Quantifying Chaos:

Practical Estimation of the Correlation Dimension

Thesis by
James Theiler

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1988

(Submitted 11 September 1987)

and I have learned from experience
with mathematics and love
that neither is pure.

ACKNOWLEDGEMENTS

I wish to acknowledge Professor Noel R. Corngold for his clear perspective, wise counsel, infinite patience, and uncountable criticisms (mostly constructive), as my thesis advisor. In the same breath, I acknowledge Mark Muldoon for many fruitful discourses and digressions.

Various conversations with Jim Havstad, Peter Grassberger, Henry Greenside, and J. Doyne Farmer have proved useful and interesting. I was first introduced to nonlinear differential equations by J. Patrick Miller.

Finally, I would like to thank my parents, without whom *dot dot dot*, and all my family and friends and relatives, too numerous to mention, and too generous to forget.

Contents

Tables and Figures

Outline

I. INTRODUCTION AND OVERVIEW

II. CORRELATION DIMENSION

A. Initial remarks

B. Geometrical effects

ABSTRACT

Be it a physical object or a mathematical model, a nonlinear dynamical system can display complicated aperiodic behavior, or "chaos." In many cases, this chaos is associated with motion on a strange attractor in the system's phase space. And the dimension of the strange attractor indicates the effective number of degrees of freedom in the dynamical system.

In this thesis, we investigate numerical issues involved with estimating the dimension of a strange attractor from a finite time series of measurements on the dynamical system.

Of the various definitions of dimension, we argue that the correlation dimension is the most efficiently calculable and we remark further that it is the most commonly calculated. We are concerned with the practical problems that arise in attempting to compute the correlation dimension. We deal with geometrical effects (due to the inexact self-similarity of the attractor), dynamical effects (due to the nonindependence of points generated by the dynamical system that defines the attractor), and statistical effects (due to the finite number of points that sample the attractor). We propose a modification of the standard algorithm, which eliminates a specific effect due to autocorrelation, and a new implementation of the correlation algorithm, which is computationally efficient.

Finally, we apply the algorithm to chaotic data from the Caltech tokamak and the Texas tokamak (TEXT); we conclude that plasma turbulence is not a low-dimensional phenomenon.

CHAPTER ONE

1. INTRODUCTION

1.1 Chaos

From economics [1] to epidemiology [2], the paradigm of deterministic chaos has been invoked to describe the irregular cycles and fluctuations that are observed in the physical world. According to the paradigm, chaotic motion can often be characterized as motion over a "strange attractor" in the system's phase space. The strange attractor is a complicated self-similar set whose (typically fractional) dimension indicates the effective number of degrees of freedom in the system. With varying success, evidence for the existence of these strange attractors (and in some cases, estimates of their dimensions) has been sought in a wide variety of physical systems [3]: in optical systems, such as lasers [4] and bistable devices [5]; in electrical circuits with nonlinear oscillators [6] or with p-n junctions [7]; in solid state phenomena, such as spin waves [8] and electronic transport [9]; in biological systems [10] such as stimulated cardiac cells [11], human electroencephalograms [12], giant squid axons [13], and slime mould [14]; in chemical systems such the Belousov-Zhabotinskii reaction [15], and in worldwide climatic patterns over the past million years [16]. Physicists and engineers have long sought models for hydrodynamical turbulence, and much of the work in the field of experimental nonlinear dynamics has concentrated on fluid mechanical systems; here, the most notable successes have been with Rayleigh-Bénard convection [17], Couette-Taylor flow [18], and Robert Shaw's celebrated dripping faucet [19]. In each of these hydrodynamical examples, low dimensions were observed for systems just beyond the onset of

chaos. It is not known whether low-dimensional chaos can provide an adequate description of more fully deveoped turbulence. Our own investigations were stirred by the desire to characterize the turbulence exhibited by a tokamak plasma [20,21].

Numerical methods have been developed for estimating the dimension of a strange attractor directly from a time series of measurements taken of the system. Though these methods have only recently been developed, they are sufficiently general that they can be applied to any system or simulation that can provide a time series.

Despite their popularity and wide applicability, these dimenson algorithms are subject to a profuse assortment of errors. A thorough analysis of these errors, and of the biases and limitations inherent in these algorithms, has been lacking. This thesis addresses that lack. We concentrate on the correlation dimension, as it is the most efficiently computable and (therefore) the most frequently used. We attempt to isolate and identify the causes of various "effects" that can lead to inefficient, imprecise, or incorrect estimates of dimension.

1.1.1 Simple systems with complicated behavior

The central observation of nonlinear dynamics is that simple deterministic systems with a few degrees of freedom can display chaotic behavior. This observation is credited to Lorenz [22], who in 1963 exhibited a system of three ordinary differential equations that were inspired by a model for convective flow in the earth's atmosphere. His analysis was primarily numerical, and his

computations revealed a complicated aperiodic flow. That such irregular behavior should arise from a simple deterministic system of equations came as a surprise to the physics community.

We contrast these "simple deterministic" systems to stochastic systems, which evolve probabilistically in time. From the point of view of classical physics, all systems are, strictly speaking, deterministic. But some systems (such as a roomful of air molecules with $\sim 10^{23}$ degrees of freedom) are so complicated that it is virtually impossible to specify their states exactly, and so a stochastic description is necessary.

On the other hand, we point out that although a system may in principle have many degrees of freedom available to it, not all of them are necessarily used. A macroscopic fluid technically has $\sim 10^{23}$ degrees of freedom, and a continuum model for the fluid will have an infinite number of degrees of freedom (though in a Navier-Stokes fluid, viscosity limits the continuum degrees of freedom to an upper bound which increases with the Reynolds number [23]). There are *many* microscopic degrees of freedom, but in the case of laminar flow, for instance, the microscopic motion is sufficiently collective that only a few degrees of freedom are needed to describe the macroscopic motion.

1.1.2 Sensitive dependence on initial conditions

A fundamental property of deterministic nonperiodic flow is its sensitive dependence on initial conditions. If the initial state of the system is known absolutely and with infinite precision, then the future of the deterministic system can be predicted with absolute precision. But if there is any uncertainty

at all in the initial condition, then that uncertainty is magnified with time. For this reason, the long-term future of such systems is essentially unpredictable.

For this reason also, stochastic methods are often useful in analyzing such systems. This may seem paradoxical at first, for it is because these nonlinear systems are *not* stochastic that they are interesting. However, as the uncertainty in the initial conditions grows with time, the system effectively "forgets" its distant past, and states of the system well separated in time act as though statistically independent.

