Information Dimension and the Probabilistic Structure of Chaos

J. Doyne Farmer

Dynamical Systems Collective, Physics Department, UC Santa Cruz, Santa Cruz, California

Z. Naturforsch. **37a**, 1304–1325 (1982); received May 18, 1982

The concepts of entropy and dimension as applied to dynamical systems are reviewed from a physical point of view. The information dimension, which measures the rate at which the information contained in a probability density scales with resolution, fills a logical gap in the classification of attractors in terms of metric entropy, fractal dimension, and topological entropy. Several examples are presented of chaotic attractors that have a self similar, geometrically scaling structure in their probability distribution; for these attractors the information dimension and fractal dimension are different. Just as the metric (Kolmogorov-Sinai) entropy places an upper bound on the information gained in a sequence of measurements, the information dimension can be used to estimate the information obtained in an isolated measurement. The metric entropy can be expressed in terms of the information dimension of a probability distribution constructed from a sequence of measurements. An algorithm is presented that allows the experimental determination of the information dimension and metric entropy.

Introduction

Let us start out with a quotation from Abbot's "Flatland" [1]:

"Dimension implies direction, implies measurement, implies the more and the less."

"As one of your Spaceland poets has said, we are all liable to the same errors, all alike the Slaves of our respective Dimensional prejudices."

A Square

There is no unique notion of dimension. In the familiar territory of Euclidean space, we may all be slaves to a different aspect of dimension and yet arrive at the same conclusions. But, if we venture into the realm of the bizarre, distinct notions of dimension diverge. In the domain of chaos, deterministic structure amplifies the uncertainty inherent in measurement, until only probabilistic information remains. To comprehend the strange objects that inhabit this world, we must expand our concept of dimension to encompass chance as well as certainty.

When dealing with a dissipative dynamical system we may begin with a Euclidean phase space of initial conditions of large or even infinite dimension; after some time passes, the transients relax, some modes may damp out, and the point in phase space that describes the state of the system approaches an attractor. The number of degrees of freedom is thereby reduced. Dimension provides the natural vehicle to make the notion of "degrees of freedom" more precise.

Following A Square [1], we will distinguish three distinct intuitive notions of dimension:

(1) Direction: Euclidean 3-space is naturally described by 3-tuples, and is not homeomorphic to Euclidean 2-space. Following Poincaré, Brouwer, Menger, and Urysohn, this notion naturally leads to the modern definition of *topological dimension* [2].

(2) Capacity (the more and the less): The volume of a cube varies as the third power of its side, whereas the "volume" of a square varies as the second. Following Hausdorff and Besicovitch, we are led to the *fractal dimension*, beautifully described in the book by Mandelbrot [3].

(3) Measurement: Here we come unstuck from pure geometry. Measurement leads to probability, and probability leads to entropy, or to take the positive aspect, information. The scaling rate as the length of resolution varies defines the *information dimension*. This little known quantity was originally defined by Balatoni and Renyi [4]. Exploring its properties and arguing for the utility of this concept are the central goals of this paper.

For simple, predictable attractors such as fixed points, limit cycles, or 2-tori, the separate notions of dimension converge; by any reasonable definition these attractors are of dimension 0, 1, or 2 respectively. Chaotic (strange) attractors, however, pose a more difficult problem. At any fixed level of precision, most information about initial condi-

Reprint requests to Center for Nonlinear Studies, MS B258 L.A.N.L., Los Alamos, New Mexico 87545, USA.

0340-4811 / 82 / 1100-1304 $ 01.30/0. — Please order a reprint rather than making your own copy.

Dieses Werk wurde im Jahr 2013 vom Verlag Zeitschrift für Naturforschung in Zusammenarbeit mit der Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. digitalisiert und unter folgender Lizenz veröffentlicht: Creative Commons Namensnennung-Keine Bearbeitung 3.0 Deutschland Lizenz.

This work has been digitalized and published in 2013 by Verlag Zeitschrift für Naturforschung in cooperation with the Max Planck Society for the Advancement of Science under a Creative Commons Attribution-NoDerivs 3.0 Germany License.

Zum 01.01.2015 ist eine Anpassung der Lizenzbedingungen (Entfall der Creative Commons Lizenzbedingung „Keine Bearbeitung") beabsichtigt, um eine Nachnutzung auch im Rahmen zukünftiger wissenschaftlicher Nutzungsformen zu ermöglichen.

On 01.01.2015 it is planned to change the License Conditions (the removal of the Creative Commons License condition "no derivative works"). This is to allow reuse in the area of future scientific usage.

tions is lost in a finite time τ. For times greater than τ, knowledge of the future is limited to the information contained in the probability distribution of points on the attractor [6]. By assigning a dimension to a probability distribution, the information dimension provides a probabilistic notion of dimension. As we shall see, chaotic attractors can have a self similar structure in their probability distributions that causes their information dimension to differ from their fractal dimension. Just as Mandelbrot [3] calls a set a *fractal* if the fractal dimension exceeds the topological dimension, we will call a probability distribution a *fractal measure* if the fractal dimension exceeds the information dimension.

The direct physical relevance of the information dimension is in measurement. Knowledge of the information dimension of an attractor allows an observer to estimate the information gained when a measurement is made at a given level of precision. This is to be contrasted with the new information gained in a series of measurements that are *not* isolated. In this case there are typically causal relationships between the measured values so that the average information is less than that gained in an isolated measurement. The metric entropy is by definition the upper bound on the information acquisition rate. The metric entropy also estimates the rate at which the accuracy of a prediction of the future decays as the time for prediction increases. We will take positive metric entropy as the definition of chaos. Information dimension and metric entropy are related concepts. Information dimension deals with isolated measurements, and metric entropy with sequences of measurements, so it is natural to discuss them together.

This paper is a synthesis of old and new results. By presenting new results integrated into the background of previous results, I hope that the context of the new results will be clear, and also that logical gaps in previous results will be filled so that they form a coherent whole. Most of the new results are contained in the first section, which develops the concept of information dimension as applied to chaotic attractors. New results include the application of information dimension to dynamical systems, invention of "Cantor's Density", computations on the Henon map, calculations of the information dimension of an asymmetric Cantor set, and computations of the information dimension of the asymptotic probability distribution of the logistic equation. The second section is primarily a review of the metric entropy, exceptions being the development of an algorithm for information dimension and metric entropy, and the expression of the metric entropy in terms of the information dimension.

Part I: Isolated Measurements

Background: Partitions, Measures, and Measurement

To begin the discussion, consider a measuring instrument with a uniform minimum *scale of resolution* ε. For a ruler, for example, ε is the distance between adjacent graduations. If a measuring instrument is assigned to each of the N real variables of a dynamical system, the graduations of these instruments induce a *partition* of the phase space. (A partition $\beta = \{B_i\}$ of a set S is a collection of nonempty, nonintersecting measurable sets B_i that cover S, see Figure 1). Thus, there is an element of the partition corresponding to every possible outcome of a measurement.

Note: A measuring instrument will typically induce a partition only on a region of the phase space, which we will assume to contain the attractors of interest. Also, although measurements of classical quantities can be made continuously they are limited by observational noise, i.e. random perturbations that occur as a measurement is being made. Although continuous, these observations are inevitably recorded as rational numbers with a finite number of digits. In the following discussion assume that the number of digits recorded is commensurate with the quality of an instrument, so that the induced partition forms a good model for the measurement process.

In the limit of perfect resolution the values of the N variables of the dynamical system at any given time may be represented by a point in the phase space. We will refer to this hypothetical point as the *state* of the system. Throughout this paper, assume that the motion in phase space is bounded, and that a sufficient time has passed without a perturbation of the system so that the state is close to an attractor. This means that the probability to find the state in an element of the partition not containing part of the attractor is negligibly small.

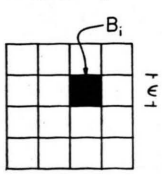

Fig. 1. A schematic drawing of a uniform two dimensional partition. ε is the minimum scale of resolution, and B_i is an element of the partition, corresponding to the outcome of a measurement.

By making repeated measurements at random time intervals, a histogram can be made to estimate the frequency of occurrence $\bar{P}_i$ of the ith element of the partition. The set of probabilities $\{\bar{P}_i\}$ is called the *coarse grained asymtotic probability distribution*. Take the limit as the scale of resolution goes to zero, and the number of samples goes to infinity. Assume that for any fixed set C on the attractor, the sum of $\bar{P}_i$ over elements of the partition covering C approaches a number $\bar{\mu}(C)$. This defines a measure $\bar{\mu}$ on the attractor, with a corresponding probability density $\bar{P}(x)$.

$$\bar{\mu}(C) = \int_C \bar{P}(x)\,\mathrm{d}x\,. \tag{1}$$

Thus, for any partition $\beta = \{B_i\}$, $\bar{P}_i = \bar{\mu}(B_i)$. For convenience, assume that within the basin of attraction, almost every initial condition produces the same measure $\bar{\mu}$. (Numerical experiments indicate that in many cases these are good assumptions.) Let the phase space flow φ be defined as $x(t) = \varphi^t x(0)$, where $x(t)$ and $x(0)$ are the states at time t and time 0. φ^{-t} is the inverse of φ^t. By construction the measure $\bar{\mu}$ is invariant under the action of the flow, i.e.

$$\bar{\mu}(C) = \mu(\varphi^{-t}C) \tag{2}$$

for any set C and time t. Equation (2) is nothing more than a statement that the flow must conserve probability, i.e., the total probability associated with a given set must be equal to the probability of the sets that are mapped into it by the flow.

Lebesgue measure is simply the N dimensional volume on the phase space, constructed by letting $P(x) = 1$. Equation (2) can be made into a recursive equation in order to provide a possible means of constructing an invariant measure. Picking a fixed time interval $t = 1$,

$$\mu_{i+1}(C) = \mu_i(\varphi^{-i}(C))\,. \tag{3}$$

Letting μ_0 be Lebesgue measure, in many cases (3) is known to approach an invariant measure equal to the measure $\bar{\mu}$ of time averages [6]. This will be assumed throughout. Rewritten in terms of the probability density, equation (3) is called the Frobenius-Perron equation.

$$P_{i+1}(y) = \sum_{x=\varphi^{-1}y} \frac{P_i(x)}{|\varphi^1 x|}\,, \tag{4}$$

where $|\varphi^1 x|$ is the Jacobian determinant of $\varphi^1 x$. (For a continuous flow, φ has a unique inverse and the sum in (4) is unnecessary, but for discrete mappings the inverse may not be unique.)

Information Dimension

The amount of information gained in making a measurement depends on the a priori knowledge of the observer making the measurement, i.e. information depends on its context [7]. The context that we shall concentrate on here is that of an observer who has a knowledge of the equations of motion, and all the information that can (in principle) be extracted from them. This observer is therefore capable of computing a coarse grained asymptotic probability distribution $\bar{P}_i$, and can obtain a good approximate knowledge of the measure $\bar{\mu}$.

Suppose that this observer makes an isolated measurement of the state of the system. "Isolated measurement" means that no measurements have been made recently, so that the observer has no ability to predict the state of the system other than the knowledge gained by knowing that it is near a given attractor. The information she gains upon making a measurement is [7]

$$I(\varepsilon) = -\sum_{i=1}^{n(\varepsilon)} \bar{P}_i \log \bar{P}_i\,, \tag{5}$$

where $n(\varepsilon)$ is the number of cells with nonzero probability. The information is written $I(\varepsilon)$ to emphasize its dependence on the scale of resolution. *All the statements in this paper concerning information could equivalently be couched in terms of the entropy* $H(\varepsilon) = -I(\varepsilon)$.

