J. Theiler

Estimating the Fractal Dimension of Chaotic Time Series

Fractals arise from a variety of sources: they have been observed in nature and on computer screens. An intriguing characteristic of fractals is that they can be described by noninteger dimensions. The geometry of fractals and the mathematics of fractal dimension provide useful tools for a variety of scientific disciplines—in particular, the study of chaos.

A chaotic dynamical system exhibits trajectories that converge to a strange attractor. The fractal dimension of this attractor counts the effective number of degrees of freedom in the dynamical system and thus quantifies the system's complexity. This article reviews the numerical methods that have been developed to estimate the dimension of a physical system directly from the system's observed behavior.

Introduction to Chaos

Chaos is a good thing, change is what comes of it.
—Septima Poinsette Clark

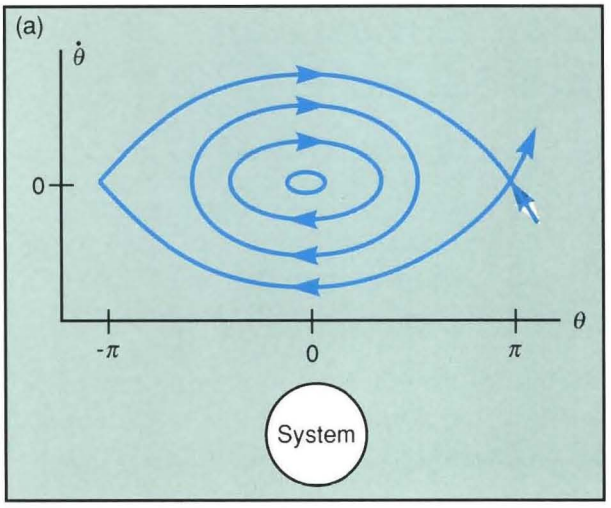
Chaos is the complicated behavior of simple deterministic equations, and complicated behavior is ubiquitous in nature. This notion is provocative, for it suggests that the irregular fluctuations exhibited in nature (or at least some of them) can be explained in simple terms. Indeed, researchers have observed chaotic behavior in systems ranging from a dripping faucet to the human body.

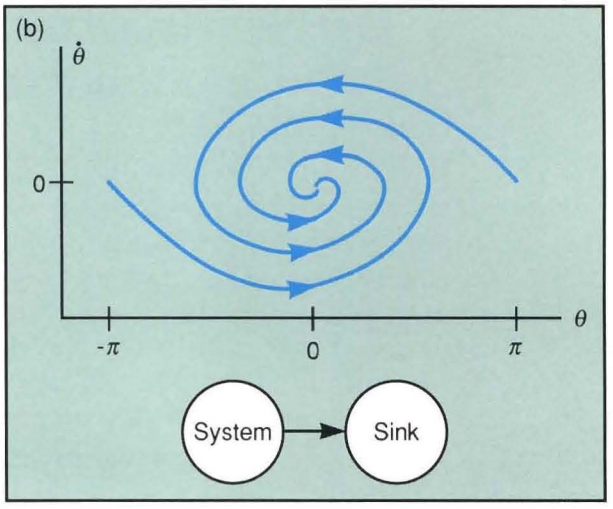
The study of chaos shows that even systems that obey simple laws can exhibit exotic and unpredictable behavior. The solution of the equations that describe the motion of a simple system can be an aperiodic trajectory for which long-term prediction is virtually impossible. If the system is dissipative, e.g., if friction is present, the trajectory converges to a subset of the system's *phase space*, which is the set of all instantaneous states available to the system. The subset might be a type of fractal that is referred to as a strange attractor. (Fractals are geometric forms with irregular patterns that repeat themselves at different scales. The forms consist of fragments of varying size and orientation but similar shape.) Although details of the trajectory have a sensitive dependence on the initial conditions, the geometrical structure of the strange attractor is robust. The fractal dimension of the strange attractor corresponds to the number of active degrees of freedom in the system.

Researchers have developed numerical methods for detecting and quantifying deterministic chaos. The algorithms first reconstruct the phase space directly from the observations and then estimate the fractal dimension of the observed trajectory. The application of the algorithms remains something of an art; practitioners do not agree fully on what constitutes a reliable estimate. Nonetheless, the promise of the algorithms is difficult to resist: by merely observing a single component of a potentially complex physical system, one can actually count the active degrees of freedom in the system. Thus, given an observation of irregular motion, it is possible to answer the question, is it chaos or is it noise?

Self-Organization, Dissipation, and Counting the Degrees of Freedom

One usually measures the complexity of a physical system by the number of degrees of freedom the system possesses. However, it is possible for the many nominal degrees of free-





dom that make up the complex system to be combined into a few effective degrees of freedom. This collective behavior, which in linear systems would be called normal modes, is more generally termed *self-organization*. Much of the interest in self-organization as a principle has been stimulated by the writings of Hermann Haken [1], an early and enthusiastic advocate. "A system is self-organizing," according to Haken, "if it acquires a spatial, temporal, or functional structure without specific interference from the outside." The effect appears in the laminar flow of a fluid, the spiral arms of a galaxy, or the regular ticking of a mechanical clock.

Self-organization is intriguing because thermodynamics seems to forbid it: a self-organizing system decreases its own entropy. For a closed system, the first and second laws of thermodynamics demand that energy be conserved, and that entropy not decrease. But for an open system, these thermodynamic laws do not apply, and the system can undergo self-organization.

Consider the pendulum. An ideal frictionless pendulum (Fig. 1[a]) swings with an amplitude and frequency that do not vary with time. However, the effect of an external perturbation is to put the system into another orbit and make the pendulum swing at a new amplitude and frequency. In this closed system, the long-term motion is dependent on the initial condition, and there is no *attractor* for the motion.

On the other hand, a pendulum with friction (Fig. 1[b]) does exhibit an attractor, albeit a trivial one. Independent of its initial condition, the pendulum will eventually come to rest. Perturbing the pendulum will start it swinging again, but the motion will eventually dampen out and the pendulum will stop. This final fixed-point motion is called an attractor because it is stable to perturbation.

Finally, consider a windup pendulum clock (Fig. 1[c]). Here the pendulum is both damped and driven. It will continue to swing at a fixed frequency (at least until the spring runs out) even if its motion is perturbed. A perturbation may affect the *phase* of the motion, but its amplitude and frequency will be altered only temporarily. The attractor is a *limit cycle*, and the stability of the limit cycle is what keeps the clock running at the correct rate.

The pendulum in the windup clock is an open system that receives energy from the spring and dissipates energy through friction. Thus energy is not conserved except as a long-term average. The pendulum qualifies as a self-organizing system because the property of swinging at a stable frequency is a function of the pendulum itself and does not depend on the details of the input energy source.

In general, self-organization arises in dissipative dynamical systems whose post-transient behavior involves fewer degrees of freedom than the full system. For the pendulum, there are two nominal degrees of freedom: angular position and angular momentum. But when the pendu-

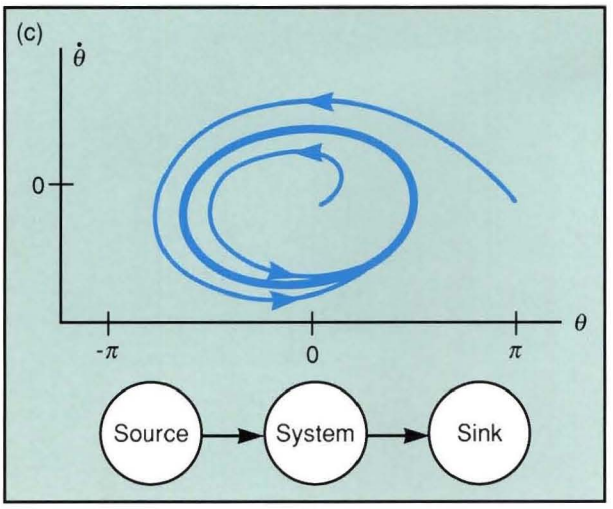


Fig. 1—Self-organized motion of the pendulum: (a) Orbits of an isolated frictionless pendulum are closed curves in phase space. The orbit of the pendulum's motion depends on the initial conditions, and the effect of an external perturbation is to put the system into a different orbit. (b) The inclusion of friction leads to a system that dissipates all of its energy. All orbits evolve toward a fixed-point attractor that corresponds to the pendulum hanging vertical and motionless. (c) The inclusion of both friction and a driving force such as that which might be applied by a windup spring leads to a system with a limit-cycle attractor. Even in the presence of external perturbations, the system tends toward an orbit on the limit cycle. Thus the amplitude and frequency are intrinsic to the system, and they depend neither on the initial conditions nor on the details of the perturbations.

lum is in its stable limit cycle, only one degree of freedom—the phase angle—is required to describe the state of the system. The system is *attracted* to a lower-dimensional phase space, and the dimension of this reduced phase space is the number of active degrees of freedom in the self-organized system.

Attractors are not confined to fixed points and limit cycles. For nonlinear systems with three or more degrees of freedom, *strange attractors* may arise. (If the nonlinear system is a discrete map, then only two degrees of freedom are needed. Figure 2 shows the strange attractors obtained from the Ikeda map [2] and the Hénon map [3].) Motion on such an attractor can be highly irregular and essentially unpredictable.

Even systems that are nominally complex (i.e., systems that have many available degrees of freedom) may relax to a low-dimensional attractor. Behavior that is irregular can be low-dimensional or it can be essentially stochastic. Distinguishing between the two types of irregular behavior is the motivation for quantifying chaos. The estimation of dimension from a time series is one way to detect and quantify the self-organizational properties of natural and artificial complex systems.

Nonlinear Dynamical Systems

Strange attractors arise from nonlinear dynamical systems. Physically, a dynamical system is anything that moves. (And if it doesn't move, then it's a dynamical system at a fixed point.) Mathematically, a dynamical system is defined by a *state space* (also called phase space) $\mathbf{R}^M$ that describes the instantaneous states available to the system, and an *evolution operator* ϕ that tells how the state of the system changes with time. M is the number of degrees of freedom in the system; ϕ can be thought of as the physics of the system. An element $\mathbf{X} \in \mathbf{R}^M$ specifies the current state of the system. For a pendulum, $\mathbf{X}$ might represent the two coordinates of angular position and angular momentum. If the pendulum contains a windup spring, $\mathbf{X}$ might also contain a coordinate for the instantaneous torque being applied by the spring. The evolution operator is a family of functions $\phi_t: \mathbf{R}^M \rightarrow \mathbf{R}^M$ that map the current state of the system into its future state at a time t units later. The operator ϕ_t satisfies $\phi_0(\mathbf{X}) = \mathbf{X}$ and $\phi_{t+s}(\mathbf{X}) = \phi_t(\phi_s(\mathbf{X}))$. A dynamical system is *nonlinear* if in general $\phi_t(c_1\mathbf{X}_1 + c_2\mathbf{X}_2) \neq c_1\phi_t(\mathbf{X}_1) + c_2\phi_t(\mathbf{X}_2)$, where c_1 and c_2 are constants.

The function ϕ_t can be defined either as a discrete map or in terms of a set of ordinary differential equations. Researchers have also studied partial differential equations and differential delay equations. (In these last two cases, M is infinite).