Lorenz, who was a meteorologist, found that his extremely simplistic model of the weather exhibited a fundamental unpredictability. It is reasonable to speculate that a more realistic model would be at least as unpredictable as Lorenz's model (and that real weathermen should not be held too much to blame for their legendary inaccuracy). By showing that predictability requires more than identifying and enumerating all the degrees of freedom of a system, Lorenz effectively closed a door on the hope that a big enough computer could "solve" weather, and opened the door to a new branch of physics: deterministic chaos.

1.2 Dynamical Systems

A *dynamical* system is a system whose state may change in time. If the change from one state to another is governed by probabilistic considerations, the system is stochastic. A *deterministic* dynamical system is a dynamical system whose state in the future can be predicted from its state in the past.

1.2.1 Evolution operators

In particular, a deterministic dynamical system has an *evolution operator* f, which specifies how the system changes in time. If X_0 is the state of the system at some initial time $t=0$, then the state at time t is given by $X_t = f^t X_0$. Clearly, $f^{s+t} = f^s f^t$. Here we will give some examples of deterministic dynamical systems.

1.2.1.1 Ordinary Differential Equations

A system of m autonomous ordinary differential equations (ODE's) provides the archetypal model for a dynamical system. The state is represented by the m dimensional vector $\vec{x} = (x_0, x_1, \ldots, x_{m-1})$, and the dynamical time evolution is given by a system of ODE's

$$\frac{d\vec{x}}{dt} = \vec{A}(\vec{x}). \tag{1.1}$$

In terms of the evolution operator, we write $\vec{x}(t) = f^t(\vec{x}(0))$, where $\vec{x}(t)$ is the solution of the ODE. The standard example due to Lorenz [23] is:

$$\begin{aligned} \frac{dx}{dt} &= -\sigma(x+y) \\ \frac{dy}{dt} &= -xz + rx - y \\ \frac{dz}{dt} &= xy - bz. \end{aligned} \tag{1.2}$$

A typical trajectory, for the parameter values $\sigma = 10$, $b = 8/3$, and $r = 28$, is shown in Figure 1.1.

1.2.1.2 Maps

By considering the operator f acting at discrete intervals Δt of time, we obtain the map $f \equiv f^{\Delta t}$. We usually take $\Delta t = 1$, so

$$\vec{x}_n = f(\vec{x}_{n-1}) = ff \ldots f(\vec{x}_0) = f^n(\vec{x}_0). \tag{1.3}$$

A particularly well-studied example is the Hénon map [24]:

$$\begin{aligned} x_{n+1} &= 1 - ax_n^2 + y_n \\ y_{n+1} &= bx_n. \end{aligned} \tag{1.4}$$

Here the state of the system at time $t = n$ is given by the pair (x_n, y_n). Knowing the state at time n, the map (1.4) gives the state of the system at time $n+1$. Reapplication of the map to (x_{n+1}, y_{n+1}) gives the state at time $n+2$. By iterating the map, we are able to produce the entire future of the system.

Figure 1.2 shows a typical trajectory of this map with Hénon's original values of $a = 1.4$ and $b = 0.3$.

1.2.2 Trajectories through state space

A dynamical system's *state space* (often called *phase space*) is the space whose axes represent the state variables. Since the state of a system is given by the values of the state variables, a state can be represented by a single point in state space. For instance, if m variables specify the state, the state space is $\mathbb{R}^m$.

As a system evolves in time, its state changes and so does its position in

state space. The trajectory of this position through state space describes the evolution of the system with time.

1.2.3 Conservative and dissipative dynamical systems

We classify dynamical systems into conservative and nonconservative systems, according to whether or not Liouville's theorem applies. If $E \subset \mathbb{R}^m$ is a bounded subset of the state space, then $f^t(E) \subset \mathbb{R}^m$ is the set after it has dynamically evolved for a time $t>0$. Let V(E) be the ordinary (Lebesgue) volume of E. Then, a system is conservative if volume is preserved under forward [25] evolution in time; that is,

$$V(E) = V(f^t(E)). \tag{1.5}$$

We note that for conservative systems, the ordinary volume V is invariant under evolution f. We say in this case that V is an "invariant measure."

A nonconservative system is said to be dissipative if the volume of a set E decreases as it evolves in time; that is, $V(f^t(E)) \leq V(E)$, at least for long enough time t. As an example, the Lorenz Equations (1.2) are dissipative since

$$\frac{1}{V}\frac{dV}{dt} = \frac{1}{V}\int_E \frac{\partial(\frac{dx}{dt})}{\partial x} + \frac{\partial(\frac{dy}{dt})}{\partial y} + \frac{\partial(\frac{dz}{dt})}{\partial z} = -\sigma - 1 - b < 0 \tag{1.6}$$

for σ and b positive. Usually, the decrease in volume is exponential, that is $\frac{dV}{dt} \sim -kV$, so $V(f^t(E)) \sim e^{-kt}$, and in particular $V(f^t(E)) \to 0$ as $t \to \infty$. That the volume of a chunk of state space decreases toward zero as $t \to \infty$ suggests that all the trajectories in that chunk of state space are being attracted to

something, that $f^t(E) \to \mathcal{A}$ as $t \to \infty$ for some zero volume attracting set $\mathcal{A}$.

1.3 Attractors

Intuitively, we think of the attractor as a set of points in state space toward which trajectories in a dissipative dynamical system ultimately converge. There is obvious interest in the nature of the attractor, for this describes the asymptotic behavior of the system. In this section we will show some examples of "simple" and of "strange" attractors, and then we will proceed with a formal definition of what an attractor is and what properties we expect it to have.

1.3.1 Simple Attractors

Before giving a complete definition of what an attractor is, we provide two simple examples, shown in Figure 1.3. The first is the stable static equilibrium of the dissipative dyamical system

$$\begin{aligned} \frac{dx}{dt} &= -x+y \\ \frac{dy}{dt} &= -x-y. \end{aligned} \tag{1.7}$$

As all trajectories spiral toward the origin, the origin is said to be the attractor.

In the dynamical system

$$\begin{aligned} \frac{dx}{dt} &= -x(x^2+y^2-1) - y \\ \frac{dy}{dt} &= -y(x^2+y^2-1) + x, \end{aligned} \tag{1.8}$$

trajectories converge not toward a single point but toward a limit cycle. The attractor in this case is the unit circle centered on the origin.

We note in both cases that the volume of the attractor in state space is zero. The *dimensions* of the attractors for the systems defined in Equations (1.7) and (1.8) are zero and one, respectively.