As the scale of resolution decreases, how much does $I(\varepsilon)$ increase? In the limit of small ε, the slope D_{I} of the graph of $I(\varepsilon)$ versus $|\log \varepsilon|$ is called the information dimension

$$D_{\mathrm{I}} = \lim_{\varepsilon \to 0} \{I(\varepsilon)/|\log \varepsilon|\}\,. \tag{6}$$

This definition assumes a special set of measuring instruments with uniform resolution ε. In general this is not the case; the resolution may be nonuniform, or some variables may be measured more precisely than others. Furthermore, there is nothing in the definition that need be specific to attractors of dynamical systems; the information dimension is a property of any random variable with a probability measure defined on it. The following, more technical definition seeks to gener-

ally define D_I in an unambiguous manner:

The diameter $d(C)$ of a set C is $d(C) = \sup_{x,y} d(x, y)$, where $d(x, y)$ is the distance between any two points x and y in C. The diameter d of a partition $\beta = \{B_i\}$ is $d(\beta) = \sup_{B_i} d(B_i)$. A partition of diameter ε is labeled $\beta(\varepsilon)$. The information of β with respect to a measure μ is

$$I(\beta) = -\sum_i \mu(B_i) \log \mu(B_i) . \tag{7}$$

Let M be a manifold with metric d, and let x be a random variable with measure μ. Define $I(\varepsilon)$ as the minimum information possible for a partition of diameter ε, i.e. $I(\varepsilon) = \inf_{\beta(\varepsilon)} I(\beta(\varepsilon))$. The information dimension of (M, d, μ) is

$$D_I = \lim_{\varepsilon \to 0} \{\bar{I}(\varepsilon)/|\log \varepsilon|\} . \tag{8}$$

The information dimension of an attractor is computed by imposing a metric on the dynamical system and using the invariant measure defined by (1). The dependence on the metric is clear from the presence of ε in the definition; nevertheless, we conjecture that the information dimension remains invariant under smooth coordinate transformations, providing that the measure induced by the coordinate transformation is used to compute the information dimension of the transformed attractor.

The information dimension was defined by Balatoni and Renyi [4] in 1956. They refer to it as simply "the dimension of a probability distribution".

Once the information dimension of an attractor is known, without taking the trouble to accumulate or calculate an asymptotic probability distribution, a rough estimate can be made of the quantity of information gained in making a measurement. The accuracy of an experimental instrument is usually given in terms of the signal to noise ratio S. Taking the simplest case in which the extent L of the attractor is roughly the same in every direction, and adjusting the units of measurement so that $L = 1$, the number of bits of experimental resolution is $\log S = |\log \varepsilon|$ (logs to base 2). From (6), the new information gained in a measurement is the order of $D_I \log S$.

$$I \sim D_I \log S . \tag{9}$$

Note: The definition of information dimension, (6) or (8), takes advantage of the fact that, since the limit is taken as $\varepsilon \to 0$, the partition need not cover the attractor efficiently at finite levels of resolution. The definition of information dimension depends on a ratio of $I(\varepsilon)$ to $|\log \varepsilon|$, and is independent of the units used to measure ε. Thus, there is no loss of generality in taking $L = 1$.

A clear distinction should be made between observational noise and external noise. Observational noise, which determines the signal to noise ratio S, occurs only during the measurement process, and at least in the classical limit, does not affect the operation of the dynamical system. External noise is caused by random perturbations that do affect the dynamical system. If the level of external noise exceeds the level of observational noise, then at fine resolution the observer will no longer be measuring properties of the dynamical system; the extra information gained will be information about the perturbations, and will not scale according to D_I. As Shaw has pointed out [6], fractal structure is effectively truncated by the presence of external noise.

Relation to Fractal Dimension

This definition of information dimension is reminiscent of the notion of capacity, or fractal dimension [3], defined for an attractor as follows: Let $n(\varepsilon)$ be the minimum number of balls of diameter ε needed to cover an attractor. The fractal dimension D_F is:

$$D_F = \lim_{\varepsilon \to 0} \{\log n(\varepsilon)/|\log \varepsilon|\} . \tag{10}$$

Note: This does not depend on a measure. Nonetheless, the similarity is clear: If all of the $n(\varepsilon)$ elements of a partition are equally probable, $I(\varepsilon) = \log n(\varepsilon)$. Furthermore, given a cover of balls that minimizes $n(\varepsilon)$, a partition covering the attractor of diameter $\leqq \varepsilon$ can be formed by systematically removing the regions where the balls intersect. $\operatorname{Log} n(\varepsilon) \geqq I(\varepsilon)$, which implies that $D_F \geqq D_I$. The fractal dimension is an upper bound on the information dimension.

The fractal dimension provides a means of estimating the information gained in a measurement when the asymptotic probability distribution is assumed to be uniform, i.e. using Lebesgue measure. To an observer who knows the number of cells to cover the attractor, but does not know their relative probability, the amount of new information gained in a measurement is $I = \log n(\varepsilon) \sim D_F \log S$.

Note: The fractal dimension is a property of a set, rather than a measure, so properly speaking it always refers to the support of the measure, which roughly speaking is the set of points x that have $P(x) \neq 0$, where $P(x)$ is an associated probability density. More precisely, the support of a measure μ is the set of points x such that every open set C containing x has $\mu(C) > 0$.

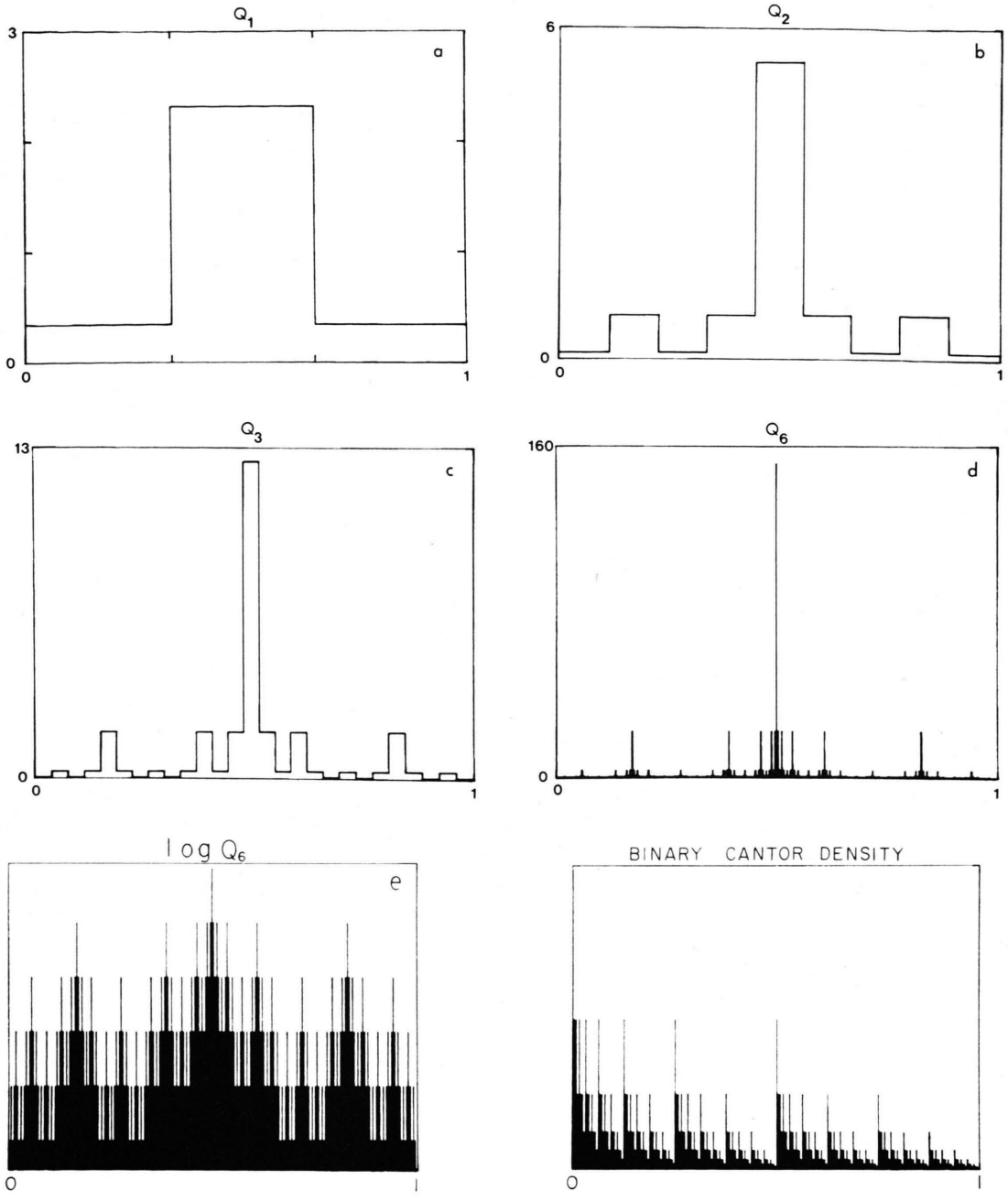

Fig. 2. "Cantor's density" has an information dimension $D_I = \log 2/\log 3$, but the fractal dimension of its support is 1 (see Eq. (12)). (a) $Q_1(x) = 3p_0 \simeq 0.3406$ on (0,1/3) and (2/3,1), and $Q_1(x) = 3p_m \simeq 2.3187$ on (1/3, 2/3). Using a uniform three piece partition, the information $I(1/3)$ is one bit. (b) To form Q_2, the probabilities are redistributed among 9 pieces (see Eq. (7), $i = 2$). Using the 3^2 element partition, $I = 2$ bits. (c) $i = 3$, $I = 3$ bits. (d) $i = 6$. Because of the finite linewidth of the plotter, this is indistinguishable from the limiting case except for the vertical scale (in the limit the spikes become infinite).

Fig. 3. The binary Cantor density. This is similar to Fig. 2, but the interval is partitioned into 2^n rather than 3^n pieces.

Some Examples

(1) Continuous Density

Renyi has proven [8] that a D dimensional random variable with a continuous probability density has an information dimension equal to D. ("D dimensional" here means that the variable is a D-tuple, or more generally, that it is defined over a D dimensional manifold.) This result can be intuitively understood as follows: To compute the information contained in a probability density at finer and finer resolution, it must be broken up into smaller and smaller pieces. Continuity requires that if a very small piece is blown up, it will look approximately uniform. Once a sufficiently fine level of resolution is reached, when the resolution is doubled, each partition element is divided into 2^D new elements of uniform probability. Thus for a continuous density, the information dimension and the fractal dimension are both integers whose dimension is equal to D.

(2) "Cantor's Density"

Since the fractal dimension is an upper bound for the information dimension, any probability density whose support has a noninteger fractal dimension may also have a noninteger information dimension. For example, take Cantor's classic set and treat it as a probability density. Begin with a uniform density on the interval [0, 1], and set the probability of the middle third to zero. Then set the probability of the middle third of each remaining piece to zero, etc. The fractal dimension and the information dimension are both equal to log 2/log 3.

By modifying this construction slightly, however, it is easy to make a probability density whose information dimension and fractal dimension are different. As mentioned in the introduction, we will refer to such objects as fractal measures. To construct an example, rather than deleting the middle third of the interval, make it more (or less) probable. Let the probability of each of the two outer pieces be p_0, and the probability of the piece in the middle be p_m. The first approximation to "Cantor's density" is $Q_1(x) = (3p_0, 3p_m, 3p_0)$. (The factor of 3 appears to make $\int_0^1 Q_1(x)\,dx = 1$, which also implies that $p_0 = (1 - p_m)/2$.) Form $Q_2(x)$ by dividing each piece again into thirds, and redistributing the probability within each of these nine pieces so that the ratios within each third are the same as those of Q_1 (see Figure 2). Alternatively, Q_i can be defined as follows: Round the value of x down to the nearest i digit base 3 number, i.e. $x = .s_1 s_2 \ldots s_i$, where $s = 0$, 1, or 2. Let j be the number of these digits that are 1. Select a value for p_m, and let

$$Q_i(x) = 3^i p_m^j ((1 - p_m)/2)^{i-j}. \tag{12}$$

The limiting density is $Q_\infty(x) = \lim_{i\to\infty} Q_i(x)$. The support of this density is the entire interval, and has a fractal dimension of one, but $Q_\infty(x)$ has an information dimension

$$D_I = \frac{p_m \log \frac{1}{p_m} + 2\,p_0 \log \frac{1}{p_0}}{\log 3}. \tag{13}$$

There are two possible values of p_m that give an information dimension of log 2/log 3. A little algebra shows that these correspond to the fixed points of the binary entropy function

$$H(p) = p \log \frac{1}{p} + (1 - p) \log \frac{1}{(1 - p)}. \tag{14}$$

The unstable fixed point, $p = 0$, corresponds to the classic Cantor set; the stable fixed point $p \cong .7716$ gives the probability density plotted in Figure 2. Since within any interval points can be found that have any number of ones in their base three expansion, Cantor's density (12) is discontinuous everywhere, and has singularities that are dense in the interval.