Dissipative Dynamics

A system is dissipative if the volume of a reference chunk of phase space tends to zero as $t \rightarrow \infty$ (Fig. 3). In other words, if $\mathbf{B}$ is a bounded subset of $\mathbf{R}^M$ with ordinary (Lebesque) volume $V(\mathbf{B})$, then

$$\lim_{t \rightarrow \infty} V(\phi_t(\mathbf{B})) = 0. \qquad (1)$$

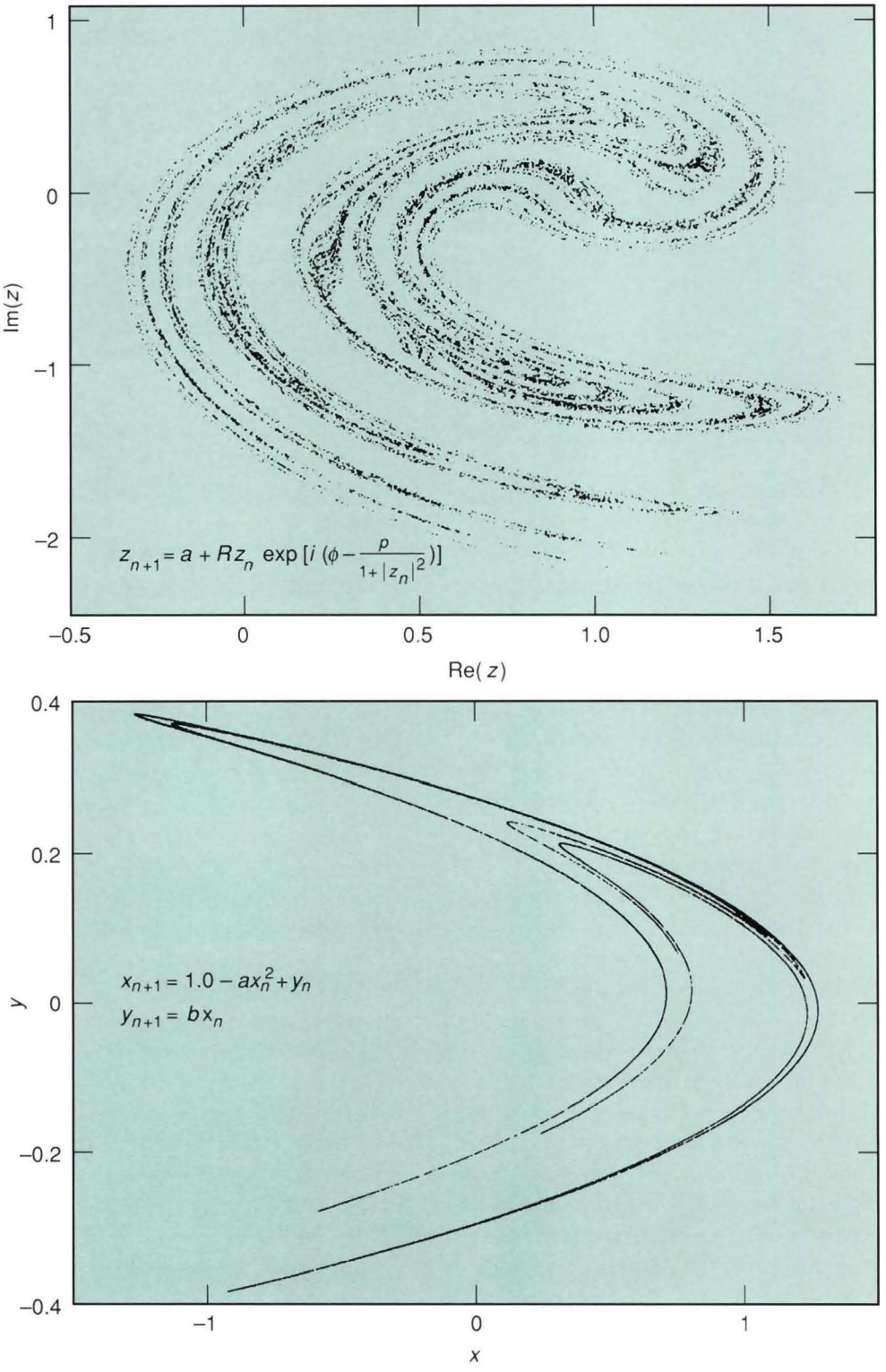


Fig. 2—Two strange attractors: (top) The Ikeda map is a complex map that derives from a model of the plane-wave interactivity field in an optical ring laser [2]. The map is iterated many times, and the points ($Re(z_n)$, $Im(z_n)$) are plotted for $n \geq 1000$, *where* $Re(z_n)$ *represents the real part of the complex number* z_n *and* $Im(z_n)$ *represents the imaginary part. Here* $a = 1.0$, $R = 0.9$, $\phi = 0.4$, *and* $p = 6$. *The fractal dimension of this attractor is approximately 1.7. (bottom) The Hénon map [3] with* $a = 1.4$ *and* $b = 0.3$ *gives an attractor that has a fractal dimension of about 1.3. Note how much thinner the attractor of lower dimension appears.*

Usually the dissipation is exponential, at least on the average for large t; i.e., $V(\phi_t(\mathbf{B})) \sim V(\mathbf{B})e^{-\Lambda t}$, where Λ is the rate of dissipation. In a conservative system, the volume is constant with time. Although conservative dynamical systems can also exhibit chaos, only dissipative dynamical systems have strange attractors.

The *attractor* **A** of a dynamical system is the subset of phase space toward which the system evolves. Informally, one might write $\phi_t(\mathbf{B}) \to \mathbf{A}$ as $t \to \infty$. An initial condition $\mathbf{X}_0$ that is sufficiently near the attractor will evolve in time so that $\phi_t(\mathbf{X}_0)$ comes arbitrarily close to the set **A** as $t \to \infty$. Since the volume $V(\phi_t(\mathbf{B}))$ goes to zero, the attractor is a zero-volume set. It is possible, however, for a small volume to grow longer even as it is growing thinner: the volume converges to zero but the attractor is not just a single fixed point (Fig. 3[c]). In fact, the set **A** can be a fractal and, in that case, the attractor is said to be strange.

The Natural Invariant Measure

Equation 1 implies that $V(\mathbf{A})$, the phase-space *volume* of an attractor, is zero for a dissipative system. To quantify the dynamics of an attractor requires first the introduction of a new measure μ that is concentrated on the attractor. The measure μ is defined with respect to the attractor **A** such that subsets **B** of the state space $\mathbf{R}^M$ are associated with real values $\mu(\mathbf{B})$ that represent how much of the set **A** is contained in **B**. The measure μ reflects the varying density over the set **A**, and $\mu(\mathbf{B})$ can be intuitively regarded as the mass of **B**.

A particularly useful measure μ is one that is invariant to the dynamics of the system. The proper way to define this is to write

$$\mu(\mathbf{B}) = \mu(\phi_{-t}(\mathbf{B})), \qquad (2)$$

where $\phi_{-t}(\mathbf{B}) \equiv \{\mathbf{X} \in \mathbf{R}^M : \phi_t(\mathbf{X}) \in \mathbf{B}\}$. In general, Eq. 2 is not enough to define a unique invariant measure for a dynamical system. For instance, a measure that is concentrated on an unstable fixed point satisfies Eq. 2 but has little to do with the typical post-transient motion of the system.

The physically relevant measure for a dynamical attractor counts how often and for how long a *typical* trajectory visits various parts of the set (Fig. 4). Here

$$\mu(\mathbf{B}) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathrm{I}_{\mathbf{B}}(\phi_t(\mathbf{X}_0))\,dt, \qquad (3)$$

where $\mathbf{X}_0$ is a typical initial condition, T is the total time length of the trajectory, and $\mathrm{I}_{\mathbf{B}}(\mathbf{X})$ is the indicator function for **B**: it is unity if $\mathbf{X} \in \mathbf{B}$ and zero otherwise. For virtually all initial conditions $\mathbf{X}_0$, Eq. 3 defines a unique invariant measure on the attractor that is called the *natural invariant measure*.

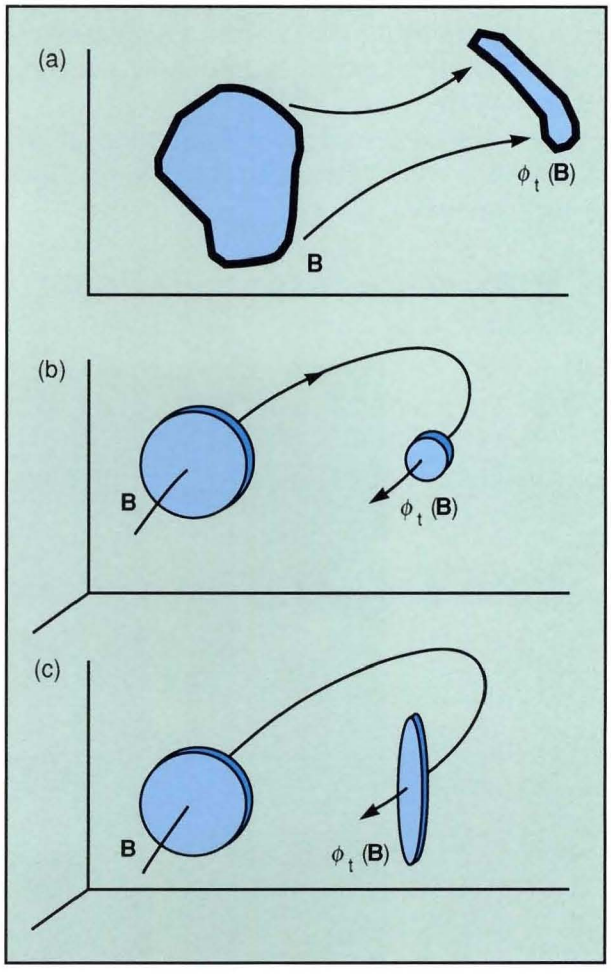


*Fig. 3—Phase-space contraction: (a) The volume V of a reference-chunk **B** of phase space decreases as the system evolves. (b) Contraction of phase space in directions perpendicular to the motion leads to limit cycles and other non-strange attractors. (c) Contraction of phase space in some directions with expansion in other directions in such a way that the phase-space volume is decreasing leads to a chaotic attractor.*

Sensitivity to Initial Conditions

The hallmark of a chaotic system is the sensitivity of the system's individual trajectories to their initial conditions. (The overall structure of the attractor, by contrast, is robust with respect to changes in initial conditions.) The sensitivity is usually quantified in terms of the Lyapunov exponents and the Kolmogorov entropy [4]. The Lyapunov exponents measure the rate of exponential divergence of nearby trajectories and the Kolmogorov entropy measures the rate of information flow in the dynamical system.

Long-term predictions of chaotic systems are virtually impossible. Even if the physics (i.e., ϕ) of a chaotic system is known completely, errors in measuring the initial state propagate exponentially.

Consider two nearby initial conditions $\mathbf{X}_0$ and $\mathbf{X}_0 + \varepsilon$ and evolve both forward in time. A Taylor series expansion gives

$$\phi_t(\mathbf{X}_0 + \varepsilon) = \phi_t(\mathbf{X}_0) + \mathbf{J}(t) \cdot \varepsilon + \mathrm{O}(\varepsilon^2)$$

where the symbol O denotes the order of the correction factor and $\mathbf{J}(t)$ is the Jacobian matrix given by the linearization of ϕ around the point $\mathbf{X}_0$:

$$\mathbf{J}(t) = \frac{\partial \phi_t(\mathbf{X}_0)}{\partial \mathbf{X}_0} = \frac{\partial \mathbf{X}(t)}{\partial \mathbf{X}(0)} .$$

The i, j element of the matrix is

$$J_{i,j}(t) = \frac{\partial X_i(t)}{\partial X_j(0)}$$

where $X_i(t)$ is the ith component of the state vector $\mathbf{X}$ at time t. The determinant of the Jacobian matrix describes the overall contraction of phase-space volume (i.e., the dissipation in the system), and the eigenvalues describe the divergence of nearby trajectories. The Lyapunov exponents (λ) quantify the average rate of expansion of these eigenvalues.

$$\lambda_n = \lim_{t \to \infty} \left(\frac{1}{t} \log |n\text{th eigenvalue of } \mathbf{J}(t)| \right).$$

By convention, the Lyapunov exponents are

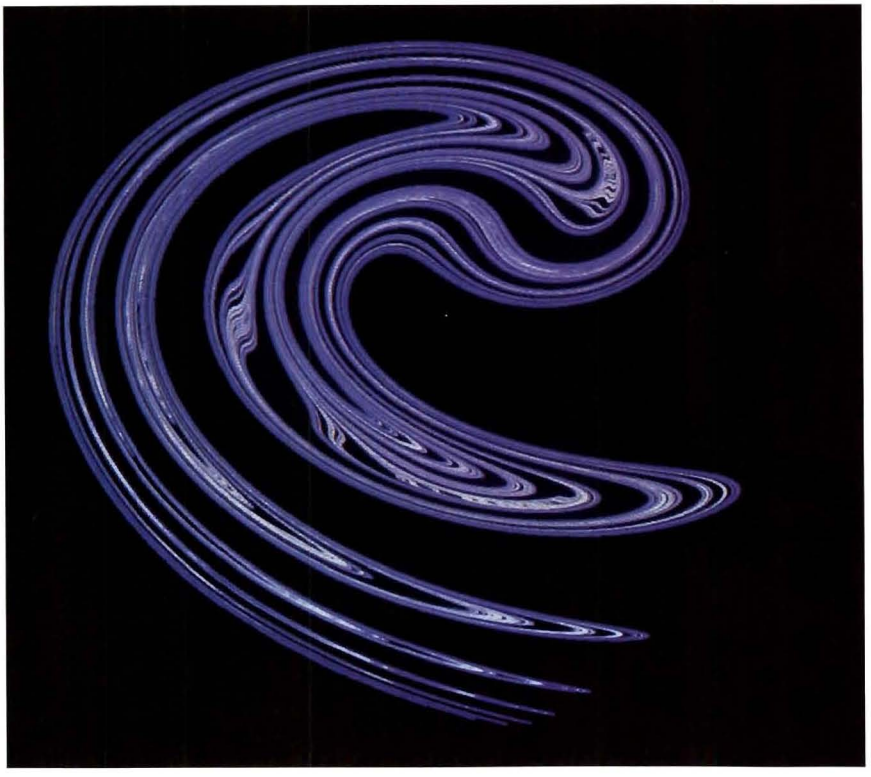


Fig. 4—Invariant measure of the Ikeda attractor. The color of the pixel indicates how often a typical orbit visits the site, from many visits (white) to few (blue) to none (black).

indexed in descending order: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$. Unless the vector that connects a pair of initial conditions happens to be precisely orthogonal to the eigenvector that is associated with the largest eigenvalue of the Jacobian matrix, then the pair will separate at a rate dominated by the largest Lyapunov exponent. And if the largest Lyapunov exponent is positive, then chaotic motion is assured. (Otherwise, trajectories will collapse to a fixed point, a limit cycle, or a limit torus.) The sum of all the exponents is negative for a dissipative dynamical system and defines the rate of phase-space contraction: $\Lambda = -(\lambda_1 + \ldots + \lambda_M)$.