1.3.2 Strange and chaotic attractors

A strange attractor is an attractor that is more complicated than a fixed point or a limit cycle. Often these volume-zero attractors have fractional dimension and complicated self-similar structure.

Most strange attractors are chaotic [26]. An attractor is said to be chaotic if nearby trajectories diverge. Since volume is always decreasing in the state space of a dissipative dynamical system, it seems at first that the attractor is the place where trajectories would converge, and this is certainly the case for stable fixed points. But Figure 1.4 shows how volume may be reduced even as nearby trajectories are diverging. We see that trajectories are converging in some directions and diverging in others. The convergence is stronger than the divergence, so the overall volume is shrinking. But if the vector separating two nearby trajectories has a nonzero component in the direction of divergence, then those two trajectories will eventually diverge. Thus, if there is *any* direction of trajectory divergence, then *most* of the trajectories will diverge.

1.3.3 Formal definition of an attractor

Informally, the attractor is what trajectories converge toward. Following Guckenheimer and Holmes [27], we begin our formal definition of an attractor with the concept of an ω-limit set.

The ω-limit set of x for f is the set of accumulation points of $f^t(x)$ as $t\to\infty$, where Y is an accumulation point of $f^t(x)$ as $t\to\infty$, if there is a sequence $t_i\to\infty$ such that $f^{t_i}(x)\to Y$.

Since ω-limit sets are defined for each initial condition X, a given evolution f may have many ω-limit sets. A more global notion of an attractor is provided by Ω, the "nonwandering set."

A point x is *nonwandering* if for every neighborhood U of x and every time T, there is a $t>T$ such that $f^t(U)\cap U\neq\emptyset$. That is, the trajectory for which x is the initial condition eventually comes back arbitrarily close to x. The nonwandering set Ω is the set of all nonwandering points x.

It can be shown that Ω is closed and that it contains all the ω-limit sets. Also, Ω is an invariant set; that is, $f^t(\Omega)=\Omega$ for all t. We do not yet want to call Ω an attractor, because it may be composed of several distinct attracting units. The usual decomposition of Ω is into "maximal topologically transitive sets." We define a closed invariant set Λ to be topologically transitive if it contains a dense orbit of f (that is, $\exists x\in\Lambda$ such that the following holds: $\forall y\in\Lambda$ and $\forall\epsilon>0$, $\exists t>0$ such that $f^t(x)$ is within ϵ of y). A topologically transitive set is maximal if there is no larger topologically transitive set that contains it. If $\mathcal{A}$ is a subset of Ω and is a maximal topologically transitive set, then we say that $\mathcal{A}$ is an attractor.

Here we mention that this is not the only formal definition that has been

proposed for attractors. A more general definition involves "chain recurrence"; we refer the reader to [27]. The important properties are just that the attractor $\mathcal{A}$ be invariant under evolution and that it contain a dense orbit.

1.3.4 Natural invariant measure

In this section, we define another measure on the state space, one that is appropriate specifically to the attractor and that is usually much more complicated than the ordinary Lebesgue volume mentioned in §1.2.3. This new measure, $\mu(E)$, is tied to the long term dynamics on the attractor; it measures how often and for how long the set E is visited by an orbit on the attractor.

For a specific orbit $\left\{f^n(X_o)\right\}_{n=0}^{\infty}$, we can define a measure

$$\mu(E) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathcal{F}(f^n(X_o)), \tag{1.9}$$

where

$$\mathcal{F}(X) = \begin{cases} 1 & \text{if } X \varepsilon E \\ 0 & \text{otherwise.} \end{cases} \tag{1.10}$$

We note that $\mu(E)=1$ if $\mathcal{A}\subseteq E\subseteq \mathbb{R}^m$ and that the measure is invariant under evolution f. That is,

$$\mu(f^{-t}(E)) = \mu(E),\ t>0, \tag{1.11}$$

where E is a subset of $\mathbb{R}^m$ and $f^{-t}(E)$ is the set of points obtained by backward evolution of the points in E during time t [28].

We also note that this measure is manifestly dependent on the initial

condition X_o. We write $\mu(E;X_o)$ to express this dependence explicitly. For instance, if X_o is a fixed point of f, then the orbit $\{f^n(X_o)\}=\{X_o\}$ induces the trivial measure

$$\mu(E;X_o) = \begin{cases} 1 & \text{if } X_o \varepsilon E \\ 0 & \text{otherwise.} \end{cases} \tag{1.12}$$

If X_o is a stable fixed point, *i.e.*, an attractor, then orbits that begin near the point X_o approach X_o, and the measure in Equation (1.12) will be valid for *all orbits* with initial conditions near X_o. That is,

$$\mu(E;X) = \begin{cases} 1 & \text{if } X_o \varepsilon E \\ 0 & \text{otherwise,} \end{cases} \tag{1.13}$$

for all X near X_o. On the other hand, if X_o is an *unstable* fixed point, then the static orbit is exceptional, and the μ defined in Equation (1.13) will not be appropriate for most X.

The "natural" invariant measure of an attractor is the measure that *is* appropriate for most initial points X. It is the measure that is induced by "typical" orbits that are dense on the attractor.

Given the natural invariant measure μ, we note that there is an ergodic property for orbits on the attractor [29]. For a function $\mathcal{F}(X)$, we define the μ-weighted spatial average [30]

$$\langle \mathcal{F} \rangle \equiv \int_{\mathbb{R}^m} \mathcal{F}(X)\, d\mu(X) \tag{1.14}$$

and note that for almost all initial points X_o,

$$\langle \mathcal{F} \rangle = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathcal{F}(f^n(X_o)). \tag{1.15}$$

It is this equality, of a spatial average on the left-hand side and a time average on the right-hand side, that we refer to when we speak of an "ergodic" property for an attractor.

Given an invariant measure μ, the *invariant distribution*, or sometimes, *invariant density*, $\rho(X)$ is defined so that for any set E,

$$\mu(E) = \int_E \rho(X)\, dX. \tag{1.16}$$

We point out that $\rho(X)$ may not be a function, *per se* (in the sense that the Dirac delta function is not a "function"), but the notation can simplify expressions involving the natural invariant measure. For instance, we can recast Equations (1.14) and (1.15) as

$$\int_E \mathcal{F}(X)\, \rho(X)\, dX = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathcal{F}(f^n(X_o)), \tag{1.17}$$

where E is any set for which $\mathcal{A} \subseteq E \subseteq \mathbb{R}^m$.

