entropy derives from the invariance properties of this quantity; in fact, the definition of metric entropy does not require a metric (but does require a measure).

[16] Except for the θ-Hausdorff dimension, for which we only obtain an upper bound.

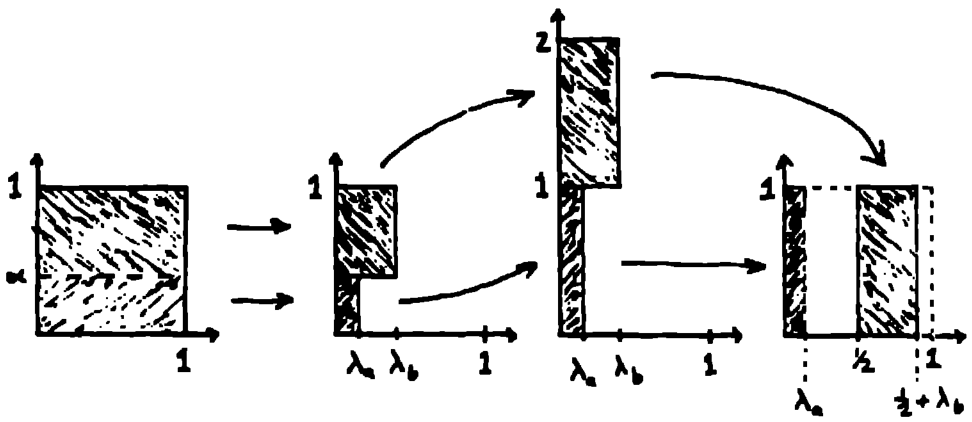

Figure 7

The Generalized Baker's Transformation.

The map to be considered is

$$x_{n+1} = \begin{cases} \lambda_a x_n & \text{if } y_n < \alpha, \\ 1/2 + \lambda_b x_n & \text{if } y_n > \alpha \end{cases} \tag{12}$$

$$y_{n+1} = \begin{cases} \dfrac{1}{\alpha} y_n, & \text{if } y_n < \alpha, \\ \dfrac{1}{1-\alpha}(y_n - \alpha) & \text{if } y_n > \alpha. \end{cases}$$

where we shall assume $0 < x_n < 1$ and $0 < y_n < 1$. If this condition is satisfied initially it is also satisfied at all subsequent iterates. Figure 7 illustrates the action of this map on the unit square. As shown in Fig. 7, we take α, λ_a, $\lambda_b < 1/2$, and $\lambda_b > \lambda_a$. Figure 8 shows the result of applying the map two times to the unit square. For Fig. 8 it is seen that, if the x interval $[0,\lambda_a]$ is magnified by a factor $\frac{1}{\lambda_a}$, it becomes a precise replica of Fig. 7(d). Similarly, if the x interval $[1/2, 1/2 + \lambda_b]$ is magnified by $\frac{1}{\lambda_b}$, a replica of Fig. 7(d) again results. This self similarity property of Eq. (12) will subsequently be used to obtain d_C, d_I, d_H, and d_L.

4.2. Lyapunov Numbers of Generalized Baker's Transformation

Now we consider the Lyapunov numbers. Equation 12(b) involves y alone and consists of a linear stretching on each of the y intervals $[0,\alpha]$ and $[\alpha,1]$. Thus almost every y initial condition in $[0,1]$ will generate an ergodic orbit in y with uniform density in $[0,1]$. The Jacobian of Eq. (12) is diagonal and depends only on y.

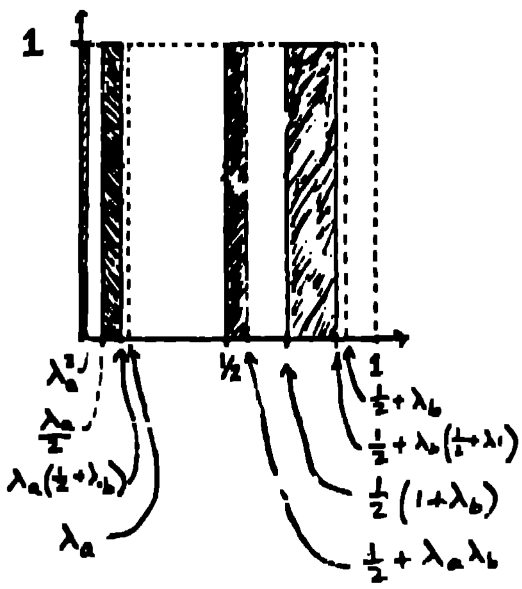

Figure 8

After applying the generalized baker's transformation to the unit square twice, only the shaded regions remain.

$$J = \begin{bmatrix} L_2(y) & 0 \\ 0 & L_1(y) \end{bmatrix}$$

where

$$L_2(y) = \begin{cases} \lambda_a & \text{if } y < \alpha \\ \lambda_b & \text{if } y > \alpha \end{cases}$$

and

$$L_1(y) = \begin{cases} \dfrac{1}{\alpha} & \text{if } y < \alpha \\ \dfrac{1}{1-\alpha} & \text{if } y > \alpha \end{cases}$$

Thus applying Eq. (6) we have

$$\lambda_1 = \lim_{n\to\infty} [L_1(y_n) \ldots L_1(y_1)]^{\frac{1}{n}}$$

or

$$\log \lambda_1 = \lim_{n\to\infty} \{ \frac{n_\alpha}{n} \log \frac{1}{\alpha} + \frac{n_\beta}{n} \log \frac{1}{\beta} \}$$

where $\beta = 1 - \alpha$, n_α is the number of times the orbit has been in the set $y < \alpha$, and n_β is the number of times the orbit has been in the set $y > \alpha$. Since for *almost* any y_1, the orbit in y is ergodic with uniform density in $[0,1]$, $\lim_{n\to\infty} \frac{n_\alpha}{n} = \alpha$, and similarly $\lim_{n\to\infty} \frac{n_\beta}{n} = \beta$. Thus

$$\log \lambda_1 = \alpha \log \frac{1}{\alpha} + \beta \log \frac{1}{\beta} \tag{13}$$

Similarly, we obtain for λ_2

$$\log \lambda_2 = \alpha \log \lambda_a + \beta \log \lambda_b . \tag{14}$$

To simplify notation in this and subsequent expressions, let

$$H(\alpha) = \alpha \log \frac{1}{\alpha} + (1-\alpha) \log \frac{1}{1-\alpha} \tag{15}$$

$H(\alpha)$ is called the ***binary entropy function*** and is the amount of information contained in a coin-toss where heads has a probability α.

The Lyapunov dimension of the attractor of the generalized baker's transformation (Eq. (12)) is

$$d_L = 1 + \frac{H(\alpha)}{\alpha \log \frac{1}{\lambda_a} + \beta \log \frac{1}{\lambda_b}} \tag{16}$$

In the following sections we compute the values of the dimensions defined in this paper, and show that all the probability dimensions take on the value given in Eq. (16).

For all but special values of λ_a, λ_b, and α, there exist unstable periodic orbits whose Lyapunov numbers are different from those given in Eqs (13) and (14).[17] Thus, in general we expect that Conjecture 2 rather than Conjecture 3 applies and $d_L \neq d_\mu$.

4.3. Capacity of Generalized Baker's Transformation

To calculate d_C we first note that the attractor is a product of a Cantor set along x and the interval [0,1] along y. Thus the capacity, or any of the other dimensions, are in the form $d_C = 1 + \bar{d}_C$, where $\bar{d}_C$ is the dimension of the attractor in the x direction. We will generally use a bar over a dimension to refer to the dimension along the x direction.

We now obtain $\bar{d}_C$ by making use of the scaling property of the generalized baker's transformation, discussed at the end of Sec. 4.1. We write $N(\varepsilon)$ as

$$N(\varepsilon) = N_a(\varepsilon) + N_b(\varepsilon),$$

where $N_a(\varepsilon)$ is the number of x intervals of length ε needed to cover that part of the attractor which lies in the x interval $[0, \lambda_a]$, and $N_b(\varepsilon)$ is the analogous quantity for the x interval $[1/2, 1/2 + \lambda_b]$. From the scaling property, $N_a(\varepsilon) = N(\frac{\varepsilon}{\lambda_a})$, and similarly $N_b(\varepsilon) = N(\frac{\varepsilon}{\lambda_b})$. Thus

$$N(\varepsilon) = N(\frac{\varepsilon}{\lambda_a}) + N(\frac{\varepsilon}{\lambda_b}) \tag{17}$$

---

[17] To see that for almost all parameter values the Lyapunov numbers of the generalized baker's transformation are not absolute, consider the special initial condition on the attractor with y value $y_1 = y_a$ where $y_a = \alpha^2(1-\alpha+\alpha^2)^{-1}$. This initial condition corresponds to one of the points on the unstable period 2 orbit, $(y_a, y_b, y_a, y_b, \ldots)$, where $y_b = \alpha^{-1} y_a$. Since $0 < y_a < \alpha < y_b < 1$, we have $\frac{n_\alpha}{n} = \frac{n_\beta}{n} = 1/2$, and the Lyapunov numbers generated by this initial condition are $\lambda_1 = (\alpha\beta)^{-1/2}$ and $\lambda_2 = (\lambda_a \lambda_b)^{1/2}$, rather than those given by Eqns. (13) and (14).