If the state $\mathbf{X}(t)$ of a system is known at time t to some accuracy, then the future can be predicted by $\mathbf{X}(t + \Delta t) = \phi_{\Delta t}(\mathbf{X}(t))$ but the predicted value will usually be less accurate. Therefore, to observe or measure the state of the system again at time $t + \Delta t$ to the original accuracy is to learn information that was previously unavailable. In this respect, the system appears to be constantly creating information. The average rate at which it does so is quantified by the Kolmogorov entropy, which is equal to the sum of the positive Lyapunov exponents [4].

Delay-Time Embedding

Confronted with a physical system, an experimentalist measures at regular and discrete intervals of time the value of some state variable (e.g., voltage) and records the time series: $x(t_0)$, $x(t_1)$, $x(t_2)$, . . . , with $x(t_i) \in \mathbf{R}$ and $t_i = t_0 + i\Delta t$. From this time series, the experimentalist attempts to infer something about the dynamics (i.e., the physics) of the system. The measurement $x(t)$ represents a projection π from the full state vector $\mathbf{X}(t) \in \mathbf{R}^M$.

$$\pi: \mathbf{R}^M \rightarrow \mathbf{R}.$$

Because a time series is one-dimensional, it is an incomplete description of a system during a time evolution. Nonetheless, many properties of the system can still be inferred from the time series.

Packard, Crutchfield, Farmer, and Shaw [5] devised an elegant and remarkably simple scheme to reconstruct the state space by embedding the time series into a higher-dimensional space. From time-delayed values of the scalar time series, vectors $\hat{\mathbf{X}} \in \mathbf{R}^m$ are created:

$$\hat{\mathbf{X}}(t) = \left[x(t), x(t - \tau), \ldots, x(t - (m - 1)\tau)\right]^{\mathrm{T}}$$

where τ (the delay time) and m (the embedding dimension) are parameters of the embedding procedure. (The superscript T denotes the transpose and the symbol $\hat{}$ denotes the reconstruction of a vector.) Here $\hat{\mathbf{X}}(t)$ represents a more complete description of the state of the system at time t than does $x(t)$ and can be thought of as a mapping from the full state $\mathbf{X}(t)$ to the reconstructed state $\hat{\mathbf{X}}(t) = \pi^{(m)}(\mathbf{X}(t))$:

$$\pi^{(m)}: \mathbf{R}^M \rightarrow \mathbf{R}^m.$$

An embedding dimension of $m > 2D + 1$, where D is the fractal dimension of the attractor, almost always ensures the reconstruction of the topology of the original attractor. That is, $\pi^{(m)}$ restricted to $\mathbf{A}$ is a smooth one-to-one map from the original attractor to the reconstructed attractor. It should be noted, however, that as long as $m > D$, the reconstructed set will almost always have the same dimension as the attractor—an assumption that is usually sufficient for the purposes of dimension estimation. Figure 5 shows that with delay-time embedding, the mapping from the actual phase space to a reconstructed phase space is generically smooth.

Delay-time embedding is a powerful tool. For example, if one believes that the brain is a deterministic system, then it might be possible to study the brain by looking at the electrical output of a single neuron. This is an ambitious example, but the point is that the delay-time embedding makes it possible to analyze the self-organizing behavior of a complex dynamical system without knowledge of the system's *full* state at any given time.

In principle, almost any delay time τ and embedding dimension $m > D$ will work (i.e., if unlimited infinitely precise data are available). However, choosing the optimal parameters for real data is a nontrivial process. For instance, if the product $(m - 1)\tau$ is too large, then the components $x(t)$ and $x(t + (m - 1)\tau)$ of the reconstructed vector $\hat{\mathbf{X}}$ will be effectively decorrelated, which will inflate the estimated dimension. On

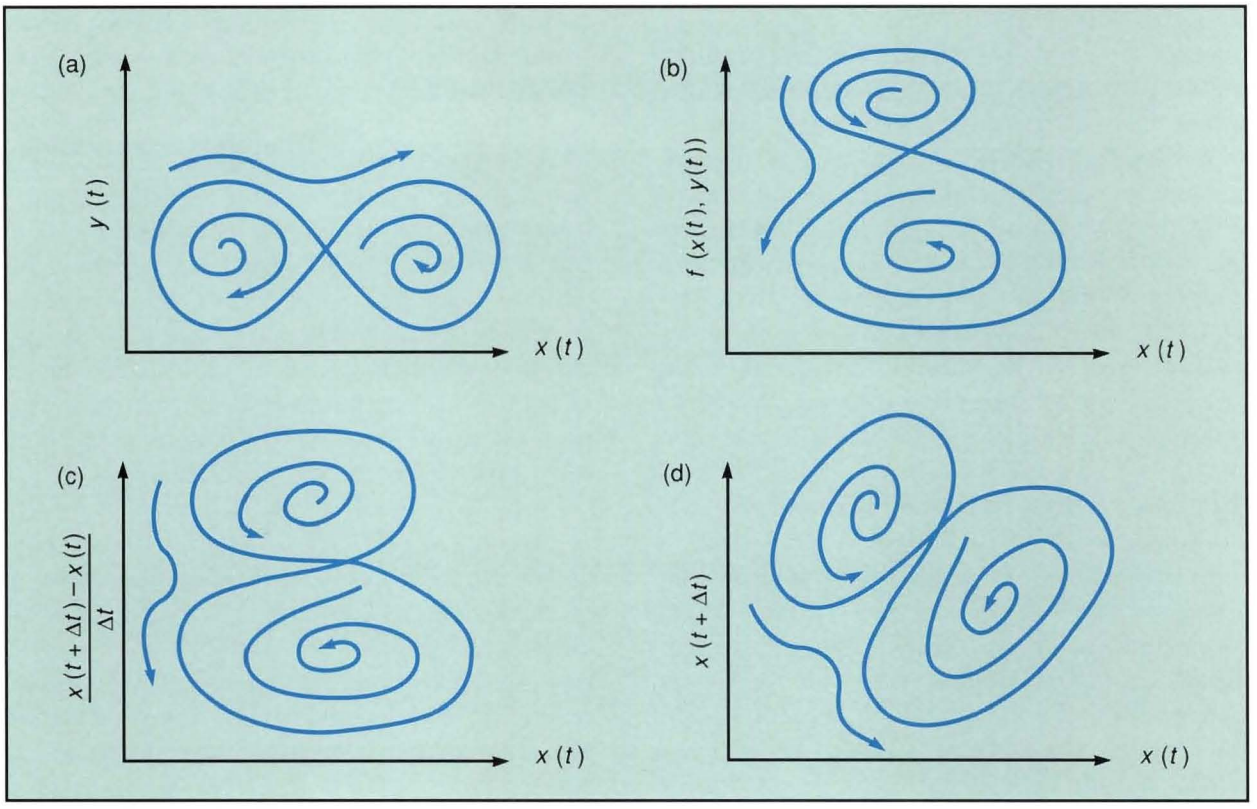


Fig. 5—Delay-time embedding. A plausibility argument illustrates that the mapping from the original phase space to a reconstructed phase space using the delay-time embedding is generically smooth. The dynamics will be assumed to be given by the differential equations $\dot{x} = f(x, y)$ *and* $\dot{y} = g(x, y)$ *where* f *and* g *are smooth functions of the actual phase-space variables* x *and* y. *Note that a succession of smooth transformations leads to the delay-time coordinates: (a) A plot of* x(t) *versus* y(t) *gives the original phase space of the system. (b) A plot of* x *versus* f(x, y) *is clearly a smooth transformation of coordinates, since* f *is a smooth function. Because* $\dot{x} = f(x, y)$, *this plot is equivalent to a plot of* x *versus* $\dot{x}$. *Furthermore, one can write* $\dot{x} = \lim_{\Delta t \to 0} [x(t + \Delta t) - x(t)]/\Delta t$. *(c) Since* x(t) *is a smooth function of* t, *it is plausible to assume that the function* $[x(t + \Delta t) - x(t)]/\Delta t$ *is a smooth function of* Δt. *Therefore, making* Δt *finite is a smooth transformation from the system in part b. (d) A linear change of coordinates from those in part c leads to the delay-time variables* x(t) *versus* $x(t + \Delta t)$. *Note that the reconstructed phase space of part d is just a smooth transformation of the original phase space in part a.*

the other hand, if $(m-1)\tau$ is too small, then the components $x(t), \ldots, x(t+(m-1)\tau)$ will all be very nearly equal and the reconstructed attractor will look like one long diagonal line. It is also inefficient to make τ very small, even for large m, because if $x(t) \cong x(t+\tau)$, successive components of $\hat{\mathbf{X}}$ become effectively redundant.

Generally speaking, one wants both τ not too much less than and $(m-1)\tau$ not too much greater than some characteristic decorrelation time. One such characteristic time is defined in terms of the autocorrelation function $A(\tau) = \langle (x(t) - \langle x \rangle)(x(t+\tau) - \langle x \rangle) \rangle$, where the angle brackets $\langle\ \rangle$ represent an average over time t. The autocorrelation time is given by $\int_{-\infty}^{\infty} A(\tau)/A(0) d\tau$. Fraser and Swinney [6] used mutual information in place of autocorrelation to provide a more sophisticated (and in many cases better) choice of characteristic time.

Typically, having chosen τ, one performs the dimension analysis for increasing values of m and looks for a plateau in the plot of D versus m.

Fractals and Fractal Dimension

Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line.

—Benoit B. Mandelbrot

The modern study of fractals originated with Mandelbrot [7], who coined the term and who remains an ardent spokesman.

Fractals are crinkly objects that defy conventional measures like length and area. Yet fractals are beguilingly far from formless. Clouds, mountains, coastlines, bark, and lightning bolts all exhibit nonsmooth shapes that can be described as fractal. Fractal objects, as Leo Kadanoff notes in Ref. 8, "contain structures nested within one another like Chinese boxes or Russian dolls." This self-similar structure is perhaps the main reason for the striking visual beauty of fractals. Self-similarity also implies a scale-invariant property. There are crinkles upon crinkles, but no preferred crinkle size.

Mathematically speaking, a set is strictly self-similar if it can be expressed as a union of sets each of which is a reduced copy of (i.e., is geometrically similar to) the full set. However, not all fractal objects exhibit this precise form. In a coastline, for instance, there is an irregular nesting of gulfs, bays, harbors, and coves that are observed over a broad range of spatial scales (Fig. 6). A magnified view of one part of the coastline may not precisely reproduce the full picture, but it will have the same qualitative appearance. A coastline displays the kind of fractal behavior that is called statistical self-similarity.