For most purposes, when we speak of the attractor, we refer to the invariant measure over the embedding space, typically $\mathbb{R}^m$. The attractor itself, the set $\mathcal{A}$, is the subset of $\mathbb{R}^m$ over which the measure is "nonzero." $\mathcal{A}$ is the "support" of the measure μ, which is defined formally as the complement of the union of all open sets for which the measure is zero.

1.4 Time series

An experimentalist, confronted with a physical dynamical system, measures at regular and discrete intervals of time the value of some state variable of the system (such as temperature or density or voltage) and records the time series: x_0, x_1, x_2, ... Many variables may be necessary to describe the state of the system; a time series represents the value of just one of the state variables (or more generally, a one-dimensional projection from the state space) as a function of time.

As such, the time series is an incomplete description of the system in its time evolution. On the other hand, many properties of the system can be inferred from the time series.

The time series depends on the initial state of the system; it is thus useful to take the mathematical viewpoint that the time series at hand is just one from an ensemble of time series, each corresponding to a different initial condition.

1.4.1 Time delay coordinates

Time series are manifestly one-dimensional. A time series describes the evolution of a single state variable, even though several variables are usually required to fully describe the state of a system. Packard, Crutchfield, Farmer, and Shaw [31] devised a delay 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 are created,

$$v_i = (x_i, x_{i-T}, \cdots, x_{i-(m-1)T}) \varepsilon \mathbb{R}^m, \tag{1.18}$$

with the delay time T and the embedding dimension m parameters of the embedding procedure. Here v_i represents a (possibly) more complete description of the state of the system at time t=i than does x_i. The dynamical information in the one-dimensional time series has been "converted" to geometric information in $\mathbb{R}^m$.

Takens [32] placed the time-delay embedding on solid mathematical footing by showing that this procedure does (almost always) reconstruct the state space of a dynamical system. To clarify this statement, we introduce some formalism.

Let $X_i \varepsilon \mathbb{R}^M$ denote the state of the system in its original (M dimensional) state space, where M may be arbitrarily large (even infinite). Let the map $\pi:\mathbb{R}^M \to \mathbb{R}$ correspond to the projection of the original system to the time series; that is, $\pi(X_i)=x_i$ is the "measurement" taken at time t=i. We note that π is in general a noninvertible many-to-one map. From π, the map $\pi_m:\mathbb{R}^M \to \mathbb{R}^m$ is constructed by the time-delay formula in Equation (1.18), so that $\pi_m(X_i)=v_i$. Here, π_m projects points in the original state space to the reconstructed state space, and particularly if M>m, we expect π_m to be many-to-one. However, if we restrict π_m to act on points in a d dimensional subspace, with $d \leq m$, and in particular to points in the d dimensional attractor $\mathcal{A}$, we may find that $\pi_m|_{\mathcal{A}}: \mathcal{A} \to \pi_m(\mathcal{A})$ is a continuous isomorphism (a one-to-one map). If this is so, the "real" attractor $\mathcal{A}$ will be topologically equivalent (isomorphic) to the reconstructed attractor $\pi_m(\mathcal{A})$.

What Takens has shown is that for large enough m, almost all projections

π lead to reconstructed attractors $\pi_m(\mathcal{A})$ that are isomorphic to the original attractor. This justifies the time-delay embedding scheme and enables us to probe the state space of a dynamical system from a one-dimensional time series.

1.4.1.1 Optimal reconstruction of the state space

Takens has shown that for virtually any T and for "large enough" m, an attractor can be reconstructed in $\mathbb{R}^m$. The obvious questions to ask are: How large is a large enough m, and what is the best T? This is something of a technical detail; the purpose of this section is merely to point to some of the work that has addressed these issues.

Mañé [33] has shown that no more than $m=2d+1$ dimensions are generically required to isomorphically embed a d dimensional attractor into $\mathbb{R}^m$. Eckmann and Ruelle [34], however, point out that as long as $m \geq d$, the reconstructed attractor $\pi_m(\mathcal{A})$, while not necessarily isomorphic to $\mathcal{A}$, will (almost always) have the same dimension.

The theorems tell us that in the limit of infinitely long ($N \to \infty$) time series, any choice of $T \geq 1$ and $m \geq d$ will reconstruct an attractor of the correct dimension. This is important to know, but another thing we'd like to know that the theorems do not tell us is this: What are the *best* (or for that matter, what are "good enough") values of T and m to use if our goal is to *estimate* the unknown d from a finite time series? The issue of optimum T has been addressed by Fraser and Swinney [35]; also by Havstad and Ehlers [36]. A variant of the embedding procedure with many more optimal parameters (essentially, each component is a linear combination of the delayed time series

values) was introduced by Broomhead and King [37].

Not knowing the dimension d of the attractor beforehand, we estimate the dimension d(m) of our reconstructed attractors for various m, increasing m until the estimated dimension d(m) saturates at some $d(m)<m$. Some authors [38] suggest increasing m until $m \geq 2d(m)+1$ as a safe rule of thumb, this presumably having been inspired by Mañé's theorem. But this is incorrect, because Mañé's theorem says nothing about *estimation* of dimension from a *finite* sample of N points.

1.4.2 Deterministic time series

Primarily, we are concerned with deterministic dynamical systems. If X_n is the full state of the system at time $t=n$, then we have $X_{n+1}=f(X_n)$, where f is the evolution operator. From the time series, on the other hand, we cannot usually predict the "next" measurement $x_{n+1}=\pi(X_{n+1})$ just from the most recent measurement $x_n=\pi(X_n)$. We say that a *time series* is deterministic if there is some function f for which

$$x_{n+1} = f(x_n, x_{n-1}, x_{n-2}, \ldots). \tag{1.19}$$

Given f and a sufficiently long list of past values x_n, x_{n-1}, x_{n-2}, ..., it is possible to generate an arbitrarily long time series.

Going the other way, trying to guess what the function f is *from the time series*, requires a certain amount of imagination [39,40,41]. But once f is known, then precise predictions of the future, *given precise measurements of the past*,

are possible.

It is not at all obvious that a deterministic dynamical system should give rise to a deterministic time series, yet Takens' demonstration that there is (almost always) an isomorphism between the full description X_n of the state and a set of consecutive values of the projected time series $(x_n, x_{n-1},...)$ shows, in fact, that these two senses of the word *deterministic* are essentially equivalent. A deterministic dynamical system generically gives rise to a deterministic time series.

1.4.3 Stochastic time series

We contrast these deterministic systems to systems that are stochastic. White noise is an important example of a purely stochastic system. In a white noise time series, $\{\epsilon_n\}$, each ϵ is chosen *independently* of its predecessors from a random distribution, which is typically, though not necessarily, a Gaussian. Predicting the future of a white noise time series is hopeless; successive values of ϵ in no way depend upon previous values.