In a similar manner, a binary Cantor density can be formed by assigning unequal probabilities to subdivisions of the interval into 2^n rather than 3^n parts. The result is plotted in Figure 3. The information dimension of this density is $H(p)$, where p is the probability of one side, and H is the binary entropy function, Equation (14). The binary Cantor density, or reshuffled forms of it, will appear later in several applications.

(3) An Asymmetric, Dissipative Baker's Transformation: Yorke's Map

Although the Cantor density presented in the previous section may seem bizarre, and at a first glance unlikely to be physically relevant, these Cantor-like probability densities are typical of chaotic attractors of dynamical systems. The following example of a dissipative Baker's transforma-

tion was introduced by Yorke [9] in order to illustrate this point:

Define $F(x, y)$ on the unit square $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$ by:

$$\begin{aligned} &\text{for} \quad 0 \leqq y_i \leqq q: \quad & x_{i+1} &= (3 - x_i)/3\,, \\ & & y_{i+1} &= y_i/(1-q)\,; \\ &\text{for} \quad q \leqq y_i \leqq 1: \quad & x_{i+1} &= (1/3)(1 - x_i)\,, \\ & & y_{i+1} &= (y_i - q)/(1-q)\,, \end{aligned} \tag{15}$$

where $0 < q < 1/2$. This mapping divides the unit square into two horizontal strips of unequal heights $p > q$, as shown in Figure 4. It compresses the square horizontally by a factor of 3, and stretches each piece vertically until each piece has a height 1 and a width 1/3. Then it slices the two pieces apart along their boundary, flips both of them about their vertical midlines, and places them as shown in Figure 4.

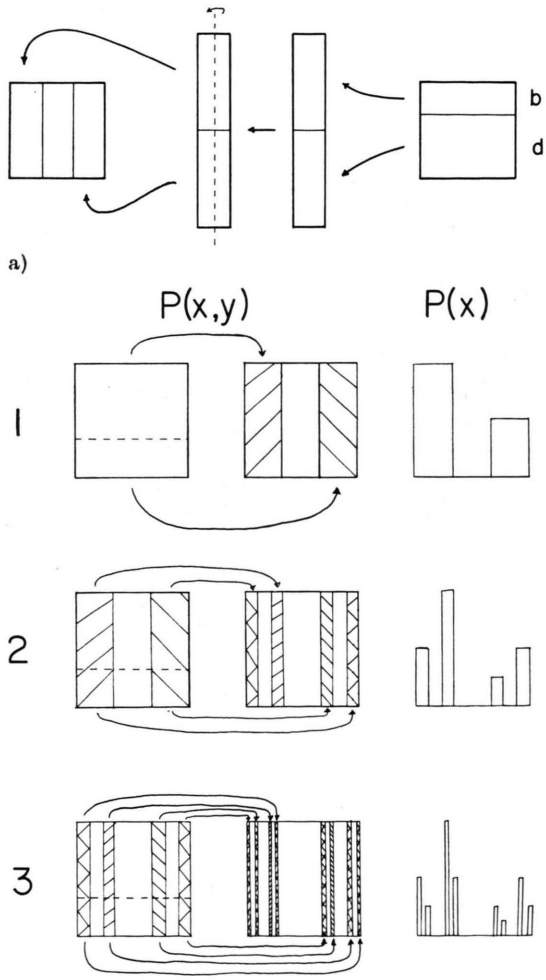

Fig. 4. (a) A geometric illustration of Yorke's Map. (b) Successive iterations of the map act on an initially uniform probability density as shown. The attractor is the Cartesian product of the classic Cantor set and the interval. The information dimension is given by (16).

Start with Lebesgue measure, i.e. uniform probability density, and use (3) to compute an invariant measure by iterating the map. After one iteration, the probability of the two pieces is different, since the thinner bottom strip, $0 < y < q$, is stretched more than the thicker top strip. Letting $p = 1 - q$, the normalized total probability of the left piece is p, and that of the right piece is q. When the mapping is repeated, four thin strips are left with nonzero probability; from left to right, their values are pq, p^2, q^2 ,and pq, as shown in Figure 4. When this process is continued indefinitely, the resulting attractor is the Cartesian product of the classic Cantor set and the interval (0, 1), and has a fractal dimension of $1 + \log 2/\log 3$. The invariant measure on this attractor is extremely nonuniform, however, as can be imagined by extrapolating Figure 4. (If the "holes" are removed, and the order of the probability appropriately reshuffled, the probability density projected onto the x axis is the binary Cantor density of Figure 3.) The information dimension can be computed by dividing the square into 3^{2n} uniform pieces for $n = 1, 2, \ldots$. For any value of n, the ratio

$$I(\varepsilon)/|\log \varepsilon| = I((1/3)^{2n})/\log 3^{2n} \quad \text{is}$$
$$1 + (p \log(1/p) + q \log(1/q))/\log 3\,.$$

Remembering that $p = 1 - q$, the information dimension D_{I} can be written as

$$D_{\mathrm{I}} = 1 + H(q)/\log 3\,, \tag{16}$$

where $H(q)$ is the binary entropy function (see (14)).

The essential feature that causes the difference between the information dimension and fractal dimension of this attractor is that the Jacobian determinant is not a constant. At a coarse level of resolution, each iteration of the mapping results in different probabilities for the "leaves" of the fractal set. Successive iterations move these leaves closer together, so that the resulting probability density has self similar structure on all scales. Since vector fields with constant Jacobians form a set of measure zero in the space of all possible vector fields, chaotic attractors with self similar structure typically have fractal measures. (Self

similar structure is not a necessary condition, however; there may be chaotic attractors that do not have an exact self similar structure that nonetheless have fractal measures.)

(4) Henon's Map

A nonconstant Jacobian determinant is not, however, a necessary condition for a chaotic attractor to have a fractal measure. For example, consider the Henon [10] mapping,

$$x_{i+1} = 1 + y_i - a x_i^2, \quad y_{i+1} = b x_i \tag{17}$$

with $a = 1.4$ and $b = 0.3$. A picture of the attractor is shown in Fig. 5, together with the two blowups as shown by Henon in [10]. The number of observable leaves of this fractal set depends on the quality of this reproduction and the keenness of your eyesight. Figure 5(d) shows a probability distribution made by resolving the blowup of Fig. 5(c) into 6 distinct leaves, and binning the total number of points that visit each leaf into a histogram. As can be seen in the figure, the probabilities of the discernable leaves are quite different.

This is true even though Henon's map is invertible and has a constant Jacobian determinant. Taken together these two properties imply that the probability density on the attractor is uniform. What is not uniform is the spacing of the leaves of the fractal set that make up the attractor. To see this, begin with a set C, and assign it a constant probability density. Iterate the map several times (see Figure 6). The set will be stretched and folded until it begins to resemble the attractor, but according to (4), the probability density on the iterated set $F^i(C)$ remains uniform. The spacing between the folds of $F^i(C)$ is not equal, however, and at a fixed level of resolution the visible leaves

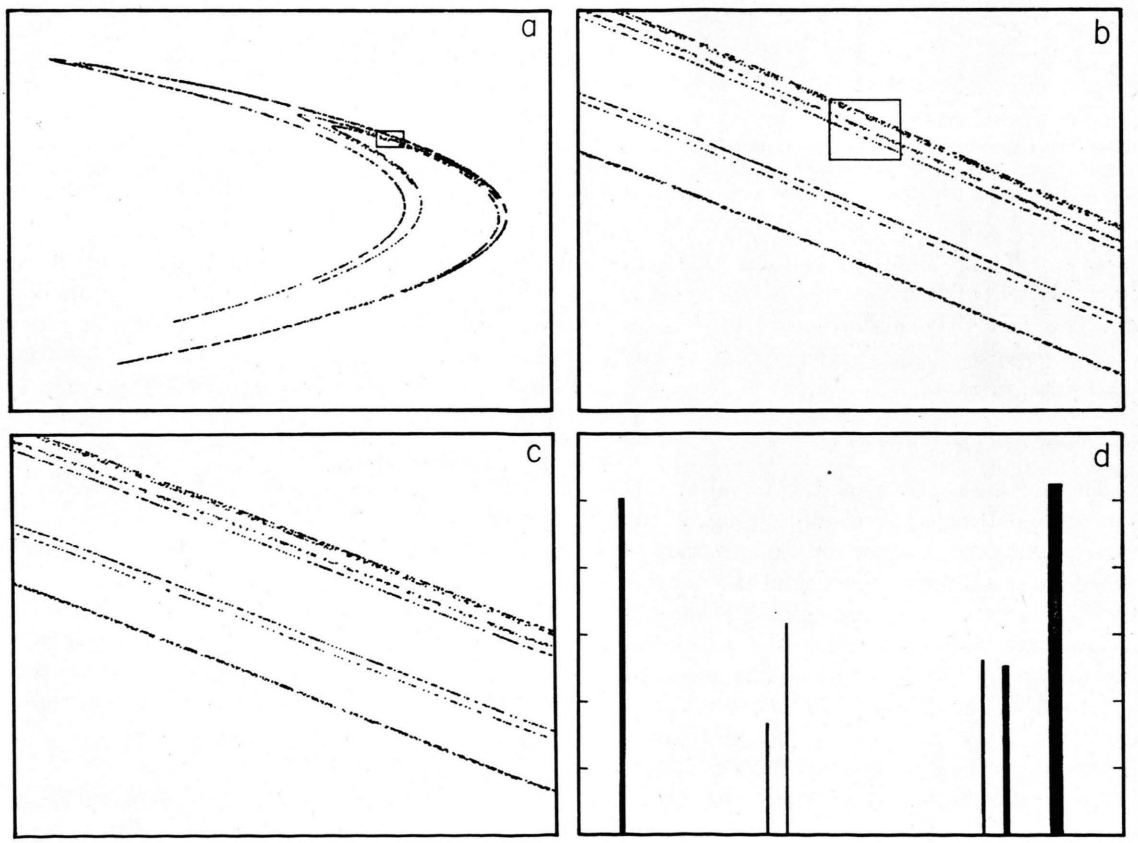

Fig. 5. The Henon attractor, see Equation (17). Figures 5(b) and 5(c) show successive blowups of the regions inside the box of the previous figure. Resolving the fractal structure of Fig. 5(c) into six "leaves", the relative probabilities of these leaves are plotted in Figure 5(d). The width of each bar is roughly the width of the corresponding leaf. This unequal distribution of probabilities persists on all scales, making the Henon attractor a "measure fractal".