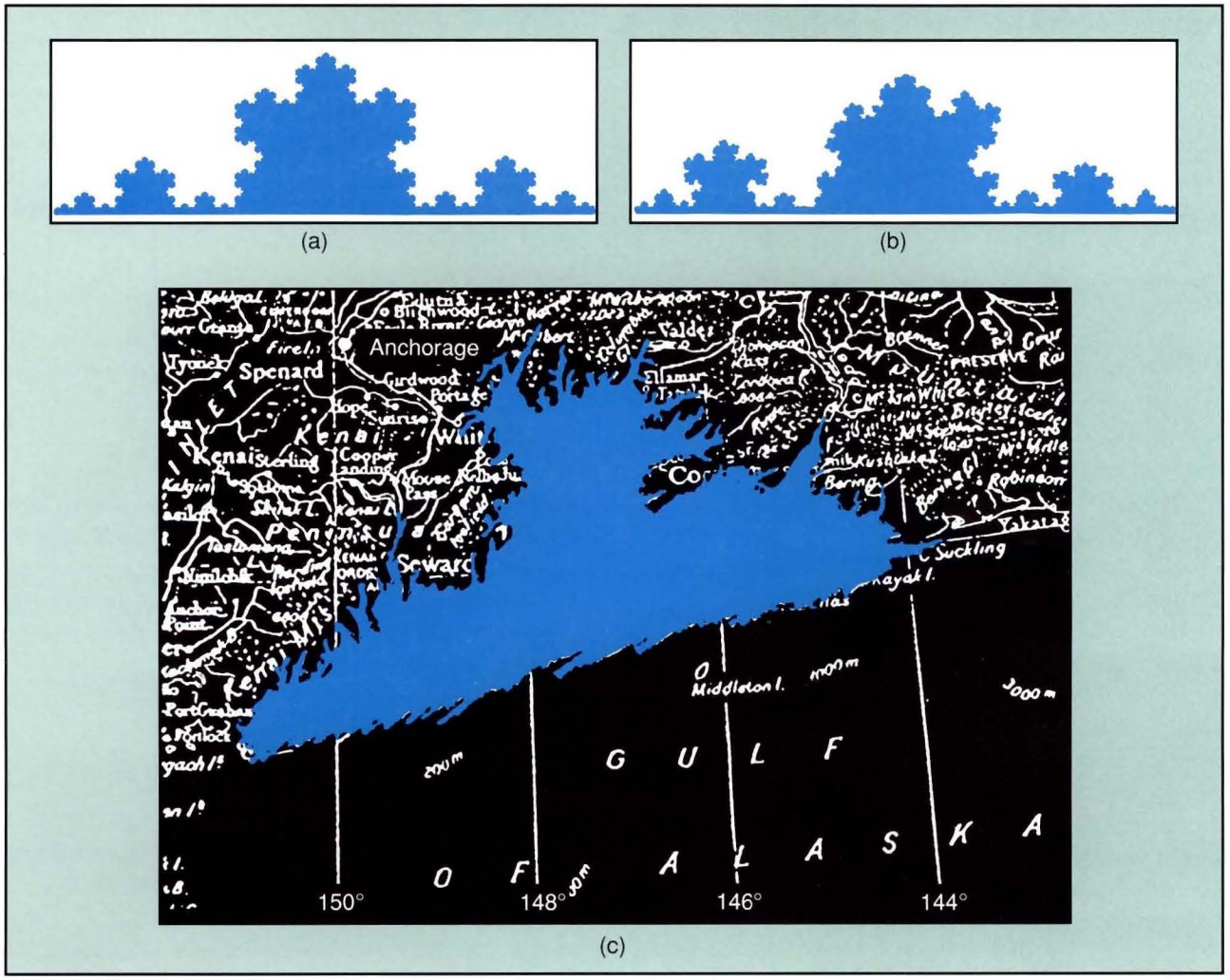


Fig. 6—Self-similar fractals: (a) The Koch curve is a fractal curve, each segment of which is a smaller copy of the full figure. (b) This variant of the Koch curve includes an element of randomness. Here, although no one segment is precisely similar to any other segment, there is a statistically *self-similar structure to the figure. (c) Alaska's Prince William Sound has been colored to highlight the fractal appearance of the coastline.*

Quantifying Fractals

The most natural quantity by which a fractal can be characterized is its fractal dimension. One motivation is described by H. Eugene Stanley [9], who has outlined a program for the practicing scientist who wants to study fractals.

> If you are an experimentalist, you try to measure the fractal dimension of things in nature. If you are a theorist, you try to calculate the fractal dimension of models chosen to describe experimental situations; if there is no agreement then you try another model.

This pithy advice applies to strange attractors as well as to solid fractal objects: fractal dimension provides a benchmark against which theory can be compared to experiment. In the case of strange attractors, however, there is further reason for wanting to know the dimension. The dimension is what counts the number of degrees of freedom in a system, and thus provides a useful measure of the system's complexity.

Definitions of Fractal Dimension

In 1919, Felix Hausdorff [10] gave a completely rigorous definition of dimension, but it is a definition that does not lend itself to numerical estimation. Much interest in the last decade has focused on the numerical study of chaos, and it is useful to consider more operational definitions of dimension, i.e., those definitions which can be translated into algorithms for *estimating* dimension from a *finite* sample of points.

In the following sections, two different ways of thinking about dimension will be discussed: first as an exponent that characterizes the scaling of a bulk with a size, then as a number that counts how many active degrees of freedom a system contains. Both definitions give the same number and both can be extended to define a generalized dimension, but each attempts to provide a different intuition for understanding what that dimension means.

Local Scaling of Bulk with Size

The geometrically intuitive notion of dimension is as an exponent that expresses the scaling of an object's bulk with its size.

$$\text{bulk} \sim \text{size}^{\text{dimension}} . \qquad (4)$$

Here the term bulk can correspond to a volume, a mass, or even a measure of information content, and the term size is a linear distance. For a line, bulk and size are the same thing, and so a line has a dimension of one. The area of a plane figure grows quadratically with its size, and so it has a dimension of two. The definition of dimension is usually cast as an equation of the form

$$\text{dimension} = \lim_{\text{size}\to 0} \frac{\log(\text{bulk})}{\log(\text{size})}$$

where the limit of size $\to 0$ is used to ensure invariance over smooth coordinate changes. The small-size limit also implies that dimension is a local quantity, and that any global definition of fractal dimension will require some kind of averaging.

Although other notions of bulk are also possible, the obvious relevant measure of bulk for a subset $\mathbf{B}$ of a dynamical attractor is the attractor's natural invariant measure $\mu(\mathbf{B})$.

A good quantity for the size of a set is its radius or its diameter, the latter of which is defined by

$$\delta(\mathbf{B}) \equiv \sup\{\|\mathbf{X} - \mathbf{Y}\|: \mathbf{X}, \mathbf{Y} \in \mathbf{B}\}$$

where sup is the supremum, or maximum, and $\|\mathbf{X} - \mathbf{Y}\|$ is the distance between points $\mathbf{X}$ and $\mathbf{Y}$. How this distance is calculated depends on the *norm* of the embedding space. If X_i is the ith component of the vector $\mathbf{X} \in \mathbf{R}^m$, then the L_s norm gives distance according to

$$\|\mathbf{X} - \mathbf{Y}\| \equiv \left(\sum_{i=1}^{m} |X_i - Y_i|^s \right)^{1/s} .$$

The most useful of these norms are L_2, the Euclidean norm, which gives distances that are rotation invariant; L_1, called the taxicab norm, which is very easy to compute; and L_∞, the maximum norm, which is also easy to compute. It is not difficult to show that fractal dimension is invariant to choice of norm.

For instance, the pointwise dimension is a local measure of the dimension of a fractal set *at*

a point on the attractor. Let $\mathbf{B}_{\mathbf{X}}(r)$ denote a ball of radius r centered at the point **X.** Define the *pointwise mass function* $B_{\mathbf{X}}(r)$ as the natural invariant measure

$$B_{\mathbf{X}}(r) = \mu(\mathbf{B}_{\mathbf{X}}(r)). \tag{5}$$

The scaling of the mass function at **X** with the radius r defines $D_p(\mathbf{X})$, the pointwise dimension at **X:**

$$D_p(\mathbf{X}) = \lim_{r \to 0} \frac{\log B_{\mathbf{X}}(r)}{\log r}.$$

$D_p(\mathbf{X})$ is a local quantity, but one can define the average pointwise dimension to be a global quantity.

$$\begin{aligned} D_p &= \int_{\mathbf{A}} D_p(\mathbf{X}) \mathrm{d}\mu(\mathbf{X}) \\ &= \lim_{r \to 0} \frac{\log H(r)}{\log r} \end{aligned} \tag{6}$$

where $H(r) = \exp\{\int_{\mathbf{A}} \log [B_{\mathbf{X}}(r)] \mathrm{d}\mu(\mathbf{X})\}$ is a weighted geometric average of the pointwise mass function $B_{\mathbf{X}}(r)$.

Information Dimension

As an alternative to the scaling of mass with size, the dimension of a set can be thought of in terms of how many real numbers are needed to specify a point on that set. For instance, the position of a point on a line can be labeled by a single real number, the position on a plane by two Cartesian coordinates, and the position in ordinary three-dimensional space by three coordinates. Here dimension is something that counts the number of degrees of freedom. For sets more complicated than lines, surfaces, and volumes, however, this informal definition of dimension needs to be broadened.

One way to extend the definition is to determine not how many real numbers but how many bits of information are needed to specify a point to a given accuracy. On a line segment of unit length, k bits divide the segment into 2^k subsegments, and therefore specify the position of a point to within an accuracy $r = 2^{-k}$. For a unit square, $2k$ bits are needed to achieve the same accuracy (k bits for each of the two coordinates specified). And similarly, $3k$ bits are needed for a three-dimensional cube. For the general case, $S(r)$ is equal to $-d \log_2 r$, where $S(r)$ is the number of bits of information that are needed to specify a position on a unit d-dimensional hypercube to an accuracy r. Solving the equation for d leads to a natural definition for the *information dimension* (D_I) of a set.

$$D_I = \lim_{r \to 0} \frac{-S(r)}{\log_2 r}. \tag{7}$$

By identifying the bulk of Eq. 4 with the expression $2^{-S(r)}$ in Eq. 7, one can still interpret D_I as a scaling of bulk with size.

If a fractal is partitioned into boxes B_i of size r, then determining the position of a point to an accuracy r requires specifying which box the point is in. The average information needed to specify one box is given by Shannon's formula

$$S(r) = -\sum_i P_i \log_2 P_i \tag{8}$$

where P_i is the probability measure of the ith box: $P_i = \mu(B_i)/\mu(\mathbf{A})$. Equation 8 can be substituted into the definition of D_I (Eq. 7):

$$D_I = \lim_{r \to 0} \frac{\sum_i P_i \log_2 P_i}{\log_2 r}.$$

The equivalence of D_I (the information dimension) and D_p (the average pointwise dimension) can be seen by treating the sum $\sum_i P_i \log_2 P_i$ as a weighted average of $\log P_i$, and comparing this quantity to the weighted average of $\log B_{\mathbf{X}}(r)$ in Eq. 6.

Box-Counting Dimension

Historically, the first numerical estimates of dimension were based on partitioning the state space into a grid of boxes, each of size r, and counting how many boxes contained points. The scaling of this number with r defines an upper bound on the Hausdorff dimension, which is referred to as the capacity, the box-counting dimension, or the fractal dimension. The last term, however, has come to be used in a generic sense for any dimension that might be nonintegral. With the grid size r and the count of

nonempty boxes $n(r)$, the box-counting dimension D_H is given by

$$D_H = \lim_{r \to 0} \frac{\log (1/n(r))}{\log r}.$$

Here the local notion of bulk is replaced with a global one: because each nonempty box contains on average $1/n(r)$ of the whole fractal, the quantity $1/n(r)$ is the average bulk of each nonempty box.

Though the box-counting dimension is not necessarily equal to the information dimension, the difference is usually small, so that for the purposes of estimating the complexity of a physical system it may not really matter which definition of dimension is used. However, precise comparisons of theory and experiment require that both are referring to the same quantity. The discrepancy will be resolved by extending the definitions to a generalized dimension.

Generalized Dimension

In computing the box-counting dimension, one makes the decision to count or not count a box according to whether there is at least one point in the box. No provision is made for weighting the box count according to how many points are inside the box.

It is possible to generalize the definition of dimension in a way that takes into account the number of points in the box. Let B_i denote the ith box and let $P_i = \mu(B_i)/\mu(\mathbf{A})$ be the normalized measure of this box. Equivalently, P_i is the probability that B_i contains a randomly chosen point on the attractor and the quantity is estimated by counting the number of points that are in the ith box and dividing by the total number of points.