More general stochastic time series combine white noise with some determinism. Here we write

$$x_n = f(x_{n-1},x_{n-2},...) + \gamma\epsilon_n, \tag{1.20}$$

with ϵ_n from a standard white noise time series. γ is the noise level of the time series; if γ is small enough, then the time series is essentially deterministic and analysis and predictions can be done. We note that even our best prediction of

the future will be in error by γ on average.

We will not have much to say about the effect of noise directly on nonlinear dynamical systems, though it is an interesting issue. The reader is referred to [42].

In Chapter Five, we do discuss the effect of additive noise in the *measurement* of a dynamical system. In constrast to Equation (1.20) above, the system x_n proceeds noiselessly according to Equation (1.19), but we measure $x_n + \gamma\epsilon_n$.

Occasionally, we find stochastic systems a useful tool for modeling certain effects that occur in analyzing deterministic systems. In Chapter Thirteen, for example, we mimic nonlinear deterministic time series with equations of the form (1.20) in which f is a *linear* function of its arguments, and the nonlinearity is taken up in the stochastic term.

1.5 Quantifying chaos: numerical diagnostics

Having embedded the time series into $\mathbb{R}^m$, we are now in a position to characterize the orbits through the reconstructed state space. There are two main approaches. We can seek out some measure of the divergence of nearby trajectories, such as the Kolmogorov entropy and the Lyapunov exponents; or we can determine how much of this embedding space is actually occupied by trajectories of the dynamical system (after transients). Do they fill out the full m dimensional space (as the trajectories of a stochastic system, for instance, would be expected to do) or are they restricted to a d dimensional attractor, with $d < m$ and d not necessarily an integer?

Both are useful measures of chaos, and both can be estimated numerically. We emphasize that these diagnostics are fully numerical, in that they use no information about the system that produced the time series. Indeed, from the point of view of these algorithms, the time series *is* the system.

1.5.1 Kolmogorov entropy and Lyapunov exponents

These are dynamical measures of chaos, in contrast to the more geometric dimension. The Kolmogorov entropy and Lyapunov exponents quantify the sensitivity of a dynamical system to initial conditions.

The Kolmogorov entropy K measures the rate at which information is "created" by the dynamical system. What is meant by this is the following. Given the evolution operator f and knowing the state of a system to an accuracy of, say, k bits, we can predict its state some time t in the future, though to less accuracy, say k−s bits. On the other hand, if we just wait that time t and observe the system to our usual accuracy of k bits, then that observation will net us s more bits of information. In this sense, the dynamical system creates an average of K=s/t bits per unit time.

The Lyapunov exponents measure the rate at which nearby trajectories diverge. As Figure 1.4 shows, this rate depends on which pair of nearby trajectories is being considered. In $\mathbb{R}^m$, then, there are m distinct rates. If $\vec{x}(t)$ is the (reconstructed) state of the system at time t, and f is the evolution operator, then the Jacobian matrix is given by the linearization of f,

$$J(t) = \frac{\partial f^t(\vec{x}(0))}{\partial \vec{x}(0)} = \frac{\partial \vec{x}(t)}{\partial \vec{x}(0)}; \tag{1.21}$$

that is,

$$J_{ij}(t) = \frac{\partial x_i(t)}{\partial x_j(0)}. \tag{1.22}$$

The *nth* Lyapunov exponent is given by

$$\lambda_n = \lim_{t\to\infty} \frac{1}{t} \log \left| nth\ eigenvalue\ of\ J(t) \right|. \tag{1.23}$$

The largest Lyapunov exponent is clearly the most important since if a pair of trajectories has any nonzero component in the eigendirection associated with the largest eigenvalue, then that rate of divergence will dominate. The Kolmogorov entropy K turns out to be given by the sum of the positive Lyapunov exponents [45].

This thesis will not have much to say about numerical estimation of Lyapunov exponents and Kolmogorov entropy from a time series. The issue is discussed in some detail by Wolf, *et al.* [46]; also, more recently in [47]. A method for estimating Kolmogorov entropy that is related to the correlation algorithm is developed in [48] and [49].

1.5.2 Dimension

Dimension is a geometric measure; it tells how much of the state space is explored by the system in its post-transient state (that is, how much of the state space is occupied by the system's attractor). Numerical estimation of dimension (in particular the correlation dimension of Takens [50] and Grassberger and Procaccia [51]) will be our primary focus. This thesis will explore criteria for determining whether and to what extent the correlation dimension algorithm

can be trusted.

Our emphasis will be on the practical. Because the "output" of the correlation algorithm is not a single number but a log-log plot of the correlation integral — or, less formally, the "C of r curve" — and because the output is often so crucially dependent on the various input parameters, the analysis has traditionally been a very human process, with a little bit of art mixed in with the science. An important concern will be with determining the optimal scaling range, which is the range of r over which C(r) is proportional to r^{ν}; equivalently, the range over which the slope of a *log* C(r) versus *log* r curve provides the best estimate of the attractor dimension ν. There is always a danger in seeking slopes on log-log plots; they often are *too* easy to find: it is an old and wise if somewhat sexist adage that on a log-log plot, even Sophia Loren lies on a straight line.

We will examine various features of strange attractors that can affect our interpretation of the C of r curve and poison our estimate of dimension. By choosing simple examples we hope to isolate and quantify the consequences of these features, so that we can diagnose symptoms and provide antidotes. We classify these effects into three main groups: geometrical, dynamical, and statistical.

1.5.2.1 Geometrical effects

Geometrical effects arise in cases of imperfect self-similarity. Points, lines, and planes look the same on all scales. Any magnification of a line still looks like a line. For many objects, however, this self-similarity is not so

precise.

For instance, a circle, looked at globally, looks like a circle; but looking at a small part of the circle, it looks like an arc, and looking closely at a very tiny part of the circle it looks like a line. If we look at an ϵ-sized chunk of the circle, then as $\epsilon \to 0$, we see what looks more and more like a self-similar line. This kind of effect is due to the finite size of the attractor, or that the attractor has an "edge." We discuss the edge effect in Chapter Six.

In Chapter Five, we point out that the effect of adding noise to the system is to prevent the $\epsilon \to 0$ limit from being taken; also in that chapter, we show that the effect of discretization on the data (which also limits the smallest scale at which the attractor can be observed) can, to leading order, be counteracted.

Chapter Seven discusses the kind of effect seen in a correlation integral when the self-similarity is discrete, as it is in most strange sets. We note for instance that the standard Cantor set exhibits self-similarity only on scales that are factors of 3 from each other.