Assuming heuristically that $N(\varepsilon) \approx k\, \varepsilon^{-\bar{d}_C}$ for small ε, substituting into Eq. (17) gives

$$k\left(\frac{1}{\varepsilon}\right)^{\bar{d}_C} = k\left(\frac{\lambda_a}{\varepsilon}\right)^{\bar{d}_C} + k\left(\frac{\lambda_b}{\varepsilon}\right)^{\bar{d}_C}$$

implying that

$$1 = \lambda_a^{\bar{d}_C} + \lambda_b^{\bar{d}_C}, \tag{18}$$

which is a transcendental equation for $\bar{d}_C$. As expected, Eqns. (16) and (18) show that, in general, $1 + \bar{d}_C = d_C \neq d_L$. However, for the special choice $\lambda_a = \lambda_b$, $\alpha = ½$, corresponding to Eq. (12) with $\lambda_1 = 2$, the two agree. Note that for this case the Jacobian matrix is constant, the Lyapunov numbers are therefore absolute, and Conjecture 3 applies.

In obtaining [illegible] in order to keep the argument simple, we have made the strong assumption [illegible] $(\varepsilon) \approx k\, \varepsilon^{-\bar{d}_C}$ for small ε, which implies the existence of the limit given in the definition of capacity, Eq. (2). We can, however, show that the limit given in Eq. (2) exists and $\bar{d}_C$ must satisfy Eq. (18) in a rigorous manner, as follows:

Define $E_C(\varepsilon)$ by

$$N(\varepsilon) = E_C(\varepsilon)\, \varepsilon^{-\bar{d}},$$

where $\bar{d}$ is defined by $1 = \lambda_a^{\bar{d}} + \lambda_b^{\bar{d}}$. Substituting this into Eq. (17) then yields

$$E_C(\varepsilon) = \bar{\alpha}\, E_C\left(\frac{\varepsilon}{\lambda_a}\right) + \bar{\beta}\, E_C\left(\frac{\varepsilon}{\lambda_b}\right) \tag{19}$$

where $\bar{\alpha} = \lambda_a^{\bar{d}}$ and $\bar{\beta} = \lambda_b^{\bar{d}}$, and are independent of ε. Notice that by definition $\bar{\alpha} + \bar{\beta} = 1$, so the above expression says that $E_C(\varepsilon)$ is a weighted average of its values at $\frac{\varepsilon}{\lambda_a}$ and $\frac{\varepsilon}{\lambda_b}$. Choose ε_1 and ε_2 so that $\varepsilon_1 > \varepsilon_2 > 0$. Since $N(\varepsilon)$ and hence $E_C(\varepsilon)$ are finite and positive for any finite ε, there exist finite non-zero numbers $B_1 > B_2 > 0$ such that $B_2 < E_C(\varepsilon) < B_1$ for $\varepsilon_1 > \varepsilon > \varepsilon_2$. We can assume that ε_1 and ε_2 are chosen so that $\frac{\varepsilon_1}{\varepsilon_2}$ is large. Since $\bar{\alpha} + \bar{\beta} = 1$, Eq. (19) implies that $B_2 < E_C(\varepsilon) < B_1$ also applies to the wider interval $\varepsilon_1 > \varepsilon > \lambda_b \varepsilon_2$. Repeating this argument increases the domain of validity of the bound to $\varepsilon_1 > \varepsilon > \lambda_b^2 \varepsilon_2$, and so on. Hence $E_C(\varepsilon)$ is bounded uniformly from above and below for arbitrarily small ε. Thus the limit of Eq. (2) exists and $\bar{d}_C = \bar{d}$. (In fact it can be shown that Eq. (19) implies that $\lim_{\varepsilon \to 0} E_C(\varepsilon)$ is a constant if $\frac{\log \lambda_a}{\log \lambda_b}$ is an irrational number.) Note that in Eq. (18), since both terms on the right hand side are monotonically decreasing, $\bar{d}_C$ obtained from solving this equation is unique.

4.4. Computation of Hausdorff Dimension

The Hausdorff dimension d_H can be calculated by an argument that is very similar to the one used above in computing the capacity. Let $\bar{d}_H = d_H - 1$, the Hausdorff dimension along x. Applying the scaling property of the map to the quantity $l_d(\varepsilon)$ (defined in Sec. 2), we obtain

$$l_d(\varepsilon) = (\lambda_a)^d\, l_d\left(\frac{\varepsilon}{\lambda_a}\right) + (\lambda_b)^d\, l_d\left(\frac{\varepsilon}{\lambda_b}\right).$$

Substituting $l_d(\varepsilon) = E_H(\varepsilon)\, \varepsilon^{(\bar{d} - d)}$ into the above equation, we again find that $E_H(\varepsilon)$ satisfies Eq. (19). Thus the limit $\varepsilon \to 0$ yields $l_d = \infty$ or $l_d = 0$ for $d < \bar{d}_C$ or $d > \bar{d}_C$, respectively. Hence, as predicted in Sec. II, the Hausdorff dimension and capacity are equal, $d_H = d_C$.

4.5. Calculation of Information Dimension

The information dimension d_I can also be calculated by a scaling argument similar to that used above in computing the capacity. Once again, let $d_I = 1 + \tilde{d}_I$ and express the summation for $I(\varepsilon)$ in Eq. (4) as the sum of contributions from the two strips in Fig. 7(d),

$$I(\varepsilon) = I_a(\varepsilon) + I_b(\varepsilon). \tag{20}$$

The total probability contained in strip $[0, \lambda_a]$ is α, and that in strip $[\frac{1}{2}, \lambda_b + \frac{1}{2}]$ is β. Assuming that it takes $N(\varepsilon)$ strips of width ε to cover the whole attractor, then from the scaling property of Eq. (12), covering the strip $[0, \lambda_a]$ at resolution $\varepsilon\lambda_a$ also requires $N(\varepsilon)$ strips. Thus

$$I_a(\varepsilon\lambda_a) = \sum_{i=1}^{N(\varepsilon)} \alpha P_i \log \frac{1}{\alpha P_i}$$

$$= \alpha \left[\log \frac{1}{\alpha} + I(\varepsilon)\right].$$

Hence, replacing $\varepsilon\lambda_a$ by ε in the above,

$$I_a(\varepsilon) = \alpha \log \frac{1}{\alpha} + \alpha\, I(\frac{\varepsilon}{\lambda_a})$$

$$I_b(\varepsilon) = \beta \log \frac{1}{\beta} + \beta\, I(\frac{\varepsilon}{\lambda_b}).$$

Thus

$$I(\varepsilon) = \alpha\, I(\frac{\varepsilon}{\lambda_a}) + \beta\, I(\frac{\varepsilon}{\lambda_b}) + H(\alpha), \tag{21}$$

where $H(\alpha)$ is given by Eq. (15). Motivated by Eq. (3), if we assume that $I(\varepsilon) = \tilde{d}_I \log \frac{1}{\varepsilon}$ for small ε, and substitute for $I(\varepsilon)$, $I(\frac{\varepsilon}{\lambda_a})$, and $I(\frac{\varepsilon}{\lambda_b})$ in the above equation we obtain

$$\tilde{d}_I = \frac{H(\alpha)}{\alpha \log \frac{1}{\lambda_a} + \beta \log \frac{1}{\lambda_b}},$$

which is in turn equal to $\tilde{d}_L$. The assumption that $I(\varepsilon) = \tilde{d}_I \log \frac{1}{\varepsilon}$ can be made rigorous in the limit as $\varepsilon \to \infty$ using an argument that is completely analogous to that used in deriving the capacity in the last part of Sec. 4.3.

We should mention that Alexander and Yorke [10] have computed the Lyapunov and information dimensions of the generalized baker's transformation for the special case $\alpha = \frac{1}{2}$, $\lambda = \lambda_a = \lambda_b$, where $\lambda > \frac{1}{2}$. In this case $d_L = 2$. For uncountably many values of λ they find that also $d_I = 2$, although there are certain special values of λ for which $d_I < 2$.

In order to calculate the other probability dimensions listed in Table 1, more information concerning the probability distribution is required. This is dealt with in Sec. V, and we therefore defer calculation of of the remaining dimensions to the sections following Sec. V.