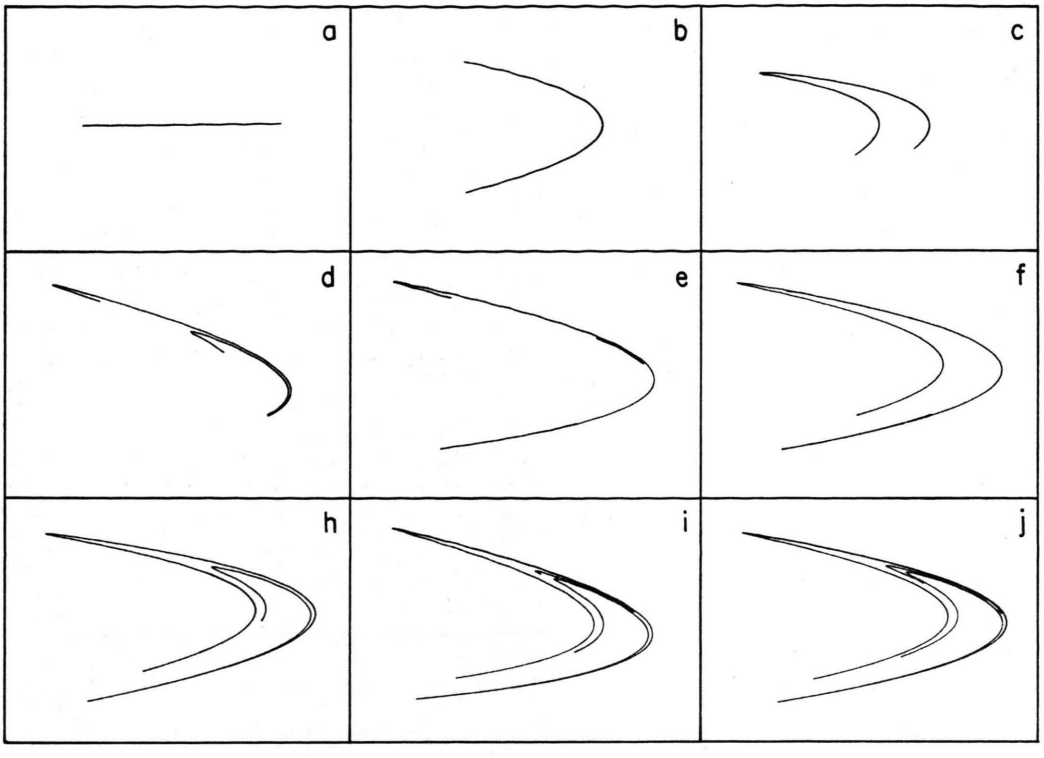

Fig. 6. Successive iterates by the Henon map of the line segment shown in Figure 6a. The spacing between the folds that form after several iterations is not constant; when evaluated at a finite level of resolution unequal numbers of them appear to merge, yielding the probability structure seen in Figure 5(d).

may contain different numbers of folds of $F^i(C)$. The total probability of each resolvable leaf appears different. The fact that this structure persists on such a microscopic level suggests the Henon attractor has a fractal measure.

(5) An Asymmetric Cantor Set

If the intersection is taken of the Henon attractor and a curve cutting transversally through it, the result is an asymmetric Cantor set. Construction of an example gives an understanding of the manner in which asymmetries in the spacing of the elements of a set can create a fractal measure. Begin as usual with the interval [0, 1], and delete the segment from 1/2 to 3/4, i.e. delete the third fourth, as shown in Figure 7(a). Then delete the third fourth of each remaining piece, and so on. The limit is an asymmetric Cantor set, of fractal dimension $D_F = \log 3/\log 4$.

Assign a uniform probability density to the points of this set. To compute the information dimension, partition the interval (0, 1) successively into 2^n equal bins, as shown in Figure 7(b). Each time the resolution is cut in half, the asymmetry causes neighboring bins to acquire different probabilities. This inequality persists on all scales. The information dimension is computed in Appendix I, and shown to be

$$D_I = \frac{3\,H\left(\frac{1}{3}\right)}{4\log 2}\,.$$

This example illustrates that the information dimension depends on the metric properties of a set (spatial distribution of the elements of the set) as well as the distribution of probability within the set. Asymmetries in either of these properties can cause a chaotic attractor to have a fractal measure.

(6) The Logistic Equation

One of the properties that is necessary for a measure to be a fractal measure is structure on all scales. As the following example shows, this is not a sufficient condition; fractal measures must have "geometrically multiplying" small scale structure.

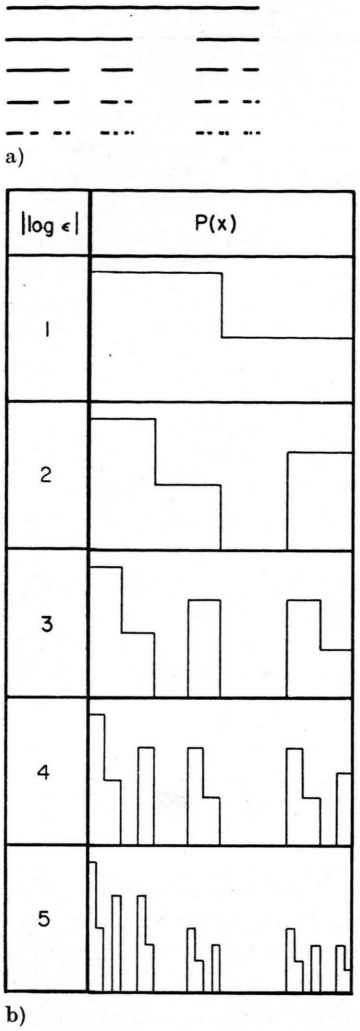

Fig. 7. (a) Successive approximations to an asymmetric Cantor set, formed by deleting the third fourth of each continuous piece. (b) Assuming the initial probability density of points on the interval is uniform, the coarse grained probability distribution is evaluated at successively finer scales of resolution $\varepsilon = 2^{-i}$, $i = 1, 2, \ldots, 5$.

Consider a mapping of the interval with a single quadratic maximum, for example, the logistic equation

$$x_{i+1} = r x_i (1 - x_i)\,, \tag{19}$$

where $0 < r < 4$, and $0 \leqq x \leqq 1$. Assume that r is set to a parameter value for which this mapping has a chaotic attractor. An invariant measure can be computed either using the Frobenius-Perron equation (4) or alternatively by picking an initial condition at random, iterating many times, and binning the result into a histogram. The result of the latter procedure at $r = 3.7$ is shown in Figure 8. Although this probability density has structure on all scales of resolution, a direct calculation with a uniform partition shows that the entropy scales linearly with $|\log \varepsilon|$ (see Figure 9). An investigation into the nature of $\bar{P}(x)$ makes this result reasonable:

Using an initial condition $P_0(x) = 1$, suppose the Frobenius-Perron equation (4) is used to construct $\bar{P}(x)$. Providing the iterates of the critical point are not asymptotically periodic, at every iteration the fact that $F(x) = 0$ at the critical point $x = z_0$ causes a new singularity to be added to $P_j(x)$. The resulting asymptotic density $\bar{P}(x)$ is not continuous, and contains an infinite number of singularities at $z_i = F^i(z_0)$. Nevertheless, since each singularity falls off on one side roughly as $x^{-1/2}$, the singularities are integrable (see Appendix II). Thus a coarse grained probability distribution $\{\bar{P}_i\}$ computed in this manner will converge in root-mean-square to the probability density $\bar{P}(x)$; the resulting measure is absolutely continuous with respect to Lebesgue measure. When ordered according to iteration from the critical point, the amplitude of the singularities decreases roughly exponentially at a rate given by the Lyapunov exponent,

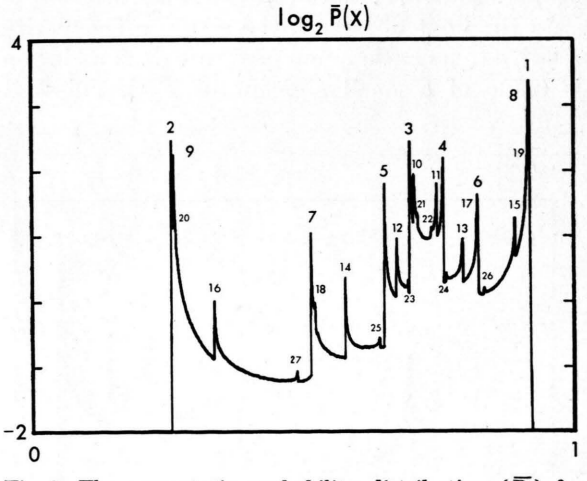

Fig. 8. The asymptotic probability distribution $\{\overline{P_i}\}$ for the logistic equation $x_{i+1} = 3.7 x_i (1 - x_i)$, obtained by iterating 10^8 times and binning the result into 1000 bins. The corresponding density $\overline{P}(x)$ contains an infinite number of integrable discontinuities, labeled in the figure according to the iterate from the critical point. Their amplitude decreases roughly exponentially. The information dimension $D_I = 1$.

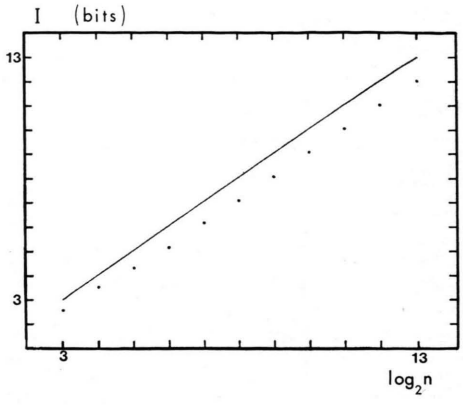

Fig. 9. The information I as a function of the logarithm of the scale of resolution. The interval is evenly divided into n equal pieces, and $\bar{P}(x)$ for the distribution shown in Fig. 8 is averaged over each piece. The dots are measured values of $I(\varepsilon)$, where $\varepsilon = 1/n$. The line is a plot of $\log n$, shown for reference, and is the information contained in a uniform $\bar{P}(x)$.

as demonstrated in Appendix II. The following approximate form may add insight into the behavior of $\bar{P}(x)$ (see Figure 10).

$$\bar{P}(x) = \sum_{i=1}^{\infty} \frac{\theta(\tilde{x}_i)\,|\tilde{x}_i|^{-1/2}}{|z_i'|^{1/2}}, \tag{20}$$

where $z_i' = [F^i(x)]'_{x=z_1}$, and $\tilde{x}_i = (x - z_i)$ multiplied by the sign of z_i. $\theta(x)$ is the step function, i.e. $\theta(x) = 0$ for $x < 0$, and $\theta(x) = 1$ for $x \geqq 0$.

In the limit of large i, the size of a singularity relative to its background goes to zero. If a blowup is made of a small segment of $\bar{P}(x)$, unless it happens to contain one of the early iterates of the critical point, it looks relatively flat. When the segment is split in half to go to a finer level of resolution, the probability of the segment will divide evenly between the two new pieces. Asymptotically (when viewed at fine resolution), this density behaves like a uniform density almost everywhere. At a coarse level of resolution, the information is less than that of a uniform density, but at fine resolution the information increases linearly, and the information dimension is one, as shown in Figure 9.

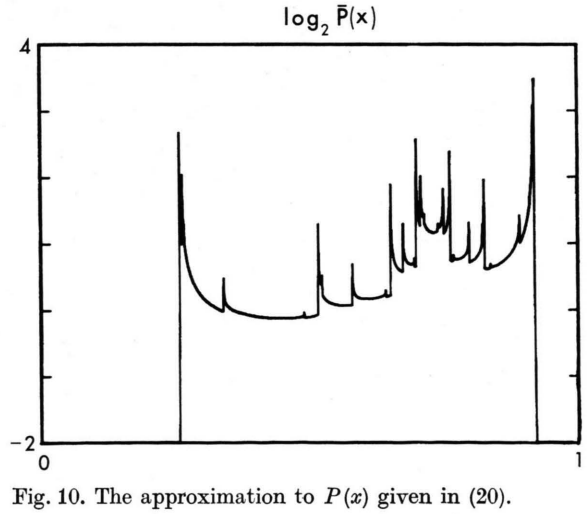

Fig. 10. The approximation to $P(x)$ given in (20).

This example illustrates the manner in which small scale structure must occur to make a fractal measure. For the logistic equation, the size of the visible spikes decreases exponentially, but the number of spikes in a given logarithmic size range is fixed, independent of scale. Compare this with Fig. 2(d) or Figure 2(e). The size of the spikes decreases exponentially, but the number of spikes visible at finer scales also increases exponentially. When a blowup is made of a small piece, structure on the scale of that piece is visible almost everywhere.