A few basic geometrical properties of the correlation integral are introduced in Chapter Three.

1.5.2.2 Dynamical effects

Dynamical effects arise from the nonrandomness in the way that points are placed on an attractor. Points on a highly chaotic attractor are distributed in a way that, although strictly deterministic, approximates randomness. When this approximation is poor, the performance of the correlation algorithm is

compromised.

Chapter Eight describes the problems involved with using the correlation algorithm to measure the dimension of quasiperiodic attractors. Basically, since quasiperiodic attractors are not chaotic, points will be distributed *too* uniformly, and the correlation integral will not be able to sample the full range of scales that would be available to it if the points were distributed more randomly.

Chapter Nine considers systems with autocorrelation (which is seen both in nonwhite noise and in deterministic chaos). The problem is that although pairs of points well separated in time act as though statistically independent, pairs that are temporally near to each other tend not to be independent, and the statistical assumptions that underly the correlation integral fail. We point out specifically what this does to the correlation integral and we suggest a remedy.

1.5.2.3 Statistical effects

Many of the problems that occur in attempting to measure the dimension of a strange attractor from a set of discrete points on the attractor are problems that fade as the number of points on the attractor is increased. In practical situations, of course, only a finite number N of points are available.

Chapter Ten is devoted to the issue of statistical error due to a finite number of points.

Chapter Eleven introduces a specific model in which the interaction of finite N and the edge effect are investigated as a function of attractor dimension. We find that the number of points needed scales exponentially with the dimension of the attractor. Chapter Eleven also provides a strategy for

estimating the optimal scaling range of the correlation integral.

Chapter Four outlines the fundamental assumptions and ensembles that underlie the statistical analysis of motion on a strange attractor.

1.5.3 Application

Chapter Twelve is concerned with schemes to speed up computation of the correlation integral, by invoking either a faster (parallel) computer or a faster algorithm. In particular, our "box-assisted correlation" algorithm can compute the full range of C(N,r) in O(N log N) time, a significant improvement over the usual $O(N^2)$ method.

Linear methods for the analysis of time series are introduced in Chapter Thirteen. We discuss what contribution these methods may make to the understanding of nonlinear systems, arguing that their (albeit limited) power has not been fully exploited in this regard.

In Chapter Fourteen, we finally apply these methods to particular physical systems. No dimension is observed for time series obtained from tokamak plasma machines. We conclude that tokamak turbulence cannot be characterized by low-dimensional chaos.

1.6 Notes and References

[1] Jean-Michel Grandmont and Pierre Malgrange. "Nonlinear economic dynamics: introduction," J. Econ. Theory 40 (1986) 3. This is the lead article in a series on nonlinear economics to which this issue (Volume 40, Number 1, October 1986)

is devoted.

[2] Willam M. Schaffer and Mark Kot. "Nearly one-dimensional dynamics in an epidemic," J. Theor. Bio. 112 (1985) 403.

[3] A review is given by N. B. Abraham, J. P. Gollub, and Harry L. Swinney. "Testing nonlinear dynamics," Physica 11D (1984) 252.

[4] R. S. Gioggia and N. B. Abraham. "Routes to chaotic output from a single-mode, dc-excited laser," Phys. Rev. Lett. 51 (1983) 650; R. S. Gioggia and N. B. Abraham. "Anomalous mode-pulling, instabilities, and chaos in a single-mode, standing-wave 3.39 micron HeNe laser," Phys. Rev. A 29 (1984).

[5] F. A. Hopf, D. L. Kaplan, M. H. Rose, L. D. Sanders, and M. W. Derstine. "Characterization of chaos in a hybrid optically bistable device," Phys. Rev. Lett. 57 (1986) 1394.

[6] J. Testa, J. Perez, and C. Jeffries. "Evidence for universal chaotic behavior of a driven nonlinear oscillator," Phys. Rev. Lett. 48 (1982) 714; James Testa. "Fractal dimension at chaos of a quasiperiodic driven tunnel diode," Phys. Lett. 111A (1985) 243.

[7] Carson D. Jeffries. "Chaotic dynamics of instabilities in solids," Phys. Scr. T9 (1985) 11.

[8] G. Gibson and C. Jeffries. "Observation of period doubling and chaos in spin wave instabilites in yttrium iron garnet," Phys. Rev. A 29 (1984) 811.

[9] E. G. Gwinn and R. M. Westervelt. "Scaling structure of attractors at the transition from quasiperiodicity to chaos in electronic transport in Ge," Phys. Rev. Lett. 59 (1987) 157; S. W. Teitsworth, R. M. Westervelt, and E. E. Haller. "Non-linear oscillations and chaos in electrical breakdown in Ge," Phys. Rev. Lett. 51 (1983) 825.

[10] A review of chaos in biology is given by P. E. Kloeden and A. I. Mees. "Chaotic phenomena," Bull. Math. Biol. 47 (1985) 697.

[11] L. Glass, M. R. Guevara, A. Shrier, and R. Perez. "Bifurcation and chaos in a periodically stimulated cardiac oscillator," Physica 7D (1983) 89.

[12] A. Babloyantz, J. M. Salazar, and C. Nicolis. "Evidence of chaotic dynamics of brain activity during the sleep cycles," Phys. Lett. 111A (1985) 152; Ivan Dvorak and Jaromir Siska. "On some problems encountered in the estimation of the correlation dimension of the EEG," Phys. Lett. 118A (1986) 63; Gottfried Mayer-Kress and Scott P. Layne. "Dimensionality of the human electro-encephalogram," preprint (1986).

[13] K. Aihara, G. Matsumoto, and M. Ichikaura. "An alternating periodic-chaotic sequence observed in neural oscillators," Phys. Lett. 111A (1985) 251.

[14] J. L. Martiel and A. Goldbeter. "Autonomous chaotic behavior of the slime mould *Dictyostelium discoideum* predicted by a model for cyclic AMP signalling," Nature 313 (1985) 590.

[15] J. C. Roux, R. H. Simoyi and H. L. Swinney, "Observation of a strange attractor," Physica 8D (1983) 257.

[16] Peter Grassberger. "Do climatic attractors exist?" Nature 323 (1986) 609. Evidently, they do not.

[17] R. P. Behringer. "Rayleigh-Bénard convection and turbulence in liquid helium," Rev. Mod. Phys. 57 (1985) 657.

[18] A. Brandstater and Harry L. Swinney. "Strange attractors in weakly turbulent Couette-Taylor flow," Phys. Rev. A 35 (1987) 2207.