V. Distribution of Probability

In this section we derive the form of the probability distribution $\{P_i(\varepsilon)\}$ associated with the natural measure μ of the generalized baker's transformation. Here P_i denotes the probability of the i^{th} cube C_i of edge ε, i.e., $P_i = \mu(C_i)$. The collection of numbers $\{P_i(\varepsilon)\}$ may be also be thought of as the result of coarse graining the natural measure. This probability distribution is interesting both for its own sake, and because it is needed to compute some of the dimensions that we are interested in. In what follows we restrict ourselves to the case in which $\lambda_a = \lambda_b \equiv \lambda_2$, which keeps the width of all the strips the same. Thus a particularly convenient partition for computing $\{P_i\}$ is the set of 2^n nonempty strips obtained by iterating the unit square n times.

Starting with a uniform probability distribution, on one application of the map two strips are produced, one with total probability α and the other with total probability β. (See Fig. 7(d).) If the map is applied again (Fig. 8), there results one strip of probability α^2, one of probability β^2, and two of probability $\alpha\,\beta$. In general, after n applications of the map, there result 2^n strips of width $(\lambda_2)^n$ and probabilities $\alpha^m\,\beta^{(n-m)}$, $m = 0,1,2,\ldots n$. The number of strips with probability $\alpha^m \beta^{(n-m)}$ is

$$Z(n,m) = \frac{n!}{(n-m)!\,m!}, \tag{22}$$

i.e., the binomial coefficient. Since we take $\alpha < ½ < \beta$, lower m corresponds to more probable strips, i.e. strips of greater natural measure. The total probability contained in these Z(n,m) strips is

$$W(n,m) = \alpha^m\,\beta^{(n-m)}\,Z(n,m). \tag{23}$$

Note the similarity to a sequence of coin tosses; Using a coin with probability α of heads and β of tails, for a sequence of n flips the total number of sequences with m occurences of heads is given by Eq. (22), and the likelihood of all such sequences is given by Eq. (23).

For large n (small ε) it is convenient to have smooth estimates for Z(n,m) and W(n,m). Using Sterling's approximation, i.e.,

$$\log n! = (n+½)\log(n+1) - (n+1) + \log(2\pi)^{½} + O(n^{-1}),$$

we obtain from Eq. (22)

$$\log Z \approx (n+½)\log(m+1) - \log(2\pi)^{½} + 1.$$

Expanding this expression in a Taylor series about its maximum value, $m = \frac{n}{2}$, yields

$$Z(n,m) \approx \frac{2^n}{\sqrt{2\pi}}\sqrt{\frac{4}{n}}\;e^{-½[\frac{4}{n}(m-\frac{n}{2})^2]}. \tag{24}$$

Similarly, from Eq. (23), W(n,m) is

$$W(n,m) \approx \frac{1}{\sqrt{2\pi n\alpha\beta}}\;e^{-\frac{(m-n\alpha)^2}{2n\alpha\beta}} \tag{25}$$

Note that, since these expressions were obtained by Taylor series expansion, Eq (24) is only valid for $|\frac{m}{n} - ½| \ll 1$, and Eq. (25) is only valid for $|\frac{m}{n} - \alpha| \ll 1$. However, since the width of these Gaussians is $O(\frac{1}{n^{½}})$, Eq. (24) is valid for most of the strips, and Eq. (25) is valid for most of the probability.

Fig. 9 shows a schematic plot of Z and W. It is clear from this figure that, for large n, almost all of the probability is contained in a very small fraction of the total number of strips. Furthermore, this situation is accentuated as ε gets smaller (n gets larger), since the width of the Gaussians given in Eqs. (24) and (25) decreases according to $n^{½}$. In the limit as $\varepsilon \to 0$ these Gaussians approach delta functions, and they do not overlap. We feel that the above properties are typical features of chaotic attractors.

5.1. Log-Normal Distribution of Probabilities

It is instructive to rewrite Eq. (25) in another form. Let $p = \alpha^m \beta^{(n-m)}$ denote the probability of a strip, and reexpress Eq. (25) in terms of $u = \log \frac{1}{p}$ rather than m. m is proportional to u, and W(n,m) becomes

$$F(u) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u-u_0)^2}{2\sigma^2}} \, . \tag{26}$$

where

$$\sigma^2 = \frac{[\alpha\beta \log \frac{\beta}{\alpha} \log \frac{1}{\varepsilon}]}{\log \frac{1}{\lambda_2}}$$

and

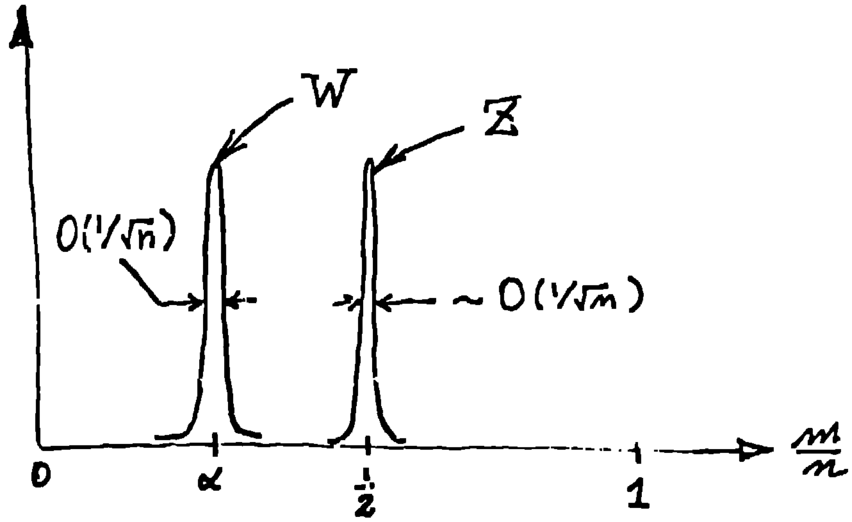

Figure 9

A schematic representation of the distribution of probabilities on the attractor. Z(n,m) is the number of cubes with probability $p = \alpha^m \beta^{(n-m)}$, and W(n,m) is the sum of the probability contained in cubes of probability p. For large n and $\frac{m}{n}$ close to its mean value, these are both approximately Gaussian distributions in $\frac{m}{n}$ whose width is proportional to n. In the limit as $n \to \infty$, W and Z become delta functions, and no longer overlap.

$$u_0 = d_L \log \frac{1}{\varepsilon} \tag{27}$$

with d_L given by Eq. (16). Eq. (26) is only valid if

$$\frac{(u-u_0)^2}{\sigma^2} \ll \log \frac{1}{\varepsilon}, \tag{28}$$

corresponding to $|\frac{m}{n} - \alpha| \ll 1$. F(u)du is the total probability contained in strips whose values of $u = \log \frac{1}{p}$ fall between u and $u + du$. Thus we see that the logarithm of p has a Gaussian distribution, or in other words, p has a log-normal distribution. We believe that this is typically true of chaotic attractors. In particular, we offer the following conjecture[18].

> *Conjecture 4:* Let A be a chaotic attractor of a p dimensional *invertible* dynamical system, and assume that this attractor has a natural measure μ. Cover A with a fixed grid of p dimensional cubes of side length ε. Assign each nonempty cube C_i probability $P_i = \mu(C_i)$, and let $U_i = \log \frac{1}{P_i}$. Let u_0 be the me .. of the numbers U_i, and let σ^2 be the variance. For typical chaotic attractors, in the limit as $\varepsilon \to 0$ values of U_i sufficiently close to the mean (in the sense of Eq. (28)) approach a Gaussian distribution. In other words, the corresponding values of P_i approach a log-normal distribution.

Note that U_i is the information obtained in a measurement that finds the orbit inside of the i^{th} cube [1,8,9]. Thus, Conjecture 4 states that for chaotic attractors the information is approximately normally distributed for small ε.

The function Z(n,m) given in Eq (24), can also be reexpressed in terms of p rather than m. When this is done, with similar restrictions to those of Eq (26), the result is also a Gaussian in terms of $u = \log \frac{1}{p}$. When recast in the more general setting of Conjecture 4, this says that the number of cubes C_i whose values U_i lie between u and $u + du$ are given by a Gaussian distribution. (Similar restrictions to those given in Conjecture 4 apply.)

VI. Computation of Probabilistic Dimensions

In this section we verify Conjectures 1 and 2 for the generalized baker's transformation by explicitly computing all of the probability dimensions defined in Sec. II. In order to simplify the computations, for all but the ϑ Hausdorff dimension we restrict ourselves to the case in which $\lambda_a = \lambda_b$ λ_a. For the ϑ-Hausdorff dimension we treat the most general case in which $\lambda_a \neq \lambda_b$, out are only able to obtain an upper bound for the dimension.