The generalized dimension D_q can be defined by

$$D_q = \frac{1}{q-1} \lim_{r \to 0} \frac{\log \sum_i P_i^q}{\log r} \qquad (9)$$

where q is a parameter that specifies the dimension being referred to. Writing the sum of P_i^q as a weighted average, $\sum_i P_i^q = \sum_i P_i (P_i^{q-1}) = \langle P_i^{q-1} \rangle$, one can associate bulk with the generalized average probability per box $\sqrt[q-1]{\langle P_i^{q-1} \rangle}$, and identify D_q as a scaling of bulk with size. (The angle brackets $\langle\ \rangle$ denote the weighted average.)

Note that for $q = 2$, the generalized average is the ordinary arithmetic average, and for $q = 3$ it is a root-mean-square average. The limit $q \to 1$ leads to a geometric average and corresponds to the information or average pointwise dimension. Finally, $q = 0$ gives the box-counting dimension.

For a uniform fractal (i.e., all P_i equal), the generalized dimension D_q does not vary with q. For a nonuniform fractal, the variation of D_q with q quantifies the nonuniformity; e.g.,

$$D_\infty = \lim_{r \to 0} \frac{\log \left[\max_i P_i \right]}{\log r}, \text{ and}$$

$$D_{-\infty} = \lim_{r \to 0} \frac{\log \left[\min_i P_i \right]}{\log r}.$$

The above equations show that the maximum dimension $D_{-\infty}$ is associated with the least dense points on the fractal; the minimum dimension D_∞ corresponds to the most dense points.

The notion of generalized dimension first arose out of a need to understand why different algorithms gave different answers for dimension. A further motivation comes from the need to characterize fractals more fully. Rather than measure just one dimension of a fractal, one can compute the full spectrum of dimensions from $D_{-\infty}$ to D_∞.

In Ref. 11, Renyi defined a generalized information measure

$$S_q(r) = \frac{1}{q-1} \log \sum_i P_i^q \qquad (10)$$

that reduces to Shannon's formula in the limit $q \to 1$. The generalized information dimension associated with the Renyi entropy is just the generalized dimension that was defined earlier by another approach. That is, from Eqs. 7 and 10, the Renyi generalized dimension is given by

$$D_q = \lim_{r \to 0} \frac{-S_q(r)}{\log r} = \frac{1}{q-1} \lim_{r \to 0} \frac{\log \sum_i P_i^q}{\log r},$$

which is the same equation as the definition given in Eq. 9.

Spectrum of Scaling Indices: f(α)

An alternative interpretation of generalized dimension is provided by the spectrum of scaling indices $f(\alpha)$, which was introduced in Ref. 12. Although $f(\alpha)$ is formally defined as a Legendre transformation of the function $\tau(q)$, in which $\tau(q) \equiv (q-1)D_q$, a transformation can be obtained by interpreting f as a kind of frequency of occurrence of pointwise dimensions α throughout the fractal.

Suppose the attractor is partitioned into many small sets, and the ith such set has a measure μ_i and a size $\delta_i < r$, where r represents the coarseness of the partition. Then α_i is defined by the local scaling $\mu_i = \delta_i^{\alpha_i}$. Let $n(\alpha, r)$ be the number of these small sets that have a scaling index between α and $\alpha + \Delta\alpha$. Finally, the spectrum $f(\alpha)$ is defined by the scaling $n(\alpha, r) = r^{-f(\alpha)}\Delta\alpha$.

Now take the case in which the attractor is partitioned into fixed-size boxes and consider the sum $\sum_i P_i^q$. The number of terms in this sum for which P_i is equal to r^α is given by $n(\alpha, r)$. Thus the sum can be written as

$$\begin{aligned}\sum_i P_i^q &= \sum_\alpha n(\alpha,r) r^{q\alpha} \\ &= \int r^{-f(\alpha)} r^{q\alpha} \mathrm{d}\alpha \\ &= \int r^{q\alpha - f(\alpha)} \mathrm{d}\alpha \\ &\sim r^\theta\end{aligned} \tag{11}$$

The exponent θ is given by $\min_\alpha \{q\alpha - f(\alpha)\}$ because the integral will be dominated by the smallest power of r when $r \to 0$. Comparing Eq. 11 to the definition of generalized dimension in Eq. 9, which gives $\sum_i P_i^q \sim r^{(q-1)D_q} = r^{\tau(q)}$, one obtains

$$\tau(q) = \min_\alpha \{q\alpha - f(\alpha)\}, \tag{12}$$

which can be rewritten as

$$f(\alpha) = \min_q \{q\alpha - \tau(q)\}. \tag{13}$$

If $\tau(q)$ and $f(\alpha)$ are differentiable, then the following simpler forms of Eqs. 12 and 13 can be used:

$$\begin{aligned}\alpha &= \frac{\partial \tau}{\partial q}, \; f = \alpha q - \tau; \text{ and} \\ q &= \frac{\partial f}{\partial \alpha}, \; \tau = \alpha q - f.\end{aligned} \tag{14}$$

Although Eqs. 12 to 14 were obtained with informal scaling arguments, the equations serve as the formal definition of the spectrum of scaling indices $f(\alpha)$.

The scaling $n(\alpha, r) \sim r^{-f(\alpha)}$ suggests interpreting $f(\alpha)$ as the dimension of the points with scaling index α. Indeed, the scaling index α is associated with the pointwise dimension D_p. The set

$$\mathbf{S}_\alpha = \{\mathbf{X} \in \mathbf{A}: \; D_p(\mathbf{X}) = \alpha\}$$

is the set of all points in $\mathbf{A}$ for which the pointwise dimension is α. The Hausdorff dimension of the set $\mathbf{S}_\alpha$ is given by $f(\alpha)$.

Figure 7 shows the relationships between D_q, α, f, and q for a typical multifractal.

A Survey of Algorithms for Computing Dimension

Notions of quantity are possible only where there exists already a general concept which allows various modes of determination.

—George Friedrich Bernhard Riemann

The previous section introduced an algorithm based on box counting. For practical computation, however, the box-counting algorithm is plagued with a number of inefficiencies, particularly at high embedding dimensions. A variety of other algorithms for estimating dimension have been developed in recent years. This section will survey those algorithms.

The most popular way to compute dimension remains the correlation algorithm of Grassberger and Procaccia [13], which estimates dimension based on the statistics of pairwise distances. The main advantage of pairwise-distance algorithms is that they permit one to probe the scaling of the attracting set to very small length scales r. It is the small-r behavior that is important, and interpoint distances can probe down to an r value as small as the nearest interpoint distance. If an equivalently small

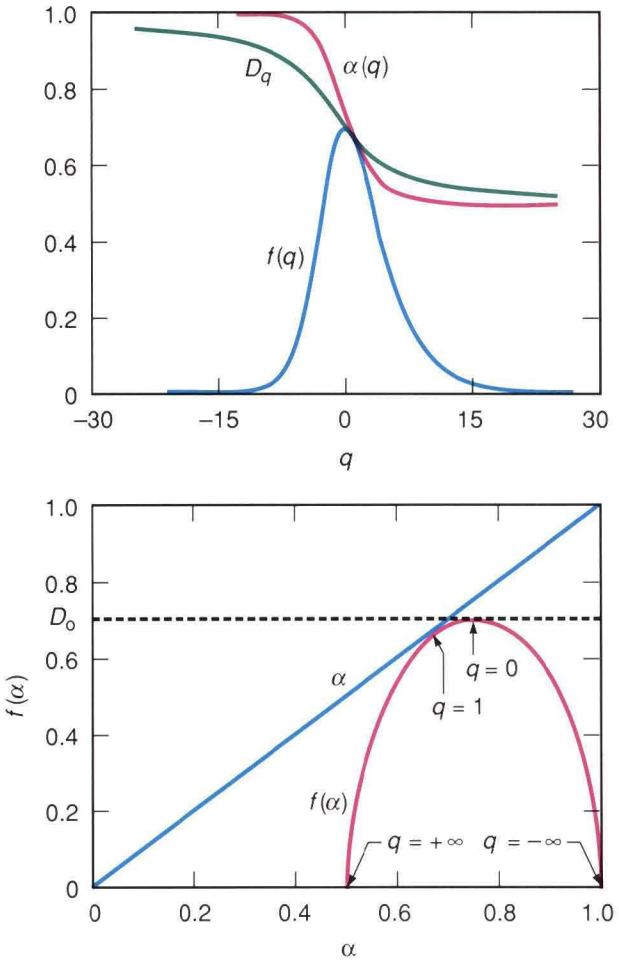


Fig. 7—Generalized dimension: (top) D_q *as a function of* q *for a typical multifractal. Also shown are both* f *and* α *as a function of* q *for the same multifractal. (bottom)* f *as a function of* α*. The curve is always convex upward, and the peak of the curve occurs at* $q = 0$*. At this point,* f *is equal to the fractal dimension* D_0*. Note that the* $f(\alpha)$ *curve is tangent to the curve* $f = \alpha$ *and the point of tangency occurs at* $q = 1$*. In general, the left branch of the curve corresponds to* $q > 0$ *and the right branch to* $q < 0$*.*

value of r were used in a box-counting algorithm, many of the tiny boxes would be improperly considered empty because there would not be enough points in the data set to fill the appropriate boxes. Furthermore, correlation algorithms have the advantage that they compute their measure of bulk for many different box sizes at once, whereas the box-counting algorithm requires that new grids be laid down for each different box size.

The box-counting and correlation algorithms are both in the class of *fixed-size* algorithms because they are based on the scaling of mass with size for boxes or grids of fixed sizes. An alternative approach uses fixed-mass boxes, usually by looking at the statistics of distances to kth nearest neighbors. Both fixed-size and fixed-mass algorithms can be applied to the estimation of generalized dimension D_q, although fixed-size algorithms do not work well for $q < 1$.

Some algorithms directly involve the dynamical properties of the strange attractor. For example, the Kaplan-Yorke conjecture (described later) relates dimension to the Lyapunov exponents. Recently, interest has focused on trying to ferret out the unstable periodic orbits of the attractor. The most direct use of the dynamics is to make predictions of the time series. Accurate predictions provide a reliable indication that the dynamics is indeed deterministic and low-dimensional.

Finally, the idea of *intrinsic dimension* is discussed. Methods in this category look for local hyperplanes that confine the data; the integer dimension of the local hyperplanes provides an upper bound on the fractal dimension of the attractor.

A number of these algorithms can be extended to compute generalized dimension, although the usefulness of this extension is, in many practical cases, limited. A generalized dimension *is* useful for quantifying the nonuniformity of the fractal and for comparing an exact and predictive theory to an experimental result. But the goal of dimension estimation is very often more qualitative in nature. One may only want to know whether the number of degrees of freedom is very large or reasonably small. To answer the question, is it chaos or is it noise? a *robust* estimate of dimension is more important than a *precise* estimate. In these cases, for example, the subtle distinction between the information dimension at $q = 1$ and correlation dimension at $q = 2$ may not be as important as the grittier issues that arise from experimental noise, finite samples, or even computational efficiency.

Average-Pointwise-Mass Algorithms

Recall the definition of the pointwise mass function in Eq. 5: $B_{\mathbf{x}}(r) = \mu(\mathbf{B}_{\mathbf{x}}(r))$, where $\mathbf{B}_{\mathbf{x}}(r)$ is

the ball of radius r around the point $\mathbf{X}$. The scaling of the quantity $B_{\mathbf{X}}(r)$ with r is what defines the local pointwise dimension at $\mathbf{X}$. The average pointwise dimension is then defined by computing the pointwise dimension $D_p(\mathbf{X})$ at each point $\mathbf{X}$ and averaging over the attractor. An alternate approach is to take some average of the pointwise masses and to define the dimension as the scaling of this average with r.