[19] R. S. Shaw. *The Dripping Faucet as a Model Chaotic System* (Aerial Press, Santa Cruz, 1985).

[20] S. J. Zweben and R. W. Gould. "Scaling of edge-plasma turbulence in the Caltech tokamak," Nucl. Fusion 23 (1983) 1625.

[21] Strange attractors *have* been observed for less-than-fully-developed turbulence in plasmas. W. Arter and D. N. Edwards. "Non-linear studies of Mirnov oscillations in the DITE tokamak: evidence for a strange attractor," Phys. Lett. 114A (1986) 84; more recently, see P. Y. Cheung and A. Y. Wong, "Chaotic behavior and period doubling in plasmas," Phys. Rev. Lett. 59 (1987) 551. In the case of more fully developed turbulence, the strange attractor has not been observed: M. L. Sawley, C. W. Simm, and A. Pochelon, "The correlation dimension of broadband fluctuations in a tokamak," Phys. Fluids 30 (1987) 129.

[22] Edward N. Lorenz. "Deterministic nonperiodic flow," J. Atmos. Sci. 20 (1963) 130.

[23] This observation of Kolmogorov was pointed out to us by Paul Dimotakis.

[24] M. Hénon. "A two-dimensional mapping with a strange attractor," Comm. Math. Phys. 50 (1976) 69.

[25] Actually, due to a mathematical subtlety which we will not elaborate, this statement is more generally correct for *backward* evolution in time. For invertible f this does not matter, but if f has more than one inverse, then all inverses must be taken. Formally, we define backward evolution so that $f^{-t}(E)=\{X\in\mathbb{R}^m \mid f^t(X)=E\}$.

[26] However, we note that not *all* strange attractors are chaotic. See, for example, Filipe J. Romeiras and Edward Ott. "Strange nonchaotic attractors of the damped pendulum with quasiperiodic forcing," Phys. Rev. A 35 (1987) 4404.

[27] John Guckenheimer and Philip Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Volume 42 of *Applied*

Mathematical Sciences (Springer-Verlag, New York, 1983), 236.

[28] See Note 25.

[29] J.-P. Eckmann and D. Ruelle. "Ergodic theory of chaos and strange attractors," Rev. Mod. Phys. 57 (1985) 617.

[30] We recall that the Stieltjes integral, when it exists, takes on the value

$$\int_E \mathcal{F}(x)\, d\mu(x) = \lim_{\epsilon \to 0} \sum_{A \in \mathcal{P}(\epsilon,E)} \mathcal{F}(x \in A)\, \mu(A), \tag{1.24}$$

where $\mathcal{P}(\epsilon,E)$ is a partition of the set E into a disjoint union of sets whose diameter is less than ϵ.

[31] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. "Geometry from a time series," Phys. Rev. Lett. 45 (1980) 712.

[32] Floris Takens. "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick*, 1980, Volume 898 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1981).

[33] Ricardo Mañé. "On the dimension of the compact invariant sets of certain non-linear maps," in *Dynamical Systems and Turbulence, Warwick*, 1980, Volume 898 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1981).

[34] See Reference 29.

[35] Andrew M. Fraser and Harry L. Swinney. "Independent coordinates for strange attractors from mutual information," Phys. Rev. A 33 (1986) 1134.

[36] James. W. Havstad and Cindy L. Ehlers. "Attractor dimension from small data sets," preprint (1987).

[37] D. S. Broomhead and Gregory P. King. "Extracting qualitative dynamics from experimental data," Physica 20D (1986) 217.

[38] Joachim Holzfuss and Gottfried Mayer-Kress. "An approach to error-estimation in the application of dimension algorithms," in *Dimension and Entropies in Chaotic Systems*, Volume 32 of *Springer Series in Synergetics* (Springer-Verlag, Berlin, 1986).

[39] J. Doyne Farmer and John J. Sidorovich. "Predicting chaotic time series," Phys. Rev. Lett. 59 (1987) 845.

[40] Alan Lapedes, (CNLS, Los Alamos) has been working on approximating the prediction function with a neural network (*private communication*).

[41] Peter Grassberger. "Information content and predictability of lumped and distributed dynamical systems," preprint (1987).

[42] An introduction and annotated bibliography is given in Chapter 9 of Hao Bai-Lin. *Chaos* (World Scientific, Singapore, 1984).

[43] It can be shown that $K \leq \sum_{\lambda_n > 0} \lambda_n$ in all cases and that equality holds in some cases; see Reference 27.

[44] Alan Wolf, Jack B. Swift, Harry L. Swinney, and John A. Vastano. "Determining Lyapunov exponents from a time series," Physica 16D (1985) 285.

[45] J.-P. Eckmann, S. Oliffson Kamphorst, D. Ruelle, and S. Ciliberto. "Liapunov exponents from a time series," Phys. Rev. A 34 (1986) 4971.

[46] Peter Grassberger and Itamar Procaccia. "Estimation of the Kolmogorov entropy from a chaotic signal," Phys. Rev. A 28 (1983) 2591.

[47] Aviad Cohen and Itamar Procaccia. "Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems," Phys. Rev. A 31 (1985) 1872.

[48] Floris Takens. "Invariants related to dimension and entropy," in *Atas do* 13^o (colóqkio brasiliero de matemática, Rio de Janeiro, 1983).

[49] Peter Grassberger and Itamar Procaccia. "Characterization of strange attractors," Phys. Rev. Lett. 50 (1983) 346.

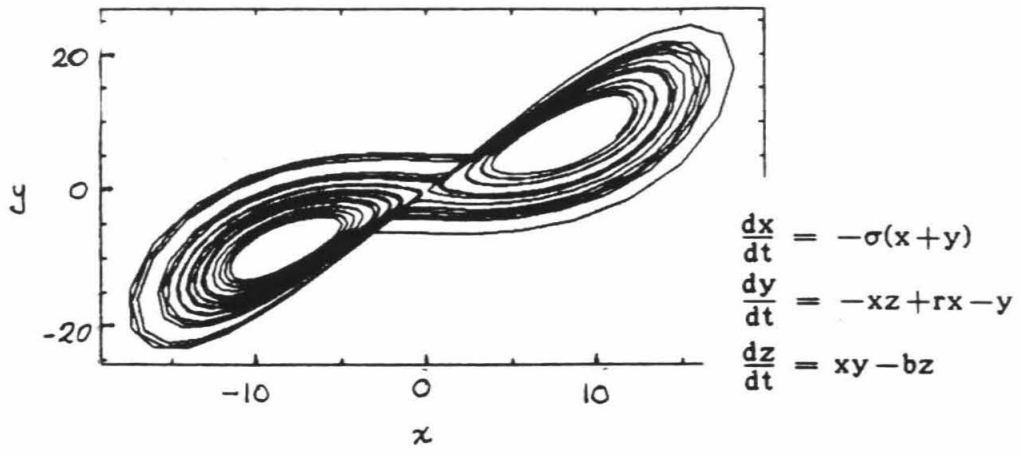


Figure 1.1. Typical trajectory of the Lorenz equations, projected onto the x-y axis. Here, $\sigma = 10$, $b = 8/3$, and $r = 28$.