6.1. Alternate Derivation of Information Dimension

Now that we know the probability distribution for the generalized baker's transformation for $\lambda_a = \lambda_b$ λ_a, we can obtain the information dimension directly from its definition. From Eq. (4) and Eq. (26) $I(\varepsilon)$ is the average value of $\log \frac{1}{P_i}$ or

[18] The form of this conjecture was developed in collaboration with Erica Jen.

$I(\varepsilon) = \int u\, F(u)\, du = u_o$.

Since from Eq. (27) $u_o = d_L \log \frac{1}{\varepsilon}$, Eq. (4) yields $d_I = d_L$ (previously shown in Sec. IV for the more general case $\lambda_a \neq \lambda_b$). Thus the mean value of the log normal distribution is simply the information contained in the probability distribution, and its scaling rate is $I(\varepsilon) \approx d_\mu \log \frac{1}{\varepsilon}$ dimension of the natural measure.

6.2. Determination of ϑ-capacity

Here we calculate $d_C(\vartheta)$ for $\lambda_a = \lambda_b = \lambda_2$. We choose ε equal to the width of a strip, $\varepsilon = \lambda_2^n$. As usual, for convenience we compute the ϑ-capacity of the attractor projected onto the x axis, i.e. $\hat{d}_C(\vartheta) = d_C(\vartheta) - 1$. The ϑ-capacity $\hat{d}_C(\vartheta)$ is defined in terms of the minimum number of intervals $N(\varepsilon;\vartheta)$ of width ε that have total natural measure at least ϑ.

$$N(\varepsilon;\vartheta) = \sum_{m=0}^{m_\vartheta} Z(n,m), \tag{29}$$

where m_ϑ is the largest integer such that

$$\sum_{m=0}^{m_\vartheta - 1} W(n,m) < \vartheta, \tag{30}$$

To find m_ϑ we use Eq. (24) and approximate the sum in Eq. (29) by an integral,

$$\vartheta \approx \frac{1}{\sqrt{2\pi\alpha\beta n}} \int_{-\infty}^{m_\vartheta} e^{-\frac{(m-n\alpha)^2}{2n\alpha\beta}}\, dm.$$

Thus for fixed ϑ we obtain

$$\frac{m_\vartheta}{n} \approx \alpha + erfc^{-1}(\vartheta) \sqrt{\frac{\alpha\beta}{n}}. \tag{31}$$

where $erfc(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}}\, dx$. Now, consider Eq. (29). The principal contribution to the sum will come from m values very close to m_ϑ, as depicted in Fig. 10. Thus we use Eqns. (24) and (25) to approximate $Z(n,m)$ as

$$Z(n,m) \approx \frac{\beta^{-n} (\frac{\beta}{\alpha})^m}{\sqrt{2\pi n\alpha\beta}}\, e^{-\frac{(m-n\alpha)^2}{2n\alpha\beta}}. \tag{32}$$

The term $(\frac{\beta}{\alpha})^m$ decreases as m decreases away from m_ϑ, and this decrease is very rapid compared to the variation of $e^{-\frac{(m-n\alpha)^2}{2n\alpha\beta}}$. Thus in performing the sum in Eq. (29), we may approximate $e^{-\frac{(m-n\alpha)^2}{2n\alpha\beta}}$ as being constant and equal to its value at $m = m_\vartheta$. Hence the only m dependent term in the sum is $(\frac{\beta}{\alpha})^m$. Since

$$\sum_{m=0}^{m_\vartheta} (\frac{\beta}{\alpha})^m \approx (\frac{\beta}{\alpha})^{m_\vartheta} \frac{\beta}{(\beta-\alpha)},$$

we find that

$$N(\varepsilon;\vartheta) \sim \beta^{-(n-m_\vartheta)} \alpha^{-m_\vartheta} n^{-\frac{1}{2}}.$$

From Eq. (4) and $n = \frac{\log \frac{1}{\varepsilon}}{\log \frac{1}{\lambda_2}}$, the above estimate of $N(\varepsilon;\vartheta)$ yields $\hat{d}_c(\vartheta) = \hat{d}_I$, in

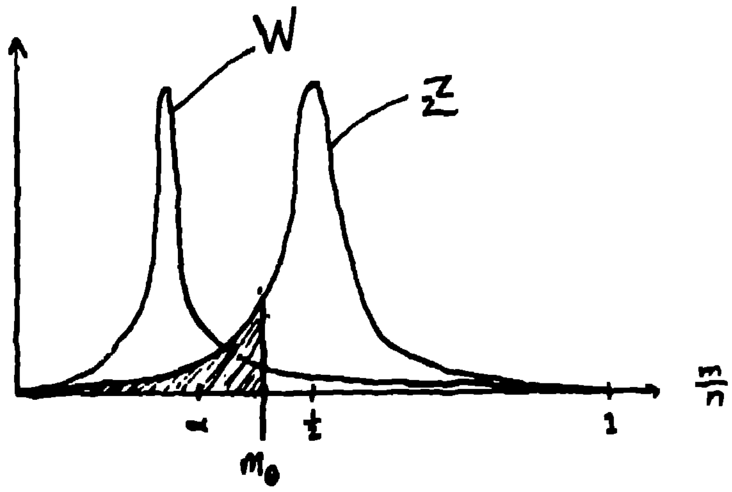

Figure 10

The principal contribution to the sum needed to compute $N(\varepsilon;\vartheta)$ (Eq. (29)) comes from values of m near m_ϑ.

agreement with Conjecture 3.

6.3. Computation of ϑ-Hausdorff Dimension

In this section we obtain an upper bound on the ϑ-Hausdorff dimension of the generalized Baker's transformation with $\lambda_a \neq \lambda_b$. (Recall that for our work in the previous section we took $\lambda_a = \lambda_b$.) We obtain an inequality for the ϑ-Hausdorff dimension by using a specific covering along x to compute the sum

$$l_d^*(\varepsilon,\vartheta) = \sum_i{}' \varepsilon_i^d ,$$

where the $\varepsilon_i < \varepsilon$ cover a fraction ϑ of the natural measure of the attractor. Our choice for the ε_i is specified below. Taking the limit as $\varepsilon \to 0$, we find that there is a value of d at which $l^*(\varepsilon,\vartheta)$ crosses over from ∞ to 0. For the partition we have chosen, we find that crossover occurs at $d = d_L$. We believe that the value we obtain is in fact the true ϑ-Hausdorff dimension. However, we cannot be sure that the particular covering we have chosen gives the lowest possible value of d, and thus we can only say that we have obtained an upper limit.

After n iterations of the map, an initially uniform probability distribution in the unit square is transformed to 2^n strips with widths $\lambda_a^m \lambda_b^{(n-m)}$ and probabilities $\alpha^m \beta^{(n-m)}$, $m = 0,1,2,\ldots,n$. As shown in Eq. (22), the number of such strips is $Z(n,m)$. We shall choose the ε_i to cover the most probable strips so that ϑ of the total probability is covered. ε for our covering is equal to the width of the widest strip, which is either $(\lambda_a)^n$ or $(\lambda_b)^n$, whichever is larger. Letting $U_d(n,m)$ be

$$U_d(n,m) = (\lambda_b^{n-m} \lambda_a^m)^d Z(n,m),$$

we have that

$$L_d^*(\varepsilon,\vartheta) = \sum_i \varepsilon_i^d = \sum_m U_d(n,m). \tag{33}$$

We still have yet to specify which m values are to be included in the sum. To do this, we expand $U_d(n,m)$ about its maximum value (as done for Z and W in Sec. V), and obtain

$$U_d(n,m) \approx \frac{[\lambda_a^d+\lambda_b^d]^n}{\sqrt{2\pi n\,\dfrac{\lambda_a^d\lambda_b^d}{(\lambda_a^d+\lambda_b^d)(\lambda_a^d+\lambda_b^d)}}}\exp-\tfrac{1}{2}\left\{\frac{\left[\dfrac{m}{n}-\dfrac{\lambda_a^d}{\lambda_a^d+\lambda_b^d}\right]^2}{\dfrac{\lambda_a^d\lambda_b^d}{n(\lambda_a^d+\lambda_b^d)(\lambda_a^d+\lambda_b^d)}}\right\} \tag{34}$$

In order to compute $L_d^*(\varepsilon,\vartheta)$, we must consider the natural measure as well as $U_d(n,m)$. Note that for the general case we are considering now with $\lambda_a \neq \lambda_b$, W(n,m) obtained in Eq. (25) continues to be the correct expression for the distribution of probabilities in each strip. Depending on the values of α, d, λ_a, and λ_b, W may peak at a value of m that is smaller, larger, or equal to the value of m at the peak of U_d. Comparing the location of the peaks of the Gaussians in Eq. (34) (for U) and in Eq. (25) (for W), we see that there are three cases:

Case 1: $\alpha < \dfrac{\lambda_a^d}{(\lambda_a^d+\lambda_b^d)}$,

Case 2: $\alpha > \dfrac{\lambda_a^d}{(\lambda_a^d+\lambda_b^d)}$, and

Case 3: $\alpha = \dfrac{\lambda_a^d}{(\lambda_a^d+\lambda_b^d)}$.