Part II: Sequential Measurements

Metric Entropy as Information Acquisition Rate

The discussion so far has concerned the amount of information gained by an observer in making a single, isolated measurement, i.e., the information gained in taking a "snapshot" of a dynamical system. We might alternatively ask how much new information is obtained per unit time by an observer who is, say, watching a movie of a dynamical system. In other words, what is the information acquisition rate of an experimenter who makes a series of measurements to monitor the behavior of a dynamical system? For a predictable dynamical system eventually new measurements provide no further information. But, if the dynamical system is chaotic, new measurements are constantly required to update the knowledge of the observer.

The metric, or Kolmogorov-Sinai entropy provides an upper bound on the information acquisition rate. Although in a context unrelated to dynamical systems, the metric entropy was originally defined by Shannon [11]. This quantity was later applied to dynamical systems, and shown to be a topological invariant by Kolmogorov [12] and Sinai [13]. To

simplify the discussion, we will follow the historical course by first defining the metric entropy of a Markov process, and then discussing it in the context of dynamical systems (see also Refs. [14—18]).

A Markov process may be thought of as an information source that randomly generates one of n symbols at discrete timesteps [19]. The text of this paper, for example, can be considered to constitute a Markov process. These symbols can be thought of as the outcome of sequential measurements. For convenience, assume that the n possible symbols are the integers from 0 to $n-1$. If the occurrence of a given symbol s depends on $m-1$ preceding symbols, then the Markov process is of mth order. Assume throughout that the Markov process is ergodic, so that time averages are independent of starting time, and are equal to ensemble averages. For an mth order process whose current symbol is s_{m} the sequence including the previous $m-1$ symbols, $(s_1, s_2, \ldots, s_{\text{m}})$, can be represented as the m digit base n fraction $.s_1 s_2 \ldots s_{\text{m}}$, abbreviated S_{m}. This number defines the state of the source. (This should not be confused with the state of a dynamical system.) The probability that a given state S_{m} will occur is

$$P(s_1, s_2, \ldots, s_{\text{m}}) = P(S_{\text{m}}).$$

Given that the source is in state S_{m}, the conditional probability that the $m+1$ symbol will be $s_{\text{m+1}}$ is $P(S_{\text{m}} | s_{\text{m+1}})$, and the amount of new information obtained on learning that this particular transition has occurred is $I = -\log P(S_{\text{m}} | s_{\text{m+1}})$. Averaging over all possible transitions from S_{m} to $s_{\text{m+1}}$, and over all possible states S_{m}, the average amount of new information gained per symbol, ΔI_{m}, is

$$\Delta I_{\text{m}} = -\sum_{S_m=0}^{1-n^{-m}} P(S_{\text{m}}) \sum_{s_{m+1}=0}^{n-1} \cdot P(S_{\text{m}} | s_{\text{m+1}}) \log P(S_{\text{m}} | s_{\text{m+1}}).$$

Making use of the facts that

$$P(S_{\text{m}}, s_{\text{m+1}}) = P(S_{\text{m}}) P(S_{\text{m}} | s_{\text{m+1}})$$

is the joint probability for S_{m} and $s_{\text{m+1}}$ to occur in succession, and that probability is conserved when the system makes a transition to a new state, i.e.

$$\sum_{s_{m+1}=0}^{n-1} P(S_{\text{m}}, s_{\text{m+1}}) = P(S_{\text{m}}) \tag{22}$$

and defining

$$I_{\text{m}} = -\sum_{S_m=0}^{1-n^{-m}} P(S_{\text{m}}) \log P(S_{\text{m}}), \tag{23}$$

the average amount of new information gained per symbol emitted by the source can be rewritten

$$\Delta I_{\text{m}} = I_{\text{m+1}} - I_{\text{m}}. \tag{24}$$

For an ergodic source, ΔI_{m} should be independent of m when m is greater than the order of the source. If the source, is not ergodic, however, ΔI_m may oscillate for large m. This problem may be avoided by taking the limit as m goes to infinity, and defining the information rate $\Delta I/\Delta t$ as

$$\Delta I/\Delta t = \lim_{m\to\infty} \{I_m/m\Delta t\}, \tag{25}$$

where Δt is the time interval between symbols. Shannon calls this simply the entropy of the source; however, we will call it the information rate to avoid confusion with other entropies. To an observer with a knowledge of the past history of the source, $\Delta I/\Delta t$ is a measure of the unpredictability of the sequences it generates.

Metric Entropy and Refinement of Initial Conditions

A sequence of numbers obtained by making a series of measurements may be thought of as the symbols emitted by a Markov source. A distinct symbol is assigned to each of the n elements of the partition induced by the measuring instrument. A measurement at time $t=0$ determines that the state of the dynamical system is located somewhere inside a given element of the partition. A finite time Δt later, another measurement may give a different outcome, indicating that the state is in a different partition element. Thus, a series of measurements generates a sequence of symbols. The information rate per unit time for this symbol sequence may be calculated by taking the limit as $m\to\infty$ of $\Delta I_m/m\Delta t = \Delta I/\Delta t$. Following Sinai [13], the metric entropy can be defined as the maximum information rate when the partition and sampling rate are varied.

$$h_\mu = \sup_{\beta,\Delta t} \lim_{m\to\infty} \{I_m/m\Delta t\}. \tag{26}$$

To an observer with optimal measuring instruments taking samples at the optimal rate, the metric entropy is the average amount of new information gained per sample.

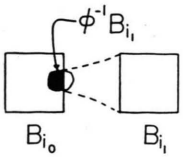

Fig. 11. A schematic illustration of the manner in which initial conditions are refined by a sequence of measurements.

A sequence of several measurements allows an initial condition to be isolated more precisely than does any one of the measurements taken alone. Suppose, for example, that a sequence of measurements taken at times $0, 1, \ldots, m$ show that the state of the system was successively in the partition elements $B_{i_0}, B_{i_1}, \ldots, B_{i_m}$. The sets B_i evolve deterministically under the action of the flow φ. Thus, for example, if at time $t=1$ the state is somewhere inside B_{i_1}, at time $t=0$ it must have been somewhere inside $\varphi^{-1}(B_{i_1})$. When the two measurements at time $t=0$ and $t=1$ are taken together, at time $t=0$ we know that the original initial condition at $t=0$ must have been somewhere inside of $\varphi^{-1}(B_{i_1}) \cap B_{i_0}$ (see Figure 11). After m measurements, it is located somewhere inside of m such intersections. In this manner a sequence of measurements can be used to refine knowledge of an initial condition.

To obtain the average rate of refinement, it is necessary to average over all possible initial conditions. This can be done by examining simultaneously the intersection of all the elements of a partition with their own inverse images. Let the intersection of two partitions α and β be the partition formed by taking all possible intersections of their elements.

$$\alpha \wedge \beta = \{A_i \cap B_j\} \tag{27}$$

for all $A_i \in \alpha$ and all $B_j \in \beta$. The intersection of a partition β with the partition formed out of its inverse images is called a refinement of the partition. If this is done m times the resulting partition, β_m, is called the mth refinement of β with respect to φ.

$$\beta_m = \beta \wedge \varphi^{-\Delta t}\beta \wedge \varphi^{-2\Delta t}\beta \wedge \cdots \wedge \varphi^{-m\Delta t}\beta\,. \tag{28}$$

(For convenience the dependence of β_m on Δt is not explicitly written.) The information $I(\beta_m)$ contained in β_m is exactly the information I_m contained in the corresponding sequence of m symbols. This motivates the standard form of the definition of metric entropy [20, 21]

$$h_\mu = \sup_{\beta,\Delta t} \lim_{m\to\infty} \{I(\beta_m)/m\Delta t\}\,. \tag{29}$$

Thus, the metric entropy may alternatively be thought of as measuring the extra amount of knowledge gained about an initial condition with each new measurement. For a predictable system, the trajectories on the average diverge at a linear or polynomial rate, and asymptotically a new measurement ceases to provide better definition of the initial condition. For chaotic systems, on the average trajectories diverge locally at an exponential rate, and each successive measurement provides new information.

The definition of metric entropy requires a measure (Ironically, it does not require a metric). To define the metric entropy of an attractor it is necessary to choose a measure on the attractor. The physically relevant choice is the measure for time averages, defined in (3). Again, we will assume that this is the same for almost every initial condition (with respect to Lebesgue measure) in a given basin. By definition a chaotic attractor has positive metric entropy.

Topological Entropy

The topological entropy is the upper bound on the information acquisition rate for an observer who knows what symbol sequences occur, but does not know their relative probability. In this case, the information contained in a refined partition β_m is $\log N_m$, where N_m is the total number of elements of β_m. For a partition consisting of n symbols (n elements), N_m is the number of distinct symbol sequences that actually occur, in contrast to the n^m symbol sequences that might occur. The topological entropy can be defined as

$$h_t = \sup_{\beta,\Delta t} \lim_{m\to\infty} \{\log N_m/m\Delta t\}\,. \tag{30}$$

This is only one of several equivalent methods of defining the topological entropy. The topological entropy is a topological invariant, and can be defined, for example, as the exponent of the rate of increase of the number of periodic orbits (not necessarily stable) as the allowed period increases. It requires neither a metric nor a measure for its definition. A more thorough discussion of topological entropy can be found in [22]—[24].

Metric Entropy as Information Dimension of P_∞

By constructing a probability density for symbol sequences on the interval (0, 1), the relationship

of the metric entropy to the information dimension (and the topological entropy to the fractal dimension) becomes clear. For an n symbol partition (n possible outcomes of a measurement), the set of probabilities $P(S_m)$ of the n^m possible symbol sequences of length m can be visually represented in the following manner:

Divide the interval (0, 1) uniformly into n^m bins. Since each of the n^m possibly symbol sequences S_m can be represented as a unique number $.s_1 s_2 \ldots s_m$, each symbol can be assigned to the bin at the position corresponding to this number. Construct a probability density function $P_m(x)$ by plotting the probability $P(S_m)$ of each sequence over its bin. Normalizing so that $\int_0^1 P_m(x)\,\mathrm{d}x = 1$,

$$P_m(x) = P(S_m)\,n^m,$$
$$S_m \leqq x < S_m + n^{-m},$$

where $S_m = 0,\ n^{-m}, \ldots, 1 - n^{-m}$. In the limit as $m \to \infty$, define $P_\infty = \lim_{m\to\infty} P_m$. Although P_∞ may be a very nasty probability density, the normalization condition insures that it induces a well defined measure μ_∞ (see Equation (1)).

Turning this construction around, $P(S_m)$ can be viewed as the result of smearing the limiting distribution P_∞ over each of n^m bins

$$P(S_m) = \int_{S_m}^{S_m + n^{-m}} P_\infty(x)\,\mathrm{d}x = \mu_\infty(i_{S_m}) \tag{32}$$

for $S_m = 0$ to $1 - n^{-m}$, with i_{S_m} the interval from S_m to $S_m + n^{-m}$. Studying symbol sequences of length m is equivalent to examining the symbol sequence probability density P_∞ at a scale of resolution of $\varepsilon = n^{-m}$. This mapping of the set of probabilities $P(S_m)$ onto the unit interval allows the metric entropy to be rewritten in terms of the information dimension of P_∞. Since $|\log \varepsilon| = m \log n$, the metric entropy h_μ can be written:

$$h_\mu = \lim_{m\to\infty} \{I_m / m \log n\} \log n,$$
$$h_\mu = D_{\mathrm{I}}(P_\infty) \log n, \tag{33}$$

where $D_{\mathrm{I}}(P_\infty)$ is the information dimension of P_∞. Log n is the amount of information per symbol if the symbols are of equal probability and do not depend on the previous symbols. $D_{\mathrm{I}}(P_\infty)$ measures the nonuniformity of the probability of symbol sequences.