The most natural such averaging strategy was introduced by Grassberger and Procaccia [13]. Here a direct arithmetic average of the pointwise mass function gives what Grassberger and Procaccia call the *correlation integral* $C(r)$:

$$C(r) = \left\langle B_{\mathbf{X}}(r) \right\rangle .$$

From this, the correlation dimension ν is defined as

$$\nu = \lim_{r \to 0} \frac{\log C(r)}{\log r} .$$

Here $B_{\mathbf{X}_j}(r)$ can be approximated from a finite data set of size N by the equation

$$\begin{aligned} B_{\mathbf{X}_j}(r) &\cong B_{\mathbf{X}_j}(N, r) \\ &= \frac{\#\left\{\mathbf{X}_i :\ i \neq j \text{ and } \left\|\mathbf{X}_i - \mathbf{X}_j\right\| \leq r\right\}}{N-1} \\ &= \frac{1}{N-1} \sum_{\substack{i=1 \\ i \neq j}}^{N} \Theta\left(r - \left\|\mathbf{X}_i - \mathbf{X}_j\right\|\right) \end{aligned} \tag{15}$$

where Θ is the Heaviside step function: $\Theta(z)$ is zero for $z < 0$ and one for $z \geq 0$. One generally excludes the distance of a point to itself (i.e., the case in which $i = j$). It is now straightforward to approximate $C(r)$ with a finite data set:

$$\begin{aligned} C(N, r) &= \frac{1}{N} \sum_{j=1}^{N} B_{\mathbf{X}_j}(r) \\ &= \frac{1}{N(N-1)} \sum_{j=1}^{N} \sum_{\substack{i=1 \\ i \neq j}}^{N} \Theta\left(r - \left\|\mathbf{X}_i - \mathbf{X}_j\right\|\right) \\ &= \frac{2}{N(N-1)} \sum_{j=1}^{N} \sum_{i=j+1}^{N} \Theta\left(r - \left\|\mathbf{X}_i - \mathbf{X}_j\right\|\right). \end{aligned} \tag{16}$$

In words, Eq. 16 can be written in the following way:

$$C(N, r) = \frac{\#\text{ of distances less than } r}{\#\text{ of distances altogether}} . \tag{17}$$

Thus the correlation algorithm provides an estimate of dimension based purely on the statistics of pairwise distances. Not only is this an elegant formulation, but it has the more substantial advantage that it probes the attractor to much finer scale than, for instance, the box-counting algorithm. For N points, the $C(N, r)$ function of the correlation method ranges from $2/N^2$ to 1. In a logarithmic sense, this range is twice that available to the box-counting method's $n(N, r)$ function, which ranges from 1 to N. The correlation integral also exhibits twice the logarithmic range of the single-point mass functions $B_{\mathbf{X}}(N, r)$.

A more general average $G_q(r)$ than the direct arithmetic average used above is given by the following equation:

$$G_q(r) = \sqrt[q-1]{\left\langle B_{\mathbf{X}}(r)^{q-1} \right\rangle} . \tag{18}$$

The scaling of this average with r, i.e., $G_q(r) \sim r^{D_q}$, gives the generalized dimension D_q:

$$D_q = \lim_{r \to 0} \frac{\log G_q(r)}{\log r} .$$

From Eqs. 15 and 18, $G_q(r)$ can be approximated with a finite set of points by

$$G_q(N, r) = \left[\frac{1}{N} \sum_{i=1}^{N} \left[B_{\mathbf{X}_i}(N, r) \right]^{q-1} \right]^{1/(q-1)}$$

where $B_{\mathbf{X}_i}(N, r)$ is defined in Eq. 15. The formula is easiest to evaluate when $q = 2$, in which case it reduces to the simple form of Eq. 16. The formula, however, does not work well for $q \leq 1$ because the term raised to the power of $q - 1$ is often zero for r much larger than the smallest interpoint distance.

Grassberger [14] suggested a q-point correlation integral defined by counting q-tuples of points that have the property that every pair of points in the q-tuple is separated by a dis-

tance that is less than r.

$$C_q(N, r) = \frac{1}{N^q} \#\left\{(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_q}): \right.$$
$$\left. \|\mathbf{X}_n - \mathbf{X}_m\| \le r \text{ for all } n, m \in \{i_1, \ldots, i_q\}\right\},$$

which has the scaling behavior $C_q(r) \sim r^{(q-1)D_q}$. Thus

$$D_q = \frac{1}{q-1} \lim_{r\to 0} \lim_{N\to\infty} \frac{\log C_q(N, r)}{\log r}. \qquad (19)$$

Equation 19 can in principle be applied to all integer $q \ge 2$. However, the equation's implementation is unwieldy for $q \ge 3$ because the number of q-tuples (which is given by N^q) grows very rapidly with N when $q \ge 3$.

The kth-Nearest-Neighbor (Fixed-Mass) Algorithms

The average pointwise mass functions described in the previous section consider the variation of mass inside fixed-size boxes. In contrast to that approach, the nearest-neighbor algorithms consider the scaling of sizes in fixed-mass boxes.

An early implementation of this notion in a chaotic dynamics context was suggested by Termonia and Alexandrowicz [15]. Here one computes $\langle r_k \rangle$, the average distance to the kth nearest neighbor, as a function of k. Let $R(\mathbf{X}, k)$ denote the distance between point $\mathbf{X}$ and its kth nearest neighbor. (By convention, the zeroth nearest neighbor is the point itself, so $R(\mathbf{X}, 0) = 0$ for all $\mathbf{X}$.) The average is

$$\langle r_k \rangle = \frac{1}{N} \sum_{i=1}^{N} R(\mathbf{X}_i, k).$$

The scaling $\langle r_k \rangle \sim k^{1/D}$ defines the dimension.

Badii and Politi [16] considered the moment γ of the average distance to the kth-nearest neighbor. They recommended keeping k fixed and computing a dimension function $D(\gamma)$ from the scaling of average moments of r_k with total number of points N:

$$\langle r_k^\gamma \rangle \sim (k/N)^{\gamma / D(\gamma)}.$$

This dimension function $D(\gamma)$ is related to the generalized dimension by the implicit formulas

$$\gamma = (q-1)D_q; \quad D(\gamma) = D_q. \qquad (20)$$

Grassberger [17] derived Eq. 20 independently and used the Γ-function to provide a small k correction to the scaling of $\langle r_k^\gamma \rangle$ with k:

$$\langle r_k^\gamma \rangle \sim \left[\frac{\Gamma(k + \gamma/D)}{\Gamma(k)k^{\gamma/D}}\right](k/N)^{\gamma/D}.$$

The term contained in the square brackets approaches unity as k becomes large.

Algorithms That Use Dynamical Information

Let $T_{\mathbf{X}_0}(r)$ be the r-recurrence time of an initial condition $\mathbf{X}_0$. This time is the minimum amount necessary for the trajectory of $\mathbf{X}_0$ to come back to within r of $\mathbf{X}_0$. Thus the inverse recurrence time, which is related to the time the trajectory is within r of $\mathbf{X}_0$, provides an estimate of the pointwise mass function. In other words, one expects

$$\frac{1}{T_{\mathbf{X}_0}(r)} \sim B_{\mathbf{X}_0}(r).$$

Consequently, the scaling of the average recurrence time with r can be related to the generalized dimension [14]:

$$\sqrt[1-q]{\left\langle T_{\mathbf{X}}(r)^{1-q} \right\rangle} \sim \sqrt[q-1]{\left\langle B_{\mathbf{X}}(r)^{q-1} \right\rangle} \sim r^{D_q}.$$

Another algorithm—the Kaplan-Yorke conjecture [18]—relates the fractal dimension of a strange attractor to the Lyapunov exponents. The conjectured formula defines what has come to be called the Lyapunov dimension (d_L):

$$d_L = j + \frac{\sum_{i=1}^{j} \lambda_i}{|\lambda_{j+1}|}$$

where $\{\lambda_1, \ldots, \lambda_m\}$ are the Lyapunov exponents in decreasing order, and j is given by

$$j = \sup\left\{k: \sum_{i=1}^{k} \lambda_i > 0\right\}.$$

The conjecture is that the Lyapunov dimension corresponds to the information dimension D_q, where $q = 1$.

When the equations of motion are available, the Jacobian matrix $\mathbf{J}(t)$ can be computed directly and the direct computation improves the calculation of Lyapunov exponents and the resulting Lyapunov dimension. In this case, computing the Lyapunov dimension is the best strategy for estimating dimension. In fact, it is the only approach that has been shown to work with dimensions larger than 10.

In discussing algorithms for predicting the future of time series, Farmer and Sidorowich [19] suggested several ways that the prediction algorithms might be used to make more accurate and reliable dimension estimates. One promising possibility is to use *bootstrapping*, in which a short time series of length N is extrapolated far into the future, and the resulting time series is used in a conventional dimension algorithm. This method is appropriate when only a small data set is available and computation time is not the limiting factor.

A second possibility is to scale the prediction error with N, the number of points from which the prediction is made. The N-scaling can be related to the fractal dimension. Moreover, it has the desirable property that if it is observed—i.e., if the prediction error noticeably reduces with increasing N—then the system can reliably be considered low-dimensional.

Instead of trying to determine the exact fractal dimension, other algorithms count the number of degrees of freedom in a more approximate way by finding the minimum embedding dimension (called the intrinsic dimension) for the time-series data. The estimate is necessarily coarser, but possibly more reliable.

The idea of seeking minimum embedding dimensions was first discussed in the context of dynamical systems by Froehling, Crutchfield, Farmer, Packard, and Shaw [20]. In Ref. 20, the authors attempted to find a linear fit of points in a local region to a D-dimensional hyperplane of an m-dimensional reconstructed embedding space. Broomhead and King [21] suggested using singular-value decomposition to make the fit to the D-dimensional hyperplane more efficient. The prediction algorithms of Farmer and Sidorowich [19, 22] can also be used to estimate intrinsic dimensions. Basically, the smallest embedding dimension for which good predictions can be made is a reliable upper bound on the number of degrees of freedom in the time series.

Practical Estimation of the Correlation Dimension

No estimate is more in danger of erroneous calculations than those by which a man computes the force of his own genius.

—Samuel Johnson

The most popular, and for many purposes the best, algorithm for estimating dimension is the correlation algorithm of Grassberger and Procaccia [13]. For one thing, the correlation algorithm is very easy to implement: one computes the correlation integral $C(N, r)$ merely by counting distances. Equation 17 defines the estimated correlation integral $C(N, r)$ as the fraction of distances less than r. From this the correlation dimension is in principle given by

$$\nu = \lim_{r \to o} \lim_{N \to \infty} \frac{\log C(N, r)}{\log r} . \tag{21}$$

However, a variety of practical issues and potential pitfalls come with making an estimate from finite data.

Extraction of the Slope

Extracting dimension directly from the correlation integral according to Eq. 21 is extremely inefficient, since the convergence to ν as $r \to 0$ is logarithmically slow.

Taking a slope solves this problem. One can either take the local slope

$$\begin{aligned} \nu(r) &= \frac{\mathrm{d} \log C(N, r)}{\mathrm{d} \log r} \\ &= \frac{\mathrm{d}C(N, r)/\mathrm{d}r}{C(N, r)/r} , \end{aligned} \tag{22}$$

or one can fit a chord through two points on the

curve in the following way:

$$\nu(r) = \frac{\Delta \log C(N,r)}{\Delta \log r} = \frac{\log C(N,r_2) - \log C(N,r_1)}{\log r_2 - \log r_1}. \quad (23)$$

However, implementing the chord strategy requires the choice of two length scales r_2 and r_1. The larger scale is limited by the size of the attractor, and the smaller scale is limited by the smallest interpoint distance.

The direct difference in Eq. 23 fails to take full advantage of the information available in $C(N, r)$ for values of r between r_1 and r_2. To use those values, one might be tempted to fit a slope by some kind of least-squares method, but this approach is problematic. Unweighted least squares are particularly poor, because the estimate of $C(r)$ by $C(N, r)$ is usually much better for large r than for small r. Weighted fits can compensate for this effect, but there is still a problem because successive values of $C(N, r)$ are not independent. Since $C(N, r + \Delta r)$ is equal to $C(N, r)$ plus the fraction of distances between r and $r + \Delta r$, it is a mistake to assume that $C(N, r+\Delta r)$ is independent of $C(N, r)$. An estimate of dimension will depend on whether $C(N, r)$ is sampled at logarithmic intervals (r = 0.001, 0.01, 0.1, . . .) or at uniform intervals (r = 0.001, 0.002, 0.003, . . .). Furthermore, the error estimate that the least-squares fit naturally provides will have very little to do with the actual error in the dimension estimate.