Cases 1 and 2 may be shown to be equivalent as follows. From the case 2 inequality and the fact that $\alpha + \beta = 1$, we obtain $\beta < \dfrac{\lambda_b^d}{(\lambda_a^d+\lambda_b^d)}$. But, if we define m' by $m = n - m'$, and change the sums over m to sums over m', then the roles of (α,λ_a) and (β,λ_b) are interchanged, and case 2 is converted to case 1. We shall not consider case 3 here; suffice it to say that it does not alter the results obtained from consideration of cases 1 and 2. Therefore it is sufficient to compute the ϑ-Hausdorff dimension for case 1.

For case 1, selecting a the best covering of intervals that contain ϑ of the total probability is easy. Since W remains valid, we get a covering that includes ϑ of the total natural measure by including intervals whose value of m is less than m_ϑ, just as we did for the computation of ϑ-capacity. Furthermore, since U_d peaks at a larger value of $\frac{m}{n}$ than W does, this selection gives the smallest value of L_d^*. The situation is analogous to the computation of ϑ-capacity, except that here the role of Z is played by U_d (cf. Fig. 10). To evaluate

$$L_d^* = \sum_{m=0}^{m_\vartheta} U_d(n,m) \tag{35}$$

we note that, as for the analogous evaluation for ϑ- capacity in the previous subsection, the principal contribution to the sum comes from m values close to m_ϑ. Thus we approximate $U_d(n,m)$ as

$$U_d(n,m) \approx \frac{\lambda_a^{dm}\lambda_b^{d(n-m)}}{\alpha^m\beta^{(n-m)}}\,W(n,m)$$

with W approximated by Eq. (25). Proceeding as in Sec. 6.1 we obtain an estimate for $L_d^*(\varepsilon,\vartheta)$

$$L_d^*(\varepsilon,\vartheta) \approx n^{1/2}\left(\frac{\beta}{\lambda_b^d}\right)^{m_\vartheta - n}\left(\frac{\alpha}{\lambda_a^d}\right)^{-m_\vartheta},$$

or

$$\log [l_d^*(\varepsilon,\vartheta)] \approx -n\,[d-(\bar{d}_L)]\log\left(\frac{1}{\lambda_2}\right),$$

for $\varepsilon \to 0$ (i.e., $n \to \infty$) we obtain $l_d^*(\vartheta) = 0$ for $d > \bar{d}_L$ and $l_d^*(\vartheta) = \infty$ for $d < \bar{d}_L$. Thus remembering that $d_H(\vartheta) = \bar{d}_H(\vartheta) + 1$,

$$d_H(\vartheta) \le d_L. \tag{36}$$

As already mentioned, we expect that the above inequality is really an equality. This expectation is reinforced by the fact that when $\vartheta = 1$ we recover the exact expression for the Hausdorff dimension computed in Eq. (18). To see that this is true, replace m_ϑ in Eq (35) by n. From the form of U_d, this sum is simply the binomial expansion of $(\lambda_a^d + \lambda_b^d)^n$. As $n \to \infty$, this quantity is 0 or ∞ for $d > \bar{d}_H$ or $d < \bar{d}_H$, where $\bar{d}_H$ satisfies $\lambda_a^{\bar{d}_H} + \lambda_b^{\bar{d}_H} = 1$, which is the same as Eq. (18). That is, for the specific choice of ε_i that we have used, we obtain the correct value of d_H. Since the same choice of the ε_i was used in obtaining $d_H(\vartheta)$, it seems plausible that the equality might apply in Eq. (36).

6.4. Computation of the Pointwise Dimension

We now consider the pointwise dimension for the generalized baker's transformation with $\lambda_a = \lambda_b < \frac{1}{2}$, and we show that d_p exists and is equal to d_L.

As previously noted in Sec. IV, application of the map n times to the unit square produces 2^n strips of widths $(\lambda_a)^n$. (Recall that we are assuming $\lambda_a = \lambda_b$.) In order to compute the pointwise dimension, we choose a point x at random with respect to the natural measure μ, comput the natural measure contained in an ε ball centered about x, i.e. $(\mu(B_\varepsilon(x))$, and compute the ratio of $\mu(B_\varepsilon(x))$ to $\log \frac{1}{\varepsilon}$ in the limit as ε goes to zero. The simplest case for this computation occurs when $\lambda_a < \frac{1}{4}$, so that the gaps between strips are bigger than the strips themselves, as pictured in Fig. 11(a). Choosing a point x at random with respect to the natural measure μ, let S_n denote the n^{th} order strip of width $(\lambda_a)^n$ that the point x lies in. Letting $\varepsilon = (\lambda_a)^n$, the natural measure contained in a ball of radius ε around x (i.e., the x interval $[x - (\lambda_a)^n, x + (\lambda_a)^n])$ will be equal to the natural measure of the strip S_n, regardless of where in the strip x lies (See Fig. 11(a).) The natural measure contained in a given strip is $\alpha^m \beta^{(n-m)}$, where $n \ge m > 0$, where m depends on the particular strip that x happens to lie in. (See Sec. V.) Thus, we have

$$\lim_{\varepsilon\to\infty} \frac{\mu(B_\varepsilon(x))}{\log \varepsilon} = \lim_{n\to\infty} \frac{\log \mu(S_n)}{n \log \lambda_a} = \lim_{n\to\infty} \frac{m \log \alpha + (n-m)\log \beta}{n \log \lambda_a} \tag{37}$$

In the limit as n grows large, as shown in Sec. V (see Fig. 9), the total probability W(n,m) contained in strips of a given m value is a approximately Gaussian centered about $\frac{m}{n} = \alpha$. Thus, in the limit as $n \to \infty$ it becomes overwhelmingly likely that $\frac{m}{n} = \alpha$. Thus for almost every x with respect to the natural measure μ, $\lim_{n\to\infty} \frac{m}{n} = \alpha$. (This is just a statement of the law of large numbers.) Putting this into Eq (37) gives

$$\lim_{n\to\infty} \frac{\mu(B_\varepsilon(x))}{\log \varepsilon} = \frac{\alpha \log \alpha + \beta \log \beta}{\log \lambda_a} = \frac{H(\alpha)}{\log \frac{1}{\lambda_a}} = d_l \tag{38}$$

(See Eqs (15) and (16).)

To extend this computation of the pointwise dimension to the case that $\frac{1}{2} > \lambda_a > \frac{1}{4}$, for any $\lambda_a < \frac{1}{2}$ choose a k such that $\lambda_a^{k+1} \le \frac{1}{2} - \lambda_a$ (e.g., for $\lambda_a \le \frac{1}{4}$ this relation is satisfied for any $k \ge 0$, for $\lambda_a \le 0.305...$, for any $k \ge 1$; etc.). Then we can show $\mu(B_\varepsilon(x)) \sim \alpha^{k}\mu(S_n)$,

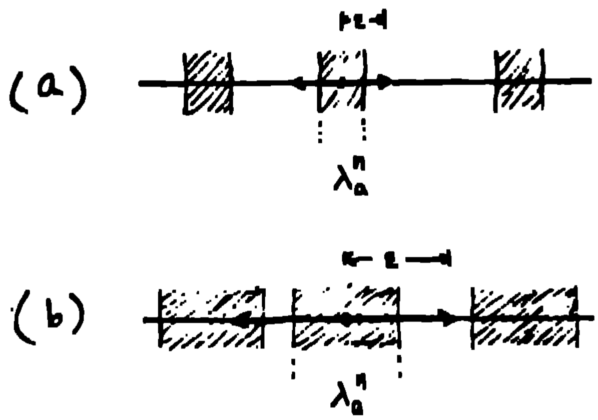

Figure 11

(a) Computing the pointwise dimension for the case that $\lambda_a < \frac{1}{4}$. (b) The case $\lambda_a > \frac{1}{4}$, in which the computation is a little more complicated.

where, without loss of generality, we have assumed $\alpha \leq \beta$. Since $B_\epsilon(x) \supset S_n$, we have also $\mu(B_\epsilon(x)) \geq \mu(S_n)$. Thus $\mu(S_n) \leq \mu(B_\epsilon(x)) \leq \alpha^{-k}\mu(S_n)$ which with Eq (37) yields $d_p = d_\mu$. (Our evaluation of Eq (37) holds not for $\epsilon \to 0$ but rather holds for the restricted set of $\epsilon = \lambda_a^n$, $n = 1,2,\ldots$; however, it is not hard to show that that in fact implies Eq (38) for every sequence of ϵ going to 0.)