In a manner similar to the metric entropy, the topological entropy can be expressed as

$$h_t = D_{\mathrm{F}}(P_\infty) \log n, \tag{34}$$

where $D_{\mathrm{F}}(P_\infty)$ is the fractal dimension of the support of P_∞ (see also [27]).

Examples:

For numerical computations of the information rate the reader may wish to see Shimada [25] (Lorenz equations), Kaufman et al. [26] (Chirikov's map), Crutchfield and Packard [27] (logistic equation), and Curry [28] (Henon map). The principal purpose of the following examples is to make clear the nature of the metric entropy and its relationship to the information dimension.

(1) The Binary Shift

Take a number between 0 and 1, multiply it by two, and truncate it so that it is still between 0 and 1. Written in terms of a base two representation:

$$.101100111 \to .01100111\ ? \tag{35}$$

Alternatively, this can be thought of as the one dimensional map shown in Figure 12. If the digit that is shifted past the decimal point on a given iteration is 1, we can say that this Markov source has generated a 1; if it is 0, it has generated a zero. Equivalently, the interval can be partitioned into two equal pieces, and the elements of the partition assigned the symbols 0 and 1. This source emits one bit of information per iteration. The refinements of this partition, found by computing the inverse images of .5, divide the interval into 2, 4, 8, ... equal pieces of equal probability for $m = 0, 1, 2, \ldots$. The information contained in the mth refinement is $\log 2^m = m$. Since possible symbol sequences are equally likely, the symbol sequence density P_∞ is a constant, and has an information dimension of one. Thus, using any of the formulations that have been discussed here, the information rate is one bit per iteration (see [6] and [21]).

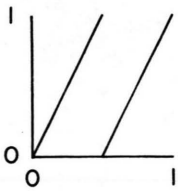

Fig. 12. The binary shift map (see Equation (35)).

This map with the partition above is a model of a fair coin flip. If the interval is partitioned into two pieces of unequal length p and $1-p$, however, the coin is not fair. In this case the density P_∞ is not uniform; heads is weighted differently than tails, so that in passing from sequences of length m to sequences of length $m+1$, the probability in each bin of width 2^{-m} is split unequally between two bins of width 2^{-m-1}. The resulting density P_∞ is the binary Cantor density, shown in Figure 3. The information production rate is the information dimension of P multiplied by log 2; the information rate is therefore $H(p)$, where $H(p)$ is the binary entropy function (see Eqs. (14) and (15)). The equal partition $p=1/2$ gives the maximum value $H(1/2)=1$, corresponding to the metric entropy of the map.

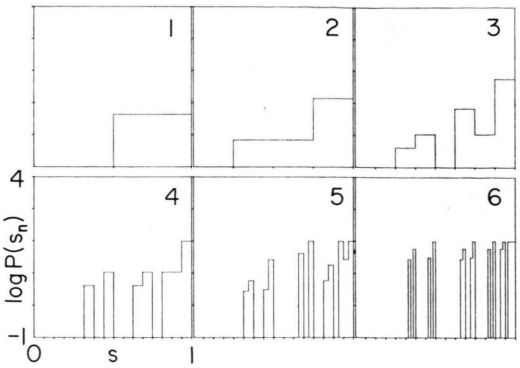

Fig. 13. P_∞ for the logistic equation (19) with $r=3.7$, taken from [27]. The interval is partitioned into two equal parts, which are assigned the symbols 0 and 1. Symbol sequences $S_m = .s_1\, s_2 \ldots s_m$ are generated by iterating the map, and binned to approximate $P(S_m)$. Successive approximations to P_∞ are shown in the figure. The metric entropy is the information dimension of P_∞, and the topological entropy is the fractal dimension (using $\log_2$).

(2) The Logistic Equation Revisited

In a similar manner it is possible to compute the information production rate of the logistic equation, (19). For $r=4$, the logistic equation is symmetric and two onto one. The invariant probability density $\bar{P}(x)=1/\pi(x(1-x))^{-1/2}$ is also symmetric. As pointed out by Shaw [6], assuming also that there are no stable periodic orbits, these facts imply that the metric entropy with respect to $\bar{P}(x)$ is one bit per iteration, since in order to reconstruct a trajectory going backwards in time a binary decision must be made at each iteration. For $r \neq 4$, the information rate can be computed numerically by making a partition of the map and iterating many times. A histogram can be made of the relative frequency of occurrence of the symbol sequences, to approximate P_∞, as shown in Figure 13. This figure shows the characteristic behavior of a fractal measure as the scale of resolution is varied; compare to Fig. 7(b), for example. As demonstrated numerically by Crutchfield and Packard [27], the information rate is maximized for a two element partition when the division is made at the critical point $x=1/2$.

Equivalently, the symbol sequence probability distribution $P(S_m)$ can be demonstrated by computing the measure $\bar{P}(x)$ of points on the interval (Fig. 8) via any convenient method, computing the inverse iterates of the critical point to get a refined partition, and summing $\bar{P}(x)$ over each of the elements of the refined partition to calculate the probabilities $P(S_m)$ of the corresponding symbol sequences. If the elements of the refined partition β_m are B_{m_i}, the probabilities of the symbol sequences S_m are

$$P(S_m) = \int_{\beta_m} \bar{P}(x)\, dx = \bar{\mu}(B_{m_i}). \tag{36}$$

(3) Yorke's Map Revisited

Let us reconsider Yorke's map (15) in the context of sequential measurements. Partition the square into two pieces by dividing it along the line $y=q$, as shown in Figure 14. The preimages of this partition are horizontal strips whose widths are distributed according to a binomial distribution, and are shown on the left side of Figure 14. Now, consider a sequence of measurements at successive

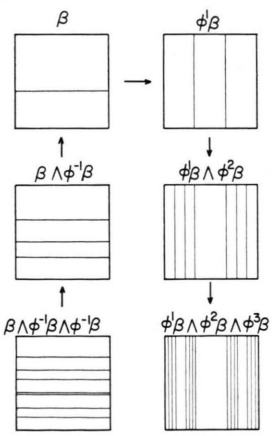

Fig. 14. Iterates of the partition of Yorke's map shown in the upper left corner (refer to Eq. (15) and Figure 4). The preimages of the partition are shown on the left side, and the images on the right side. When a combination of images and preimages is used, an initial condition can be determined arbitrarily well.

iterations of the map made at times

$$-m, \quad -m+1, \ldots, 0, \ldots, m-1, m.$$

If the map is used to compare all the preimages of this sequence of measurements at the same past time $-m$, the y value at $t=-m$ can be determined to arbitrary precision as m increases. From the example of part 1, Fig. 4, we know that the measure μ is uniform vertically, so the measure of the preimages is just proportional to their areas. Except for a reshufflling of their order, this distribution of probabilities is exactly the binary Cantor density (Figure 3). Thus the information rate is $H(q)$, the binary entropy function of q (Equation 14).

Now, consider the images of this partition. These are the uniform vertical bars shown on the left side of Fig. 14, and are exactly the successive approximations to the attractor shown in Figure 4. If the images of each measurement are compared at the same future time m, the horizontal position, and hence the "leaf" of the Cantor set that the state occupies at time m can be determined to arbitrary precision. Although the areas of the images are the same, their probability is not; the relative probabilities are the same as those plotted on the attractor in Figure 4.

The agreement between these probability distributions is not a coincidence. Since Yorke's map is invertible, either preimages or images of the partition can be used to make refinements. Conservation of probability requires that the measure of each element of the partition remain the same under the action of the map. If both images and preimages are used to evaluate all the measurements at the same time $t=0$, both the horizontal and the vertical position can be determined arbitrarily well as m grows large. A partition such as this one with the property that every point on the attractor can be represented arbitrarily well is called a generator. From Sinai's theorem [21], the information rate computed using a generator is equal to the metric entropy.

This example illustrates the ambiguities that are possible in constructing P_∞. On one hand, using the two symbol partition shown on the left of Fig. 13, P_∞ is exactly the binary Cantor density. If, on the other hand, the three symbol partition shown on the right of Fig. 13 is used, then P_∞ is just the same as the density $\bar{P}(x)$ on the attractor projected onto the x axis, as shown in Figure 4. Although these two versions of P_∞ have different information dimensions, the rule given by eq. (33) gives the same metric entropy.

Predicting the Future

Taken together the metric entropy and information dimension can be used to estimate the length of time that a physical prediction remains valid. The information dimension allows an estimate to be made of the information contained in an initial measurement, and the metric entropy estimates the rate at which this information decays.

As we have already seen, the metric entropy is an upper bound on the information gained in each measurement of a series of measurements. But, if each measurement is made with the same precision, the information gained must equal the information that would have been lost had the measurement not been made. Thus, the metric entropy also quantifies the initial rate at which knowledge of the state of the system is lost after a measurement.

Now, to make the notion of "information contained in a prediction" more precise: Let $\beta=\{B_i\}$ be a partition. For convenience, refer to B_i, the elements of the partition, as simply outcomes. Let

$$p_{ij}(t) = \begin{array}{l}\text{probability that a measurement at time } t \text{ has outcome } B_j \\ \text{if a measurement at time 0 had outcome } B_i.\end{array} \tag{37}$$

$p_{ij}(0)$ must satisfy

$$\begin{aligned} p_{ij}(0) &= 1 \quad \text{if} \quad i=j; \\ p_{ij}(0) &= 0 \quad \text{if} \quad i \neq j. \end{aligned} \tag{38}$$

If B_i is known, the information gained on learning of outcome B_j is $-\log p_{ij}(t)$. With no initial information, the information gained from the measurement is determined solely by the asymptotic measure μ, and is $-\log \bar{\mu}(B_j)$. The extra information gained using a prediction from the initial data is the difference of the two, or

$$I_{\text{rel}} = \log \{p_{ij}(t)/\bar{\mu}(B_j)\}. \tag{39}$$

This must be averaged over all possible measurements B_j at time t, and all possible initial measurements B_i. The measurements B_j are weighted by their probability of occurrence $p_{ij}(t)$, and the initial

measurements are weighted by $\bar{\mu}(B_i)$. This gives

$$I(t) = \sum_i \bar{\mu}(B_i) \cdot \sum_j p_{ij}(t) \log \{p_{ij}(t)/\bar{\mu}(B_j)\}. \tag{40}$$

$I(t)$ is the average amount of information contained in a prediction made a time t into the future.

Using the invariant measure $\bar{\mu}$, $p_{ij}(t)$ can be expressed in more geometrical terms: At time t the initial measurement B_i evolves into $\varphi^t B_i$; the probability of outcome B_j at time t is therefore $\bar{\mu}(\varphi^t B_i \cap B_j)$ (see Figure 15). Dividing by $\bar{\mu}(B_i)$ to make $p_{ij}(0)=1$, $p_{ij}(t)$ can be written in the alternate form

$$p_{ij}(t) = \bar{\mu}(\varphi^t B_i \cap B_j)/\bar{\mu}(B_i). \tag{41}$$

Combining (40) and (41) get

$$I(t) = \sum_{i,j} \bar{\mu}(\varphi^t B_i \cap B_j) \cdot \log \{\bar{\mu}(\varphi^t B_i \cap B_j)/\bar{\mu}(B_i)\,\mu(B_j)\}. \tag{42}$$

At time $t=0$, $\varphi^t B_i \cap B_j = B_i$ if $i=j$, and equals zero otherwise. Thus $I(0)$ is exactly the information contained in an initial condition, given by (7) or (5).

An attractor is mixing [19] if

$$\lim_{t\to\infty} (\varphi^t B \cap A) = \bar{\mu}(A)\bar{\mu}(B) \tag{43}$$

for any sets A and B on the attractor. Intuitively, this just means that B is "smeared" throughout the attractor by the flow, so that the probability of finding a point originating in B inside of A is just the original probability of B, weighted by the probability of A. For an attractor that is mixing, substituting (43) into (42) implies that $I(\infty)=0$.