Because of the limitations of the least-squares approach, Takens [23] developed a method in which the intermediate-distance information could be incorporated in a consistent way. Using the theory of best estimators, Takens derived an estimator for ν that makes optimal use of the information in the interpoint distances $r_{i,j}$. In fact, the method uses all distances less than an upper bound r_0 so that a lower cutoff distance need not be specified.

$$\nu(r_0) = \frac{-1}{\left\langle \log\,(r_{i,j} / r_0) \right\rangle} \quad (24)$$

where the angle brackets $\langle\,\rangle$ refer to an average over all distances $r_{i,j}$ that are less than r_0. In terms of the correlation integral, Eq. 24 can be written [24] as the following:

$$\nu(r_0) = \frac{C(r_0)}{\int_0^{r_0} [C(r)/r]\,dr} = \frac{\int_0^{r_0} [dC(r)/dr]\,dr}{\int_0^{r_0} [C(r)/r]\,dr}. \quad (25)$$

The more unwieldy form of Eq. 25 is meant to be compared to Eq. 22 for the local slope.

Not only does the Takens estimator provide an efficient means to squeeze an optimal estimate out of the correlation integral, it also supplies an estimate of the statistical error, namely $\sigma_\nu = \nu/\sqrt{(N^*)}$ [23], where N^* is the number of distances less than r_0. However, this error estimate makes the assumption that the N^* distances are statistically independent, and is only valid if $N^* \le N$ [25].

Reference 26 points out that even though the Takens estimator uses all distances less than r_0, the estimator is still sensitive to oscillations in the correlation integral.

Computation

Although there are $O(N^2)$ interpoint distances to be considered in a time series of N points, most of the distances are large, i.e., on the same order as the size of the attractor. Because the short distances are the most important, and because there are so few of them, it is advantageous to organize the points on the attractor so that only the short distances are computed. The algorithm in Ref. 27 provides a way to compute the $O(N)$ shortest distances with only $O(N \log N)$ effort by placing the points in boxes of size r_0. Farmer and Sidorowich [19] organized points in a "*k–d* tree" for their prediction algorithm; such an organization of points would also be useful for the *k*-nearest-neighbor algorithms that were discussed earlier.

Comparison with Filtered Noise

It is a common and well-advised tactic to compare the correlation dimension computed

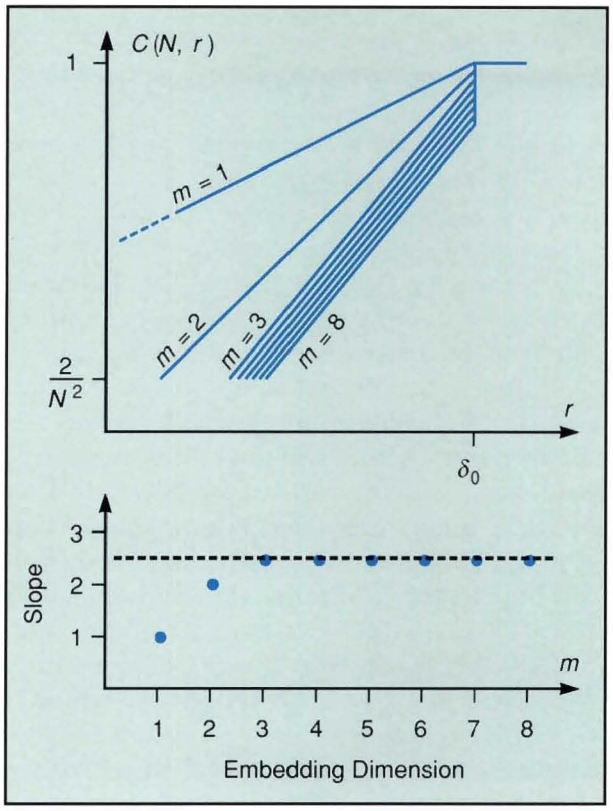


Fig. 8—Plot of an idealized correlation integral C(N, r) *versus* r *on log-log axes. In the ideal case,* C(N, r) *scales as* r^m *for the embedding dimension* $m < \nu$*, and as* r^ν *for* $m > \nu$ *over a range from* $C(N, r) = 2/N^2$ *to saturation at* $C(N, r) = 1$*. Here* ν *is somewhere between 2 and 3. This idealization, however, is only approximated by correlation integrals computed from actual samples of time-series data. The value* δ_0 *is the diameter of the attractor.*

for a given time series against a set of test data whose properties are known in advance. For example, if the given time series has a significantly different dimension than does white noise (i.e. random noise), then one can rule out the null hypothesis that the original time series is white noise. However, the original time series might still be colored noise; i.e., the time-series data might be correlated with itself (called autocorrelation).

A more stringent test is to create a time series with the same Fourier spectra as the original time series. For instance, one could take a Fourier transform of the original time series, randomize the phases, and then invert the transform. If the $C(N, r)$ obtained from the new time series is significantly different from that of the original time series, then a stronger statement can be made: not only is the original time series not white noise, it could not have been produced by any linear filter of white noise.

Sources of Error

The estimation of dimension from a time series is prone to two fundamental sources of error: statistical imprecision and systematic bias. The first type of error results directly from the finite sampling of data; as such, the error is reasonably tractable. The second type of error comes from a wide variety of sources. Indeed, that wide variety has led to great difficulty in verifying results as reliable.

The following sections present a brief survey of the kinds of problems that arise in the computation of correlation dimension.

Finite Sampling and Statistical Error

What finally limits an estimate of dimension is that only a finite number of points sample the attractor. Many of the systematic effects discussed in the following sections can be eliminated in the $N \to \infty$ limit.

First, a finite sample of N points limits the range of interpoint distances. The correlation integral, which is approximated by $C(N, r) \sim r^D$, varies from $2/N^2$ to 1, so that any fit has only this range to work with (Fig. 8). In particular, any fit over a range R of distances requires $N^2/2 \geq R^D$, so that at least $N = \sqrt{(2R^D)}$ points are required. (R is the ratio of the largest to the smallest r.) This scaling, however, is an absolute lower bound. In practice, many more points are required, due in most cases to the variety of systematic effects (discussed below) that must be overcome.

The statistical error in an estimate of dimension typically scales as $1/\sqrt{N}$, where the coefficient of the scaling depends on the fluctuations in the pointwise masses $B_{\mathbf{x}}(r)$ [25]. In special cases, the coefficient can be zero, in which case $1/N$ scaling can be observed.

Edges and Finite Size

The finite size of a compact fractal object limits the range over which $C(r)$ scales as r^ν.

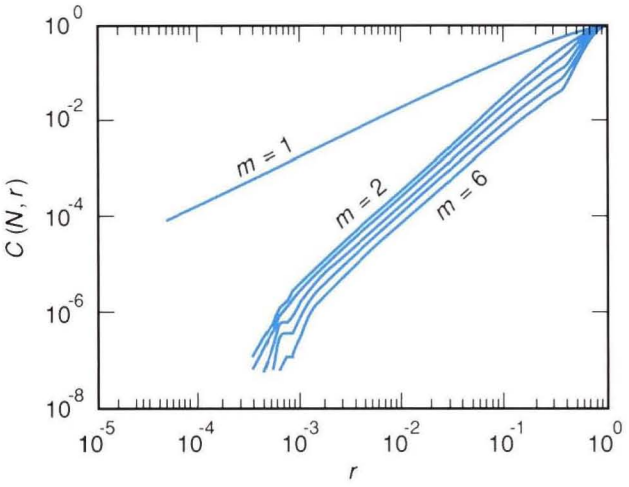


Fig. 9—An actual correlation integral for a two-dimensional chaotic attractor with embedding dimensions m = *1 through* m = *6. The finite sample size leads to poor statistics at small* r, *and the finite size of the attractor (the edge effect) limits the scaling at large* r. *Nonetheless, the slopes are more or less constant over a range of* C(N, r) *of order* N^2.

For r greater than the diameter of the attractor (δ_0), the correlation integral saturates at $C(r) = 1$. This finite-size effect is not necessarily a problem in dimension calculations. As long as the effect is confined to length scales larger than some $r_0 \cong \delta_0$, then accurate estimates of dimension can still be obtained in the $r \leq r_0$ range from the slope of a plot of log $C(r)$ versus log r (Figs. 8 and 9).

The real problem stems from the edges that finite-sized objects in $\mathbf{R}^m$ all have. The neighborhoods around points near the edge have different scaling from neighborhoods further in the interior.

Although any model of edge effect will depend on the shape of the fractal, a tractable and reasonably generic model assumes that the fractal is a uniform hypercube of unit length and dimension m. In such a case, the correlation integral can be derived exactly [24]: $C(N, r) = (2r - r^2)^m$. The local slope at r of the log-log curve is given by

$$\nu(r) = \frac{r}{C(r)}\left(\frac{dC}{dr}\right) = \frac{m(2-2r)}{2-r} \cong m\left(1 - \frac{r}{2}\right),$$

so that the relative error is $|\nu(r) - m|/m \cong r/2$. Here is an explicit demonstration of the need for the $r \to 0$ limit in an estimate of dimension.

Noise

Noise, the ultimate corrupter of measurements, is usually the first concern of the experimentalist. In the case of dimension estimation, however, the effect of low-amplitude noise is often not as significant as other effects.

One expects that the fractal scaling of bulk with size will break down at length scales equal to the noise amplitude. But unless the system amplifies noise excessively, one does not expect the scaling to be affected at length scales much larger than the noise amplitude (Fig. 10). Although noise *is* amplified along the expansion directions of a chaotic attractor, this effect does not have much influence on the dimension estimation because the noise is amplified back onto the attractor. In other words, the noise is drawn to the attractor and consequently has little effect on the scaling. Thus at relatively high SNRs there is still a good range over which a fractal may be scaled.

Because noise often possesses a much higher characteristic frequency than the deterministic attractor, it is tempting to subject the signal to a low-pass filter to reduce the noise effects. As a rule, however, the use of a low-pass filter is not recommended because the filter can actually increase the dimension of the time series [28].

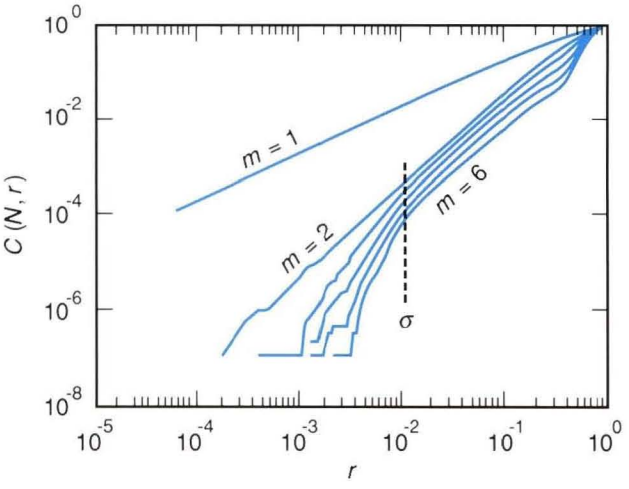


Fig. 10—The effect of noise is seen in this figure. Where σ *is the amplitude of the noise, one sees that for* $r << \sigma$, *a slope that approaches the embedding dimension* m *is observed. For* $r >> \sigma$, *the effect of the noise is unimportant.*

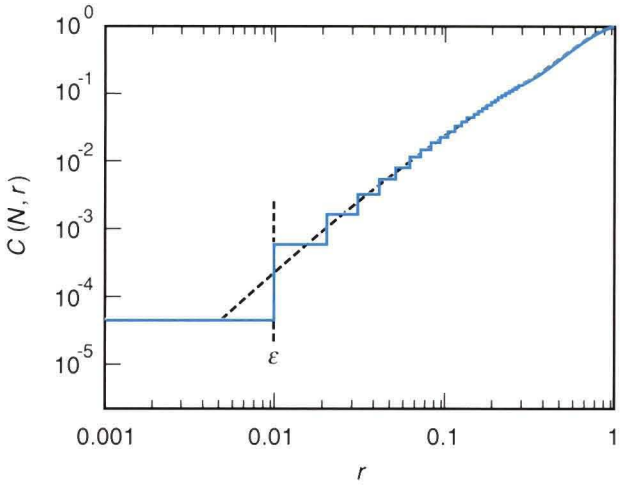


Fig. 11—The effect of discretization is to introduce stair steps into the correlation integral. The steps are all of equal width, but the log-log plot magnifies those at small r. The effect is minimized if one plots log C(N, r) versus log r for r = (k + ε/2) where k is an integer and ε is the level of discretization (dashed curve).