Thus we have shown that for the generalized baker's transformation the pointwise dimension is equal to the dimension of the natural measure. (Although we have only shown this for $\lambda_a = \lambda_b$, it is not hard to extend this result to $\lambda_a \neq \lambda_b$.)

VII. The Core of Attractors

As shown in Sec V, for the generalized baker's transformation, typically almost all of the probability is contained in a very small fraction of the total number of cubes needed to cover the attractor. In the limit as ϵ goes to zero, this fraction goes to zero. Thus, the natural measure of the attractor is *concentrated* on a subset of the attractor. We will call this subset the *core* of the attractor.

To get a better feel for why this comes about, and to see how the properties of the core are related to those of the attractor and its natural measure, consider the special case of the generalized baker's transformation where $\lambda_a = \lambda_b = \frac{1}{2}$. As we have already seen, at the n^{th} level of approximation the natural measure consists of 2^n vertical strips of probability $\alpha^m \beta^{n-m}$. For large n and $\beta > \alpha$, strips with small values of m will contain much more of the natural measure than strips with large values of m. Fig. 12 shows a plot of the n^{th} level approximation to the probability distribution as a function of x with $\lambda_a = \lambda_b = \frac{1}{2}$, and $\alpha < \frac{1}{2}$ and $n = 10$. The probability distribution looks as though it were made up of spikes, showing that already at $n = 10$ the natural measure has become quite concentrated in certain cubes (in

Figure 12

The natural probability distribution of the generalized baker's transformation projected onto the x axis, and coarse grained using intervals of width $\epsilon = 2^{-10}$. In this case $\alpha = \frac{1}{3}$, and $\lambda_a = \lambda_b = \frac{1}{2}$.

this case intervals).

To understand the form of this probability distribution, it is instructive to represent the probability distribution of these strips in terms of x rather than m. To do this, approximate x using its first n binary digits, i.e. as a binary decimal truncated after n digits. Let m be the number of ones contained in the first n digits of the binary expansion of x. The natural measure of the strip $S_n(x)$ containing x is then $\mu(S_n(x)) = \alpha^m \beta^{(n-m)}$. (See the discussion at the beginning of Sec. V.) As we have already shown (see Fig. 9), when written in terms of m, for large n the natural measure is approximately a Gaussian centered about αn, and in the limit where n is large almost all the measure is contained in strips with $m = \alpha n$. In other words, the natural measure of the generalized baker's transformation for $\lambda_a = \lambda_b = \frac{1}{2}$ is concentrated on those values of x that have 1's in their binary expansions in the fraction α, or equivalently, 0's in the fraction β. In the limit $n \to \infty$, *all* the natural measure is contained in this set, which we will call the core of the attractor.

For this case ($\lambda_a = \lambda_b = \frac{1}{2}$) the attractor is the entire unit square. The core of this attractor is dense on the attractor. In other words, any point of the attractor has points of the core arbitrarily close to it. Hence any covering of the core must also be a covering of the attractor, and visa versa. Thus the capacity of the core is the same as that of the attractor. The Hausdorff dimension, in contrast, is more subtle, and in fact, computing the Hausdorff dimension of the set of numbers whose binary expansions have a given fraction of ones is a classic problem in the study of Hausdorff dimension [31]. The Hausdorff dimension of this set is

$$d_H = \frac{H(\alpha)}{\log 2}$$

(See Eq. (15).) This result was conjectured by Good in 1941 [32] and proved by

Eggleston in 1949 [33]. Also, the Hausdorff dimension of a very similar example (involving ternary rather than binary expansions) was proven by Besicovitch in 1931 [34].

Thus, for this example we see that the Hausdorff dimension of the core is equal to the dimension of the natural measure, and the capacity of the core is equal to the fractal dimension of the attractor. For the case of diffeomorphisms of the plane, the former result has been proven by Young [11]. We suspect that this is a property of typical attractors.

VIII. An Attractor that is a Nowhere Differentiable Torus

This section contains a review of the work of Kaplan, Mallet-Paret, and Yorke [35] on the dimension of a chaotic attractor in a setting that is quite different from that of the generalized baker's transformation. The attractor described below has the same topological form as a torus, and yet is nowhere differentiable, thus providing an interesting example of the nonanalytic forms that can be produced by chaotic dynamics.

Consider the following map:

$$\begin{aligned} x_{n+1} &= 2x_n + y_n && \mathit{mod}\ 1 \\ y_{n+1} &= x_n + y_n && \mathit{mod}\ 1 \\ z_{n+1} &= \lambda z_n + p(x_n, y_n) \end{aligned} \tag{39}$$

where x and y are taken mod 1, z can be any real number, and p is periodic in x and y with period 1 and is at least five times differentiable. (For example, $p(x,y) = \cos 2\pi x$.) In order to keep z bounded, λ is chosen between 0 and 1. Note that the eigenvalues and eigenvectors of the Jacobian matrix of Eq (39) are independent of x, y, and z. Thus *every* initial condition has the same Lyapunov numbers, i.e., the Lyapunov numbers are absolute, so that in this case Conjecture 3 is relevant, and we expect that the fractal dimension and the dimension of the natural measure should be equal.

The equations for x and y are independent of z, and in fact are the classic Anasov or "cat" map [36],

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = A \begin{pmatrix} x_n \\ y_n \end{pmatrix} \quad \mathit{mod}\ 1$$

where

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

Thus, the x-y dynamics are chaotic, and are unaffected by the value of z.

To understand the shape of the attractor in the z direction, put a sample initial condition into Eq (39). For example, consider $(x_0, y_0, 0)$. z_n takes on the form

$$z_n = \sum_{k=1}^{n} \lambda^{k-1} p(x_{n-k}, y_{n-k})$$

Making use of the fact that $\begin{pmatrix} x_{n-k} \\ y_{n-k} \end{pmatrix} = A^{-k} \begin{pmatrix} x_n \\ y_n \end{pmatrix}$, and letting n go to infinity, it can be shown that the surface given by

$$z(x,y) = \sum_{k=1}^{\infty} \lambda^{k-1} p\left(A^{-k} \begin{pmatrix} x \\ y \end{pmatrix}\right)$$

is invariant and is the unique attractor of this dynamical system.

$z(x,y)$ has some very interesting properties. For $\lambda < \frac{1}{R}$, where $R = \frac{3+\sqrt{5}}{2}$, $z(x,y)$ is smooth and has dimension 2. If $\lambda > \frac{1}{R}$, however, for most choices of p, $z(x,y)$ is nowhere differentiable. A typical cross section of $z(x,y)$ is shown in Fig. 13. To understand intuitively how the nondifferentiability of $z(x,y)$ comes about, notice that $z(x,y)$ is the sum of an infinite number of periodic functions whose arguments are the successive iterates of the cat map. Unless λ is small enough to diminish the effect of higher order iterates, the value of the sum can swing wildly as x or y change.

The Lyapunov numbers of the map given in Eq. (39) are $\lambda_1 = R$, $\lambda_2 = \lambda$, and $\lambda_3 = \frac{1}{R}$, where $R = \frac{3+\sqrt{5}}{2}$, as given above. Kaplan, Mallet-Paret, and Yorke [35] have shown that there are two possibilities for the dimension of $z(x,y)$

Either

(i) $z(x,y)$ is nowhere differentiable and

$$d_C = d_L$$

or

(ii) $z(x,y)$ is differentiable and $d_C = 2$.

For given p, the nowhere differentiable case occurs for nearly every choice of λ. Thus we see that Conjecture 3 is satisfied for this example.

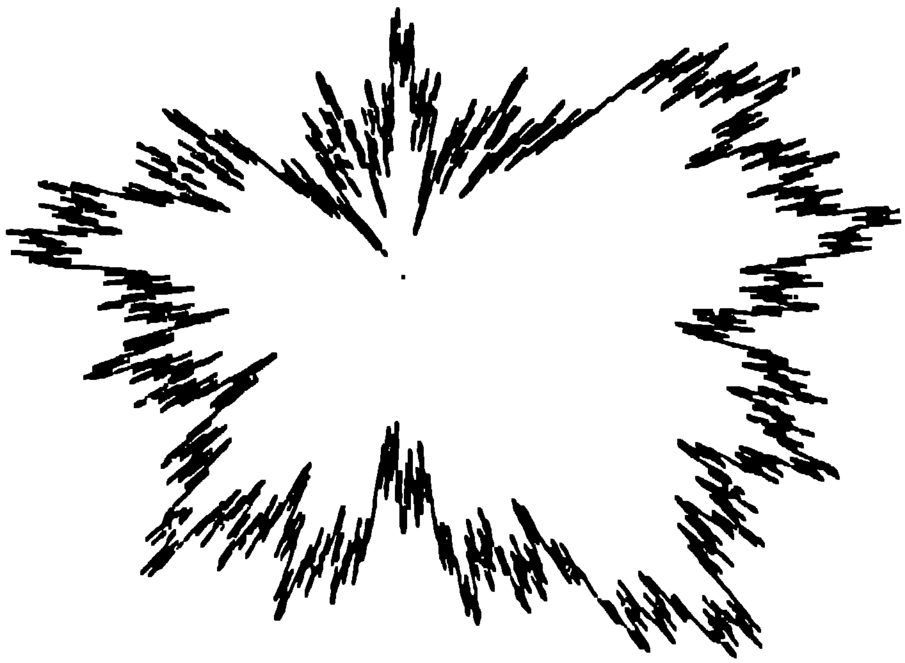

Figure 13

A cross section of a nowhere differentiable torus, made using Eq. (39) with λ chosen so that $d_C = 2\frac{1}{2}$.