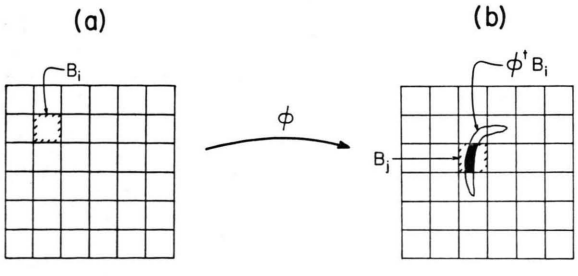

Fig. 15. (a) Assume that an initial measurement finds that the state is contained in partition element B_i. (b) Given outcome B_i at time 0, a time t later the state must be somewhere inside $\varphi^t B_i$; the probability that a measurement at time t will have outcome B_j is $\bar{\mu}(B_j \cap \varphi^t B_i)$, where $\bar{\mu}$ is the invariant measure (which gives the proper a priori weight to each outcome).

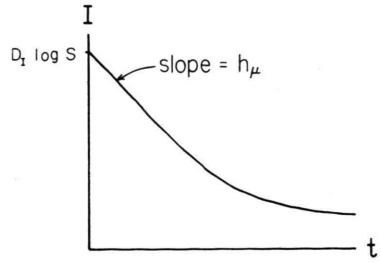

Fig. 16. The typical behavior of $I(t)$, the information contained in a prediction, for a chaotic attractor. Initially $I(t)$ decreases at a linear rate given by the metric entropy.

On a chaotic attractor, adjacent trajectories on the average diverge exponentially, and $I(t)$ initially decreases at a linear rate, as shown in Figure 16. Goldstein [29] has shown that this initial rate is equal to the metric entropy for flows with a continuous invariant measure. As demonstrated in Ref. [30], the metric entropy quantifies the rate of transverse mixing. In contrast, the longitudinal rate of mixing does not generally depend on the metric entropy. The longitudinal mixing rate determines the long time decay of $I(t)$, and many of the properties of power spectra [31].

Since the information decays initially at a linear rate, at a given level of precision the metric entropy and information dimension can be used to estimate $I(t)$ for short times. For initial measurements made with a signal to noise ratio S,

$$I(t) = I(0) - h_\mu t = D_I \log S - h_\mu t \tag{44}$$

(see Equation (9)). The initial data becomes useless after a characteristic $\tau = (D_I/h_\mu) \log S$.

Experimental Applications: Infinite Dimensions, Limited Knowledge

Thus far we have assumed a dynamical system of finite dimension N. In contrast, many important physical systems, for example fluid flow, are described by partial differential equations, which have an infinite dimensional phase space. In this case, to determine a completely arbitrary initial condition, an infinite number of real variables must be measured, corresponding to the infinite degrees of freedom possible in choosing an initial function. When the state of the system is close to an attractor, however, the number of degrees of freedom may be reduced. Throughout the following discussion we will assume that the attractors considered can be contained in a smooth manifold of dimension m_d.

When this assumption holds the number of measurements required to determine an initial condition at a given level of precision is $M \leqq 2m_d + 1$. ("Number of measurements" can refer to simultaneous measurements of M real variables, or as discussed below, a single variable can be measured M successive times.) Assume that the attractor can be embedded in M dimensions. If M is substituted for the phase space dimension N in the previous discussions, all of the quantities that refer to attractors, such as information dimension and metric entropy, are well defined for infinite dimensional systems (see Refs. [31] and [32]).

In addressing the experimental determination of the dimension of a strange attractor, Takens [32] has expressed the fractal dimension and topological entropy in a manner that illustrates their conjugate relationship. As shown below, the information dimension and the metric entropy can be expressed in an analogous form. This form provides an algorithm that (in principle) can be used for the experimental determination of the information dimension and metric entropy of a dynamical system.

In experiments it is often the case that only a few variables, corresponding to only a few of the dimensions in the phase space of the system, can be measured. The experimenters are then faced with reconstructing qualitative properties of the attractor from a projection onto only a few dimensions. Fortunately, as demonstrated in Refs. [33], [34] and [35], a valid representation of an attractor can be constructed from a projection of a trajectory onto a single dimension. For example, if $v_x(r_0, t)$ is a single component of the velocity of a fluid at a fixed point r_0 in space, a set of phase variables $(v_1, v_2, \ldots, v_K)$ can be constructed.

$$\begin{aligned} v_1(t) &= v_x(r_0, t), \\ v_2(t) &= v_x(r_0, t + \tau_1), \\ v_3(t) &= v_x(r_0, t + \tau_2), \\ &\vdots \\ v_K(t) &= v_x(r_0, t + \tau_{K-1}), \end{aligned} \tag{45}$$

where $\tau_1, \ldots, \tau_{K-1}$ are suitably chosen delay times (see Ref. [35]). We will refer to coordinates constructed by taking delayed values of a projection onto a single dimension as delay coordinates. If M of these coordinates are taken, then, as shown in [32], any given state can be represented in terms of delay coordinates.

The representation of an attractor in terms of delay coordinates can be viewed as an attractor of a dynamical system that is related to other representations by a coordinate transformation. Providing this coordinate transformation is smooth, the attractors should have the same metric entropy. This relies on the invariance of the metric entropy; if the measure, the partition, and the flow are all transformed, the metric entropy remains the same [21]. The invariance properties of the information dimension are not currently known, but for the purpose of the following discussion we will assume that the information dimension is also invariant under smooth coordinate transformations.

Let $P_{m,n}$ be the probability distribution of sequences of length m for an n symbol partition of a single variable, and let $I(P_{m,n})$ be the information contained in $P_{m,n}$. Assume the sampled values used to construct $P_{m,n}$ are taken at time intervals Δt. If the delay times τ_i are picked so that $\tau_i = i\Delta t$, then a sequence of m symbols may be thought of as a set of measured values of an m dimensional representation in terms of delay coordinates. To determine the information dimension of the attractor, the limit of $I(P_{m,n})/|\log \varepsilon|$ must be taken as the scale of resolution goes to zero, which requires that the number n of elements in the partition go to infinity. Furthermore, the number of samples m must be large enough so that the delay representation is of sufficient dimension to faithfully represent states on the attractor. Although the embedding dimension M is in general unknown, a faithful representation can be accomplished by letting m get arbitrarily large. Once $m \geqq M$, adding further samples is redundant, and the limit should converge.

To calculate the metric entropy the limit as $m \to \infty$ must be taken for a partition and sampling rate that maximize the information rate. Intuitively, this rate should be approached by an observer with accurate instruments that induce a uniformly fine partition of the phase space containing the attractor. With this assumption, the information dimension and metric entropy of an attractor can be written in the following form:

$$D_I = \lim_{m\to\infty} \lim_{n\to\infty} \{I(P_{m,n})/\log n\}, \tag{46a}$$

$$h_\mu = \lim_{n\to\infty} \lim_{m\to\infty} \{I(P_{m,n})/m\}. \tag{46b}$$

Unfortunately, neither this algorithm, nor other algorithms previously presented in the literature, are practical to implement in experiments (a possible exception is the algorithm given in [35]). For attractors of dimension greater than three, the amount of data that needs to be gathered to achieve a reasonable level of resolution is prohibitive [35]. Future applications to physical experiments await more efficient algorithms, if they exist.

If the equations of motion are known, the information dimension is much easier to compute than the fractal dimension. Kaplan and Yorke define a quantity they call the Lyapunov dimension [36, 37] which can be expressed in terms of the spectrum of Lyapunov exponents (see Ref. [32], for example, for a definition of Lyapunov exponents). They have recently conjectured that the Lyapunov dimension is generally equal to the information dimension [9]. For any dynamical system that can be simulated efficiently it is feasible to compute the relevant part of the spectrum of Lyapunov exponents. The computational time and memory requirements increase linearly with the dimension of the attractor, rather than as the power of the dimension, as they do for a direct computation of the fractal dimension. In addition, the algorithms are relatively independent of resolution. Assuming that the Kaplan-Yorke conjecture is true, Ref. [33] contains examples of computations of the information dimension using this technique, for an infinite dimensional dynamical system with attractors of dimension as great as twenty (see also the comparisons of the Lyapunov dimension and the fractal dimension made in [38]).

Conclusions

Four distinct dimensions have been discussed in this paper. The topological dimension and fractal dimension are discussed in detail elsewhere [2, 3], and not much is said about them here. The embedding dimension is the minimum dimension for an embedding into Euclidean space; for an attractor, roughly speaking it is the minimum number of independent real numbers needed to smoothly label points everywhere on the attractor. The embedding dimension is important because it gives the minimum number of modes that are needed for a physical description. For the topological, fractal, and embedding dimensions it is sufficient to consider the attractor as a set (for the second two a metric is also needed); a notion of relative probability is not necessary.

Table 1. Dynamical quantities.

		(Need measure)	Relevant Probability Density
Dimensions	D_F	D_I	
(Need metric)	Fractal Dimension	Information Dimension	Asymptotic Probability Density
	$\lim_{\varepsilon-0} \frac{\log n}{\lvert\log \varepsilon\rvert}$	$\lim_{\varepsilon\to 0} \frac{I(\varepsilon)}{\lvert\log \varepsilon\rvert}$	$P(x)$
Entropies	h_j	h_μ	
	Topological Entropy	Metric Entropy	Symbol Sequence Distribution
	$D_F(P_\infty)\log n$	$D_I(P_\infty)\log n$	P_∞

The information dimension requires both a metric and a probability measure for its definition, and provides a geometrical way to understand the amount of information gained in sampling a random variable. It can thus be used to discuss some of the probabilistic properties of attractors. Knowledge of the information dimension of an attractor of a dynamical system allows an estimate to be made of the information gained in a measurement of an initial condition. For a continuous probability density defined on a smooth manifold, the information dimension is an integer, but for probability densities with structure that multiplies geometrically as the scale of resolution decreases, it is usually not an integer.

The information dimension fills a logical gap in the classification of chaotic attractors in terms of metric entropy, fractal dimension, and topological dimension. As summarized in Table 1, the information dimension plays the same role relative to the metric entropy that the fractal dimension does relative to the topological entropy. Information dimension and metric entropy are probabilistic notions, and require a measure for their definition. Topological dimension and fractal dimension do not require a measure. Both of the dimensions require a concept of distance (metric), whereas the entropies do not. Dimensions relate to isolated measurements (snapshots), whereas entropies relate to sequences of measurements (movies). The entropies can both be expressed in terms of the corresponding

dimension of the symbol sequence distribution mapped onto the interval.

Because it weights the probabilistic aspects inherent in any physical system, the information dimension has a more direct physical interpretation than the fractal dimension. This is similarly true for the metric entropy relative to the topological entropy. Another advantage of the information dimension is that when the equations of motion are known it is more feasible to compute than the fractal dimension.

Outstanding mathematical questions: Is the information dimension an invariant under smooth coordinate transformations (assuming the metric and measure are also transformed appropriately)? Does a uniform partition always provide the correct value of the information dimension in the limit where the elements of the partition are all very fine? If not, is there an algorithm to find partitions of a given diameter that minimize the information?

Acknowledgements

I would like to give special thanks to Norman Packard and Jim Crutchfield for their help in unravelling the relationship between the dynamical quantities of Table 1, and for contributing Figure 13. I am especially indebted to Jim Yorke for illuminating phone conversations and for providing the example of (15). I would also like to acknowledge valuable discussions with John Guckenheimer, Allan Kaufman, Joe Rudnick, David Rand, and Rob Shaw. I am indebted to Richard Kaplan and the Department of Aerospace Engineering at the University of Southern California, for their gracious hospitality and the use of their PDP 11/55 minicomputer. This study was supported by the Fanny and John Hertz Foundation.

Appendix I

Information Dimension of an Asymmetric Cantor Set

This appendix contains a computation of the information dimension of the asymmetric Cantor set made by deleting third fourths, shown in Figure 7. As described in Example (5) of Part I, begin the construction of the Cantor measure with a uniform probability density on the interval [0, 1].