Discretization

Discretized time series are of the form $x_i = k_i\varepsilon$, where k_i is an integer and ε is the discretization level. Such discretization is a natural artifact of digital measuring devices. In fact, many algorithms work much faster with integers than with floating-point numbers, so that it may be computationally wise to make the conversion; the conversion involves multiplication by some large factor, followed by rounding to the nearest integer. The multiplicative factor does not affect the slope of a log-log plot, but the rounding is equivalent to a discretization.

Distances between pairs of discretized points will themselves be discrete multiples of ε. This effect is most prominent at small r; indeed, pairs of points with $r = 0$ occur with finite probability and a plot of $\log C(r)$ versus $\log r$ must deal with the $r = 0$ points. A model of points on an m-dimensional lattice, with the lattice points separated by ε, leads to a scaling of $C(r) \sim (r + \varepsilon/2)^m$ to the first order of ε [24]. This result suggests that an appropriate plot for a general discretized time series is $\log C(r)$ versus $\log (r + \varepsilon/2)$, as shown in Fig. 11.

An alternate approach [29] deliberately adds noise of amplitude ε (the process is called dithering) to the original time series and then plots the usual $\log C(r)$ versus $\log r$.

Lacunarity

Dimension is not the only way to gauge how fractal a set is. Mandelbrot [7] pointed to *lacunarity* as another measure. He describes lacunarity in the following way: for two fractals having the same dimension, the one that is more textured and appears more fractal has greater lacunarity.

From the point of view of dimension computation, lacunarity has the effect of introducing an intrinsic oscillation into the correlation integral (Fig. 12). If the range over which the slope is estimated is long enough to encompass several periods of the oscillation, then the effect of the oscillation will be minimized. On the other hand, if attempts to compute dimension are based on a local slope of the correlation integral, lacunarity can prevent the dimension estimator from converging.

Autocorrelation

Autocorrelation is very common in time-series data. For continuous signals $x(t)$, there is always some time τ over which $x(t)$ and $x(t+\tau)$ are strongly correlated. If this autocorrelation time τ is long compared to the sampling time, then an

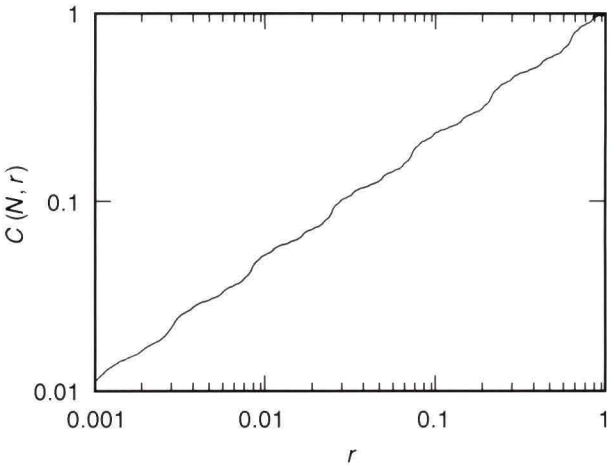


Fig. 12—Lacunarity leads to an intrinsic oscillation in the correlation integral. The oscillation inhibits accurate determinations of slope. The example here is the correlation integral of the middle-thirds Cantor set [7].

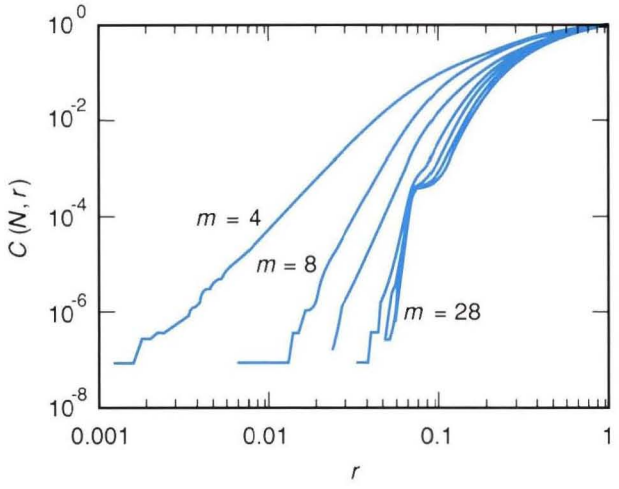


Fig. 13—Autocorrelation in the time-series data can lead to an anomalous shoulder in the correlation integral. The effect is most highly pronounced for high-dimensional attractors. In this example, the input time series was autocorrelated Gaussian noise and the correlation integral was computed for various large embedding dimensions (m = 4 to m = 28). Eq. 26 can be used to correct for this effect.

anomalous shoulder (Fig. 13) could appear in the correlation integral. The shoulder is a problem in that it can lead to inaccurate and possibly spurious estimates of dimension [30].

One solution is to increase the sampling time. The increase, however, may have adverse side effects; e.g., it could further limit the available data, and it could also affect the delay-time embedding strategy. A more effective solution can be obtained by rewriting the definition of the correlation integral from Eq. 16:

$$C(W, N, r) = \frac{2}{(N+1-W)(N-W)} \cdot \sum_{n=W}^{N-1} \left[\sum_{i=0}^{N-1-n} \Theta\left(r - \|\mathbf{X}_i - \mathbf{X}_{i+n}\|\right) \right]. \quad (26)$$

Eq. 26 computes distances between all pairs of points *except* for those that are closer together *in time* than W sampling units. Note that the case in which $W = 1$ is just the standard algorithm. Eliminating this small selection of offending pairs eliminates the anomalous shoulder without sacrificing the statistics of $O(N^2)$ distance calculations.

The definition of the correlation integral in Eq. 17 is now adjusted so that the numerator is equal to the number of distances less than r except for those distances from pairs of points closer together in time than W.

Summary

Accurate and foolproof estimation of fractal dimension remains an elusive task. Prediction-based algorithms appear to hold the most promise, but they are still in an early stage of development. The standard tool is the correlation integral of Grassberger and Procaccia. Although the algorithm is subject to a variety of practical limitations, most of them are now well known and many can be remedied, or at least compensated for. When the Grassberger and Procaccia algorithm is carefully applied, it can distinguish stochastic from low-dimensional deterministic behavior even if the latter is represented by a finite (but not too small) time series of noisy (but not too noisy) data.

Direct estimation of dimension is one way to quantify the complexity of nonlinear systems that have only a few active degrees of freedom. For systems with more than a few (in practice, more than about eight), direct estimation from a time series will probably not be possible. To understand and quantify the self-organization of more complicated systems is a more interesting and proportionately more difficult problem.

Acknowledgments

The author thanks Courosh Mehanian, Michael Holz, and Bette Korber for their assistance with the manuscript. The author also acknowledges Joyce Michaels for her help with the bibliography and Al Gschwendtner for his continuing encouragement and support. This work was sponsored by the U.S. Air Force and the Defense Advanced Research Projects Agency (DARPA).

(Editor's note—This article is adapted from a review article written for the *Journal of the Optical Society of America A.*)

References

1. H. Haken, *Information and Self-Organization: A Macroscopic Approach to Complex Systems in Springer Series in Synergetics* **40** (Springer-Verlag, Berlin, 1988).
2. S.M. Hammel, C.K.R.T. Jones, and J.V. Moloney, "Global Dynamical Behavior of the Optical Field in a Ring Cavity," *J. Opt. Soc. Am. B* **2,** 552 (1985).
3. M. Hénon, "A Two-Dimensional Mapping with a Strange Attractor," *Commun. Math. Phys.* **50,** 69 (1976).
4. J.-P. Eckmann and D. Ruelle, "Ergodic Theory of Chaos and Strange Attractors," *Rev. Mod. Phys.* **57,** 617 (1985).
5. N.H. Packard, J.P. Crutchfield, J.D. Farmer, and R.S. Shaw, "Geometry from a Time Series," *Phys. Rev. Lett.* **45,** 712 (1980).
6. A.M. Fraser and H.L. Swinney, "Independent Coordinates for Strange Attractors from Mutual Information," *Phys. Rev. A* **33,** 1134 (1986).
7. B.B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, 1983).
8. L.P. Kadanoff, "Fractals: Where's the Physics?" *Physics Today* **39: 2,** 6 (February 1986).
9. H.E. Stanley and N. Ostrowsky, Eds., *On Growth and Form: Fractal and Non-Fractal Patterns in Physics* (Martinus Nijhoff, Boston, 1986).
10. F. Hausdorff, "Dimension und äußeres Maß," *Mathematische Annalen* **79,** 157 (1919).
11. A. Renyi, *Probability Theory* (North-Holland, Amsterdam, 1970).
12. T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, and B.I. Shraiman, "Fractal Measures and Their Singularities: The Characterization of Strange Sets," *Phys. Rev. A* **33,** 1141 (1986).
13. P. Grassberger and I. Procaccia, "Measuring the Strangeness of Strange Attractors," *Physica* **9D,** 189 (1983).
14. P. Grassberger, "Generalized Dimensions of Strange Attractors," *Phys. Lett. A* **97,** 227 (1983).
15. Y. Termonia and Z. Alexandrowicz, "Fractal Dimension of Strange Attractors from Radius Versus Size of Arbitrary Clusters," *Phys. Rev. Lett.* **51,** 1265 (1983).
16. R. Badii and A. Politi, "Statistical Description of Chaotic Attractors: The Dimension Function," *J. Stat. Phys.* **40,** 725 (1985).
17. P. Grassberger, "Generalizations of the Hausdorff Dimension of Fractal Measures," *Phys. Lett. A* **107,** 101 (1985).
18. J.L. Kaplan and J.A. Yorke, *Functional Differential Equations and Approximations of Fixed Points*, in *Springer Lecture Notes in Mathematics* **730** (Springer-Verlag, Berlin, 1979), p. 204.
19. J.D. Farmer and J.J. Sidorowich, "Exploiting Chaos to Predict the Future and Reduce Noise," in *Evolution, Learning, and Cognition* (World Scientific, Teaneck, NJ, 1988), p. 227.
20. H. Froehling, J.P. Crutchfield, D. Farmer, N.H. Packard, and R. Shaw, "On Determining the Dimension of Chaotic Flows," *Physica* **3D,** 605 (1981).
21 D.S. Broomhead, R. Jones, and G.P. King, "Topological Dimension and Local Coordinates from Time Series Data," *J. Phys. A* **20,** L563 (1987).
22. J.D. Farmer and J.J. Sidorowich, "Predicting Chaotic Time Series," *Phys. Rev. Lett.* **59,** 845 (1987).
23. F. Takens, "On the Numerical Determination of the Dimension of an Attractor," in *Dynamical Systems and Bifurcations: Proceedings of a Workshop Held in Groningen, The Netherlands, 16–20 Apr. 1984*, in *Lecture Notes in Mathematics* **1125** (Springer-Verlag, Berlin, 1985), p. 99.
24. J. Theiler, "Quantifying Chaos: Practical Estimation of the Correlation Dimension," Ph.D. thesis, California Institute of Technology, Pasadena, CA, 1988.
25. J. Theiler, "Statistical Precision of Dimension Estimators," *Phys. Rev. A* **41,** 3038 (1990).
26. J. Theiler, "Lacunarity in a Best Estimator of Fractal Dimension," *Phys. Lett. A* **133,** 195 (1988).
27. J. Theiler, "Efficient Algorithm for Estimating the Correlation Dimension from a Set of Discrete Points," *Phys. Rev. A* **36,** 4456 (1987).
28. R. Badii, G. Broggi, B. Derighetti, M. Ravani, S. Ciliberto, A. Politi, and M.A. Rubio, "Dimension Increase in Filtered Chaotic Signals," *Phys. Rev. Lett.* **60,** 979 (1988).
29. M. Möller, W. Lange, F. Mitschke, N.B. Abraham, and U. Hübner, "Errors from Digitizing and Noise in Estimating Attractor Dimensions," *Phys. Lett. A* **138,** 176 (1989).
30. J. Theiler, "Spurious Dimension from Correlation Algorithms Applied to Limited Time-Series Data," *Phys. Rev. A* **34,** 2427 (1986).

JAMES THEILER is a staff member in the Opto-Radar Systems Group at Lincoln Laboratory, where he specializes in nonlinear dynamics and neural networks. Before joining the Laboratory two years ago, he worked for the Institute for Nonlinear Science at the University of California, San Diego. A member of Phi Beta Kappa, James received S.B. degrees in mathematics and physics from MIT, and a Ph.D. in physics from the California Institute of Technology. Last year, he and his wife Bette Korber became the proud parents of the brilliant and infamous Maxwell Brian Theiler.