IX. Numerical Computations

In this section we discuss some aspects of the numerical computation of dimension. First we will discuss the basic ideas behind numerical computations of dimension, secondly we will discuss some of the problems encountered in such computations, and finally we will review some previous numerical work.

The methods to compute dimension vary considerably depending on the dimension that one wishes to compute. Thus far, we are aware of numerical computations only of capacity [37-41], Lyapunov dimension [37-40], and Hausdorff dimension [42]. Of these, only the studies involving the capacity and the Lyapunov dimension were applied to attractors of dynamical systems. In each case, the computations follow from the definitions. As we shall see, the capacity is (in principle) straightforward to compute, but is in practice unfeasible to compute for all but very low dimensional attractors. The Lyapunov dimension, in contrast, is much more feasible to compute. We will begin the discussion with a description of the computation of Lyapunov dimension, and then go on to discuss the computation of capacity.

9.1. Numerical Computation of Lyapunov Dimension

The Lyapunov dimension is defined in terms of the Lyapunov numbers. (See Sec. III.) Thus, the work involved in computing Lyapunov dimension is in computing the Lyapunov numbers. Numerical methods for doing this have been discussed by Bennetin et al. [43], Shimada and Nagashima [44], and in infinite dimensions by Farmer [38]. With appropriate numerical caution, the largest k Lyapunov numbers can be computed by following the evolution of k nearby trajectories simultaneously and measuring their rate of separation. There are various numerical problems with this method, however, and a better method is to follow only one trajectory, but also follow k trajectories of the associated equations for the evolution of vectors in the tangent space. These methods have been successfully used in a variety of numerical studies.

For low dimensional cases, such as two dimensional maps or systems of three autonomous ordinary differential equations, with a modern computer and plenty of computer time, numerical computation of the dimensions we discuss here directly from their definitions is feasible, as discussed in the next subsection. Even in such low dimensional cases, however, the computation of Lyapunov dimension is by far less costly in terms of computer time and memory than the computation of other dimensions. For higher dimensional attractors it appears that only the Lyapunov dimension is computationally feasible. The key reason that the Lyapunov dimension is feasible to compute numerically even for attractors of rather high dimension (e.g. $d_L \approx 10$) is that the difficulty of the computation scales linearly with the dimension of the attractor times the dimension of the space it lies in, rather than exponentially as it does for a computation of the fractal dimension, or any of the other dimensions discussed in this paper. The memory needed to compute the largest j Lyapunov numbers is equal to the memory needed to numerically integrate the equations under study, multiplied by j+1. (Memory requirements are usually a problem only in computations involving partial differential equations.) The computer time needed is the time needed to compute a time average to the desired accuracy (which depends, among other things, on the irregularity of the natural measure of the attractor), multiplied by j+1. Fortunately it is only necessary to compute the largest Lyapunov numbers, and the number of these needed depends on the dimension of the attractor rather than the dimension of the phase space. (See Eq. (9).) This linear dependence on the dimension of the attractor has allowed computation of the Lyapunov dimension for attractors of dimension as large as twenty [38].

We should mention one disadvantage concerning Lyapunov dimension. Namely, it is not presently known how the Lyapunov dimension can be determined directly from a physical experiment. The difficulty comes about because, in some sense, in order to determine Lyapunov numbers it is necessary to be able to follow adjacent trajectories. To determine all the necessary Lyapunov numbers, it is necessary to follow some trajectories (at least one) which are not on the attractor. Thus it is not possible to compute the Lyapunov dimension by simply observing behavior on the attractor; one must perturb the system from the attractor, and do so in a very well defined way. This poses a very severe problem in the computation of dimension from experimental data, one that is not present in the computation of other dimensions.

9.2. Computation of Fractal Dimension

In principle, it is quite straightforward to use the definition of capacity, Eq. (2), to compute the fractal dimension. The region of phase space surrounding the attractor is divided up into a grid of cubes of size ε, the equations are iterated, and the number of cubes $N(\varepsilon)$ that contain part of the attractor are counted. ε is decreased and the process is repeated. If $\log N(\varepsilon)$ is plotted against $\log \varepsilon$, in the limit as ε goes to zero the slope is the fractal dimension.

The difficulty with this method is that one must use values of ε small enough to insure that the asymptotic scaling has been reached. (See Froehling et al. [40] and Greenside et al [39].) The total number of cubes containing part of the attractor scales roughly as

$$N(\varepsilon) \sim \left(\frac{1}{\varepsilon}\right)^{d_C} \tag{40}$$

Thus, the number of cubes increases *exponentially* with the fractal dimension of the attractor. To get a feel for the seriousness of this problem, plug in some typical numbers. If $\varepsilon = .01$ and $d_C = 3$, then $N \approx 10^6$, exceeding the core memory of all but the biggest current computers. Thus, computations of fractal dimension are currently not feasible for attractors of dimension significantly greater than three.

In addition, there is another potential problem involved in computing capacity. In counting cubes, how can one be sure that all the nonempty cubes have been counted? This problem is compounded by the highly nonuniform distribution of probability on an attractor. In particular, if our hypothesis that the probability is distributed log-normally is correct, in order to count the highly improbable cubes present in the wings of the distribution requires that a large number of points on the attractor must be generated. Furthermore, this number increases rapidly as ε decreases.

The conclusion is that a great deal of care must be taken in the computation of fractal dimension, and in particular, a sufficiently large number of points on the attractor must be generated to insure that low probability cubes are not left out in the determination of $N(\varepsilon)$.

Although there are as yet no extensive results on direct computations of the dimension of the natural measure, it may be easier to reliably compute than the fractal dimension.

The reason for this is that very improbable cubes are irrelevant for a computation of the dimension of the natural measure. Numerical experiments on this topic are currently in progress.

9.3. Summary of Past Numerical Experiments

In this section we summarize previous numerical experiments on dimension computation. The two studies most relevant to the topic under discussion are those of Russel et al. [37] and Farmer [38]. Both of these were made in an attempt to test the Kaplan-Yorke conjecture [7,22]. (See Section III.) In both of these studies, the capacity of chaotic attractors was computed directly from the definition The Lyapunov dimension was also computed, and compared to the capacity.

In the study of Russel et al., five examples were examined. In each case, the computed capacity agreed with the computed Lyapunov dimension to within experimental accuracy; in several of these cases, this agreement was within three significant figures. These computations were done on the Cray!, a state of the art mainframe computer; at the smallest value of $\varepsilon = 2^{-14}$, more than 10^5 cubes were counted.

The numerical experiments of Farmer were done using high dimensional approximations to an infinite dimensional dynamical system. Because the equations under study were more time consuming to integrate, and because the capacity computations were done on a minicomputer, it was only possible to achieve about two significant figures of accuracy. The computed capacity and Lyapunov dimension agreed to this accuracy at the two parameter values tested.

In 1980, Mori [23] conjectured an alternate formula relating the fractal dimension to the spectrum of Lyapunov numbers. For attractors in a low dimensional phase space, such as those studied by Russel et al. [37], Mori's formula and the Kaplan-Yorke formula (Eq. (9)) predict the same value. For higher dimensional phase spaces, however, the two formulas no longer agree. Farmer's results support the Kaplan-Yorke formula, and conclusively show that the Mori formula is incorrect for higher dimensions.

One puzzling aspect of both of these numerical experiments is the striking agreement between the computed value of capacity and the Lyapunov dimension The Kaplan-Yorke conjecture equates the Lyapunov dimension to the dimension of the natural measure, and therefore only gives a lower bound on the fractal dimension. Why, then, was such good agreement obtained between the computed capacity and the computed Lyapunov dimension? We do not yet understand the answer to this question, though further numerical experiments may resolve the question

X. Conclusions

We have given several different definitions of dimension. These divide into two types, those that require a probability measure for their definition, and those that do not. (Refer back to Table 1.) For an example that we believe is typical of chaotic attractors, i.e., the generalized baker's transformation, our computations of dimension show that all of the probabilistic definitions take on one value, which we call the dimension of the natural measure, while the definitions that do not require a probability measure take on another value, which we call the fractal dimension of the attractor. We believe that this is true for typical attractors.