Assume that the dimension can be computed using a uniform partition consisting of 2^i elements, where $i = 0, 1, 2, \ldots$. With $i = 1$, a two element partition, the total measure on the left side is $\frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} \ldots$; the total measure on the right is $\frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \ldots$; their ratio is 2:1. Letting P_j^i be the measure of the jth element of the ith partition, we get the following sequence of coarse grained probability distributions. Let n_i be the number of non-empty elements in the ith partition.

	n_i
$\{P_j^0\} = 1,$	1
$\{P_j^1\} = 1/3\ [2, 1],$	2
$\{P_j^2\} = 1/9\ [4, 2, 3],$	3
$\{P_j^3\} = 1/27\ [8, 4, 6, 6, 3],$	5
$\{P_j^4\} = 1/81\ [16, 8, 12, 12, 6, 12, 6, 9].$	8

Table A1.

Note that the n_i are the Fibonacci numbers, satisfying $n_i = n_{i-1} + n_{i-2}$. The P_j^i are given by the following recursion relation:

$$\text{For} \quad j = 1, 2, \ldots, n_{i-1}, \qquad P_j^i = 2/3\, P_j^{i-1},$$

$$\text{For} \quad j = n_{i-1} + 1, \ldots, n_i, \qquad P_j^i = 1/3\, P_{j-n_{i-1}}^{i-2}.$$

Let the information contained in the ith partition be I_i.

$$\begin{aligned} I_i &= -\sum_{j=1}^{n_i} P_j^i \log P_j^i \\ &= -\sum_{j=1}^{n_{i-1}} (2/3)\, P_j^{i-1} \log (2/3)\, P_j^{i-1} \\ &\quad -\sum_{j=1}^{n_{i-2}} (1/3)\, P_j^{i-2} \log \tfrac{1}{3} P_j^{i-2}. \end{aligned} \tag{A1}$$

A little algebra shows

$$I_i = (2/3)\, I_{i-1} + (1/3)\, I_{i-2} + H(1/3), \tag{A2}$$

where $H(p)$ is the binary entropy function, Equation (14). Let the ith approximation to the dimension be $D_i = I_i/\log 2^i = I_i/i \log 2$. Recast (2) in terms of D_i.

$$D_i = \left(\frac{1}{i \log 2}\right) \cdot \left[\frac{(i-2)}{3} D_{i-2} + \frac{2(i-1)}{3} D_{i-1} + H(\tfrac{1}{3})\right]. \tag{A3}$$

In the limit as i goes to infinitiv. D_i approaches a stable fixed point D_{I}.

$$D_{\text{I}} = \tfrac{3}{4} \frac{H(\frac{1}{3})}{\log 2}.$$

D_I is approximately equal to 0.6887, in contrast to $D_F = \log 3/\log 4 = 0.7925$.

Note that a random Cantor set formed by randomly deleting at each stage of the construction either the second fourth or the 3rd fourth of each continuous piece will have the same dimensions. Any Cantor set made this way will have the same sequence of coarse grained probability distributions as shown in Table 1, except that the order of the elements will be different. Since the information is insensitive to the order of the elements, the dimension will be the same.

Appendix II

In this Appendix I investigate some properties of the asymptotic probability density $\bar{P}(x)$ for chaotic attractors of one dimensional maps $x_{i+1} = F(x_i)$ with a single quadratic maximum. There are typically an infinite number of integrable singularities in $\bar{P}(x)$, which occur at the iterates of the critical point, and fall off on one side as $kx^{-1/2}$. The amplitude k decreases roughly as $e^{-j\lambda/2}$, where j is the order of iteration from the critical point, and λ is the Lyapunov exponent.

Assume throughout that the iterates of the critical point z_0 ($F'(z_0) = 0$) are not asymptotically periodic. This is the typical case for chaotic attractors; exceptions are, for example, $r = 4$ for the logistic equation. Assume that the asymptotic probability density can be computed using the Frobenius-Perron equation (4). Numerical computations by Rob Shaw and myself at several parameter values of the logistic equation support this assumption.

With the initial condition $P_0(x) = 1$, the jth iterate $P(x)$ of the Frobenius-Perron equation (eq. 10) will contain j singularities. The jth singularity occurs at the jth iterate of the critical point, which will be labeled z_j.

Let I_1 be a small interval of length ε_1 centered on z_1, and let I_j be the jth iterate of this interval, with length ε_j. For small ε_j

$$\begin{aligned}\varepsilon_j &= |F'(z_{j-1})\,F'(z_{j-2})\ldots F'(z_1)|\,\varepsilon_1 \\ &= |z_j'|\,\varepsilon_1\,, \end{aligned} \tag{A.4}$$

where $z_j' = [F^{j-1}(x)]'_{x=z_1}$. In the limit as $j \to \infty$, the Lyapunov exponent of z_1 is $\lambda_{z_1} = (\log|z_j'|)/j$. Quadratic mappings with a single maximum have only one attractor, and almost every point tends to this attractor. As long as the critical point is one of these typical points, the subscript can be dropped, and the Lyapunov exponent of almost every point written simply λ. Therefore $\varepsilon_j = \exp(j\lambda)\,\varepsilon_1$ for large j.

Conservation of probability implies that the total probability of I_j will be the sum of the probabilities of the inverse iterates of I_j:

$$\begin{aligned}\bar{\mu}(I_j) &= \int_{I_j} P_j(x)\,\mathrm{d}x \\ &= \int_{x \in F^{-j+1}(I_j)} P_1(x)\,\mathrm{d}x = \bar{\mu}(I_1)\,. \end{aligned} \tag{A.5}$$

The fact that F has a quadratic maximum, together with the initial condition $P_0(x) = 1$, implies that for $x \in I_1$ and I_1 sufficiently small

$$P_1(x) \simeq k_1(x - z_1)^{-1/2}\,. \tag{A.6}$$

Since I_1 is the only of the inverse iterates of I_j that contains a singularity, for ε_j small enough it dominates the contribution to the integrals in (A.5), so that for $x \in I_j$, $P_j(x) \simeq k_j(x - z_j)^{-1/2}$. Therefore, using (A.5) and (A.6),

$$\begin{aligned}\int_{I_j} P_j(x)\,\mathrm{d}x &= k_j(2\,\varepsilon_j)^{1/2} \\ &= \int_{I_j} P_1(x)\,\mathrm{d}x = k_1(2\,\varepsilon_1)^{1/2} \end{aligned} \tag{A.7}$$

and

$$k_j/k_1 = (\varepsilon_1/\varepsilon_j)^{1/2} = 1/|z_j|^{1/2} = e^{-j\lambda/2}\,.$$

Thus the amplitude k of the singularities falls off roughly exponentially at the rate given above. Notice that the larger the characteristic exponent, the faster the amplitude of the singularities decreases. This provides the motivation for the approximation for $\bar{P}(x)$ presented in (20).

[1] E. A. Abbot, Flatland: A Romance of Many Dimensions, Blackwell, Oxford 1888, 1962.
[2] W. Hurewicz and H. Wallman, Dimension Theory, Princeton 1948.
[3] B. Mandelbrot, Fractals: Form, Chance, and Dimension, Freeman, San Francisco 1977.
[4] J. Balatoni and A. Renyi, Publications of the Math. Inst. of the Hungarian Acad. of Sci. **1**, 9 (1956) (Hungarian). See English translation in "The Selected Papers of A. Renyi, **1**, p. 558, Akademia Budapest 1976.
[5] G. Bennetin, M. Casartelli, L. Galgani, A. Giogilli, and G. M. Strelcyn, Nuovo Cimento **44B**, 1883 (1978), and also **50B**, 211 (1979).
[6] R. Shaw, Z. Naturforsch. **36a**, 80 (1981).
[7] L. Brillouin, Science and Information Theory, Academic Press, New York 1956.
[8] A. Renyi, Acta Mathematica (Hungary) **10**, 193 (1959). See also A. Renyi, Probability Theory, North Holland, Amsterdam 1970.
[9] J. Yorke, private communication.

[10] M. Henon, Comm. Math. Phys. **53**, 69 (1976).
[11] C. Shannon, Bell Techn. **27**, 379 (1948).
[12] A. Kolmogorov, Dolk. Akad. Nauk SSSR **124**, 754 (1959). English summary in MR **21**, 2035.
[13] Ya. G. Sinai, Dolk. Akad. Nauk. SSSR **124**, 768 (1959). English summary in MR **21**, 2036.
[14] V. M. Alekseyev and M. V. Yakobson, Symbolic Dynamics and Hyperbolic Dynamic Systems, to appear in Physics Reports.
[15] R. Bowen, Equilibrium States and the Ergodic Theory of Anasov Diffeomorphisms, Springer-Verlag, Berlin 1975, 470.
[16] R. Bowen and D. Ruelle, Inventiones Math. **29**, 181 (1975).
[17] Ya. Sinai, Russ. Math. Surv. **166**, 21 (1972).
[18] Y. Oono, T. Kohda, and H. Yamazaki, J. Phys. Soc. Japan **48**, 3, 738 (1980).
[19] See, for example, N. Abramson, Information Theory and Coding, McGraw-Hill, New York 1963.
[20] V. I. Arnold and A. Avez, Ergodic Problems of Classical Mechanics, Benjamin, New York 1968
[21] P. Billingsley, Ergodic Theory and Information, John Wiley and Sons, New York 1965.
[22] R. Adler, A. Konheim, and M. McAndrew, Trans. Amer. Math. Soc. **114**, 309 (1965); MR **30**, 5291.
[23] R. Bowen, Trans. Amer. Math. Soc. **153**, 401 (1971).
[24] J. Guckenheimer, Comm. Math. Physics **70**, 133 (1979).
[25] I. Shimada, Prog. Theo. Phys. **62**, 61 (1979).
[26] A. Kaufman, H. Abarbanel, and C. Grebogi, Statistical Analysis of a Deterministic Stochastic Orbit, Proceedings of the Int. Conf. on Plasma Physics, Nagoya, Japan 1980.
[27] J. Crutchfield and N. Packard, UCSC preprint (1981).
[28] J. Curry, On the Entropy of the Henon Attractor, IHES preprint.
[29] S. Goldstein, Entropy Increase in Dynamical Systems, Rutgers U. Math. Dept. preprint (1980).
[30] D. Farmer, J. Crutchfield, H. Froehling, N. Packard, and R. Shaw, Ann. N.Y. Acad. Sci. **357**, 453 (1980).
[31] D. Farmer, Spectral Broadening of Period Doubling Sequences, Phys. Rev. Lett. **17**, 179 (1981).
[32] F. Takens, Detecting Strange Attractors in Turbulence, Rijksuniversiteit Groningen Mathematisch Institut, preprint (1980).
[33] D. Farmer, Chaotic Attractors of an Infinite Dimensional Dynamical System, Physica 4D (1982) 366–399.
[34] N Packard, J. Crutchfield, D. Farmer, and R. Shaw, Phys. Rev. Lett. **45**, 712 (1980).
[35] H. Froehling, J. Crutchfield, D. Farmer, N. Packard, and R. Shaw, On Determining the Dimension of Chaotic Flows, Physica 3D (1981) 605–617.
[36] J. Kaplan and J. Yorke, Functional Differential Equations and Approximation of Fixed Points. H. O. Peitgen and H. O. Walther, Eds., Springer-Verlag, Berlin 1979, p. 228, Bd. 570.
[37] P. Frederickson, J. Kaplan, and J. Yorke, The Dimension of the Strange Attractor for a Class of Difference Systems, U. of Maryland Math. Dept., to appear in J. Differential Eqs.
[38] D. Russel, J. Hanson and E. Ott, Phys. Rev. Lett. **45**, 1175 (1980).

Nachdruck — auch auszugsweise — nur mit schriftlicher Genehmigung des Verlages gestattet
Verantwortlich für den Inhalt: A. Klemm
Satz und Druck: Konrad Triltsch, Würzburg