If the probability distribution on the attractor is "coarse grained" by covering the attractor with cubes, for the generalized baker's transformation we find that the probability contained in these cubes is distributed nearly log-normally when the cubes are sufficiently small. In other words, the total probability contained in cubes whose natural measure is between $u = \log p_i$ and $u + du$ has a distribution that is nearly Gaussian, and as the size of the cubes is decreased, it becomes more nearly Gaussian. Furthermore, the number of cubes in a given interval of u also has a Gaussian distribution, but with a different mean and variance. (See Fig. 9.) As ε decreases, both of these distributions become narrower in a relative sense, in that the ratio of their variance to their mean decreases. In the limit as ε goes to zero, both distributions approach delta functions; since their means are different, in this limit the two distributions typically do not overlap. Thus, almost all of the natural measure is contained in almost none of the cubes, and the natural measure is concentrated on a core set. The capacity of the core is the fractal dimension of the attractor, while the Hausdorff dimension of the core is the dimension of the natural measure. Once again, although we have demonstrated the results mentioned in this paragraph only for the generalized baker's transformation, we feel that they are true for typical chaotic attractors.

Most of the dimensions that we have defined are difficult to compute numerically. The Lyapunov dimension, however, is much easier to compute numerically than any of the other dimensions. We compute the Lyapunov dimension for the generalized baker's transformation, and show that it is equal to the dimension of the natural measure obtained from any of the other probabilistic dimensions that we have investigated. This supports the conjecture of Kaplan and Yorke.

References

1. R. Shaw, "Strange Attractors, Chaotic Behavior, and Information Flow", Z. Naturforsch. 36a (1981) 80.
2. E. Ott, "Strange Attractors and Chaotic Motions of Dynamical Systems", Rev. Mod. Phys. *53* (1981) 655.
3. R. Helleman, "Self-Generated Chaotic Behavior in Nonlinear Mechanics", Fundamental Problems in Stat. Mech. 5, Ed. E.G.D. Cohen, (North Holland, Amsterdam and New York, 1980) 165-233.
4. J.A. and E.D. Yorke, "Chaotic Behavior and Fluid Dynamics", *Hydrodynamic Instabilities and the Transition to Turbulence*, H.L. Swinney and J.P. Gollub Eds., Topics in Applied Physics *45*, Springer-Verlag (1981) 77-95.
5. Ja. Sinai, "Gibbs Measure in Ergodic Theory", Russ. Math. Surveys 4 (1972) 21-64.
6. B. Mandelbrot, *Fractals: Form, Chance, and Dimension* (Freeman, San Francisco, 1977).
7. P. Frederickson, J. Kaplan, E. Yorke, and J. Yorke, "The Lyapunov Dimension of Strange Attractors", to appear in J. Diff. Eqns.
8. J.D. Farmer, "Dimension, Fractal Measures, and Chaotic Dynamics", to appear in *Evolution of Ordered and Chaotic Patterns in Systems Treated by the Natural Sciences and Mathematics*, H. Haken, Ed., (Springer-Verlag, Berlin, 1982).
9. J.D. Farmer, "Information Dimension and the Probabilistic Structure of Chaos", to appear Z. Naturforsch. September, 1982.
10. J. Alexander and J. Yorke, "The Fat Baker's Transformation", U. of Maryland preprint (1982).
11. L.S. Young, "Dimension, Entropy, and Lyapunov Exponents", to appear in Ergodic Theory and Dynamical Systems.
12. W. Hurwicz and H. Wallman, *Dimension Theory*, (Princeton Univ. Press, Princeton, 1948).
13. A.N. Kolmogorov, "A New Invariant for Transitive Dynamical Systems", Dokl Akad. Nauk SSSR 119 (1958) 861-864.
14. Hausdorff, "Dimension und Außeres Maß", Math Annalen *79* (1918) 157
15. R. Bowen and D. Ruelle, "The Ergodic Theory of Axiom-A Flows", Inv. Math 29 (1975) 181-202.
16. J. Balatoni and A. Renyi, Publ. Math. Inst. of the Hungarian Acad. of Sci. 1 (1956) 9 (Hungarian). English translation, Selected Papers of A. Renyi, Vol. 1 (Academiai Budapest, Budapest, 1976), p. 558. See also A. Renyi, Acta Mathematica (Hungary) 10 (1959) 193.
17. C. Shannon, "A Mathematical Theory of Communication", Bell Tech. Jour. 27 (1948) 379-423, 623-656.
18. F. Ledrappier, "Some Relations Between Dimension and Lyapunov Exponents", Laboratoire Associe au CNRS LA 224 preprint.
19. B. Mandelbrot, *The Fractal Geometry of Nature*, (W.H. Freeman and Co., San Francisco, to appear in 1982.)
20. F. Takens, "Invariants Related to Dimension and Entropy", to appear in *Atas do 13 Cologuio Brasiliero de Mathematica*.

21. T. Janssen and J. Tjon, "Bifurcations of Lattice Structure", U. of Utrecht preprint.
22. J. Kaplan and J. Yorke, *Functional Differential Equations and the Approximation of Fixed Points, Proceedings, Bonn, July 1978, Lecture Notes* in Math. 730, H.O. Peitgen and H.O. Walther, Eds., (Springer, Berlin, New York), p. 228.
23. H. Mori, "Prog. Theor. Phys. 63 (1980) 3.
24. V.I. Oseledec, "A Multiplicative Ergodic Theorem. Lyapunov Characteristic Numbers for Dynamical Systems", Trans. Moscow Math. Soc. 19 (1968) 197.
25. C. Grebogi, E. Ott, and J. Yorke, "Chaotic Attractors in Crisis", in this volume.
26. A. Douady and J. Oesterle, "Dimension de Hausdorff des Attracteurs, Comptes Rendus des Seances de L'academie des Sciences, 24 (1980) 1135-38.
27. A.N. Kolmogorov, Dolk. Akad. Nauk SSSR 124 (1959) 754. English summary in MR 21, 2035.
28. Ya. G. Sinai, Dolk. Akad. Nauk SSSR 124 (1959) 768. English summary in MR 21, 2036.
29. J. Crutchfield and N. Packard, "Symbolic Dynamics of One Dimensional Maps: Entropies, Finite Precision, and Noise", Int'l. J. Theo. Phys. 21 (1982) 433.
30. S. Pelikan, private communication.
31. P. Billingsley, *Ergodic Theory and Information*, (John Wiley and Sons, New York, 1965).
32. I.J. Good, "The Fractional Dimensional Theory of Continued Fractions", Proc. Camb. Phil. Soc. 37 (1941) 199-228.
33. H.G. Eggleston, "The Fractional Dimension of a Set Defined by Decimal Properties", Quart. J. Math. Oxford Ser. 20 (1949) 31-36.
34. A. Besicovitch, "On the Sum of Digits of Real Numbers Represented in the Dyadic System", Math. Annalen. 110 (1934) 321.
35. J.L. Kaplan, J. Mallet-Paret, and J.A. Yorke, "The Lyapunov Dimension of a Nowhere Differentiable Attracting Torus", U. of Maryland Preprint (1982).
36. V.I. Arnold and Avez, *Ergodic Theory in Classical Mechanics* (New York, 1968).
37. D. Russel, J. Hansen, and E. Ott, "Dimensionality and Lyapunov Numbers of Strange Attractors", Phys. Rev. Lett. 45 (1980) 1175.
38. J.D. Farmer, "Chaotic Attractors of an Infinite Dimensional Dynamical System", Physica 4D 3 (1982) 366-393.
39 H. Greenside, A. Wolf, J. Swift, and T. Pignataro, "The Impracticality of a Box Counting Algorithm for Calculating the Dimensionality of Strange Attractors", Phys. Rev. Rapid Communications, ? (1982).
40. H. Froehling, J. Crutchfield, J.D. Farmer, N. Packard, and R. Shaw, "On Determining the Dimension of Chaotic Flows", Physica 3D (1981) 605.
41 R. Kautz, private communication.
42. A.J. Chorin, "The Evolution of a Turbulent Vortex", Comm. Math. Phys. 83 (1982) 517-535.
43. G. Bennetin, L. Galgani, and J. Strelcyn, Phys. Rev. A 14 (1976) 2338, also see G. Benettin, L. Galgani, A. Giorgilli and J. Strelcyn, Meccanica 15 (1980) 9.
44. I Shimada and T. Nagashima, Prog. Theor. Phys. 61 (1979) 228.