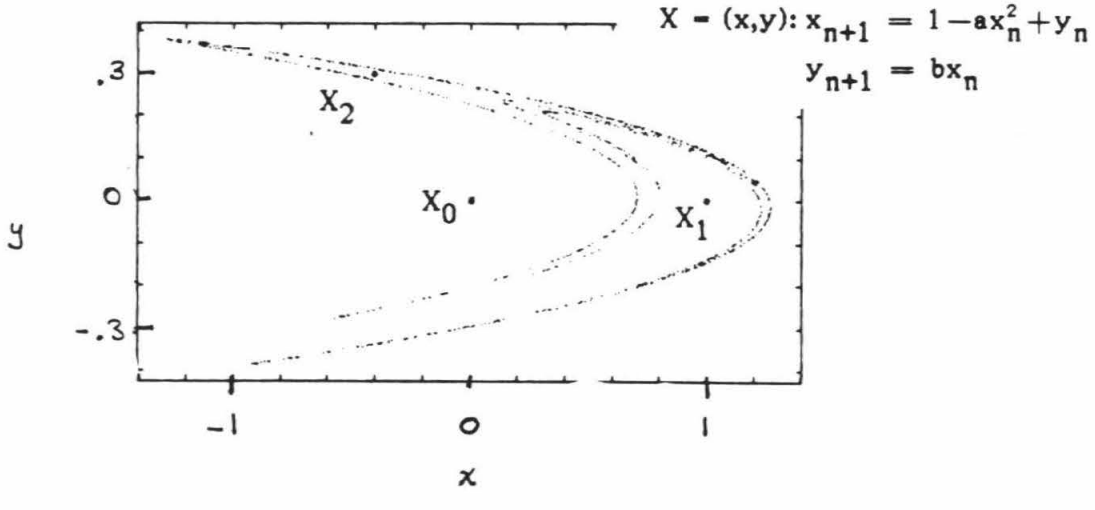


Figure 1.2 Typical trajectory of the Hénon map. The initial point (0,0) and first two iterates are labeled. The first 900 iterates are shown.

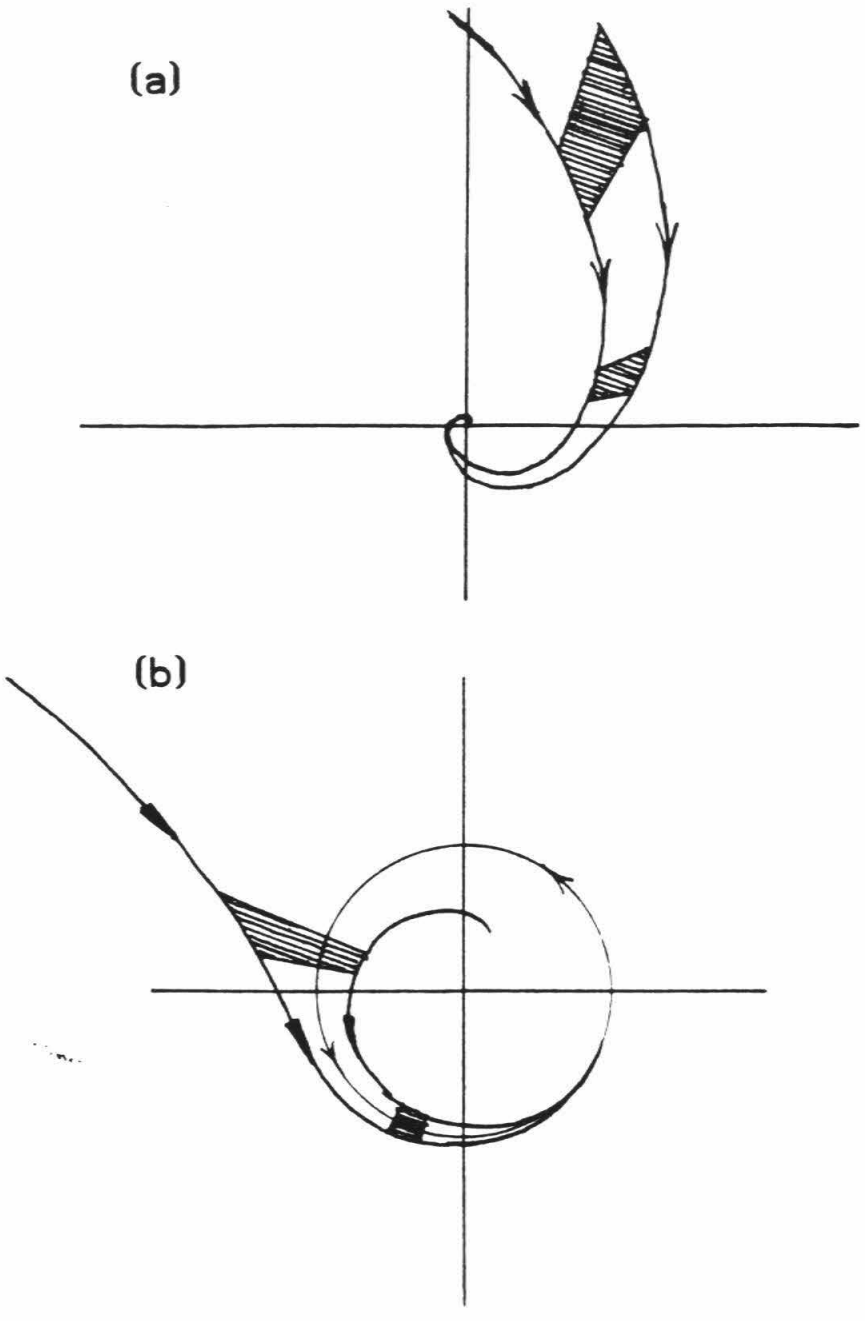


Figure 1.3 Two simple attractors: (a) limit point, (b) limit cycle. Note in both cases that a volume element shrinks as it evolves forward in time.

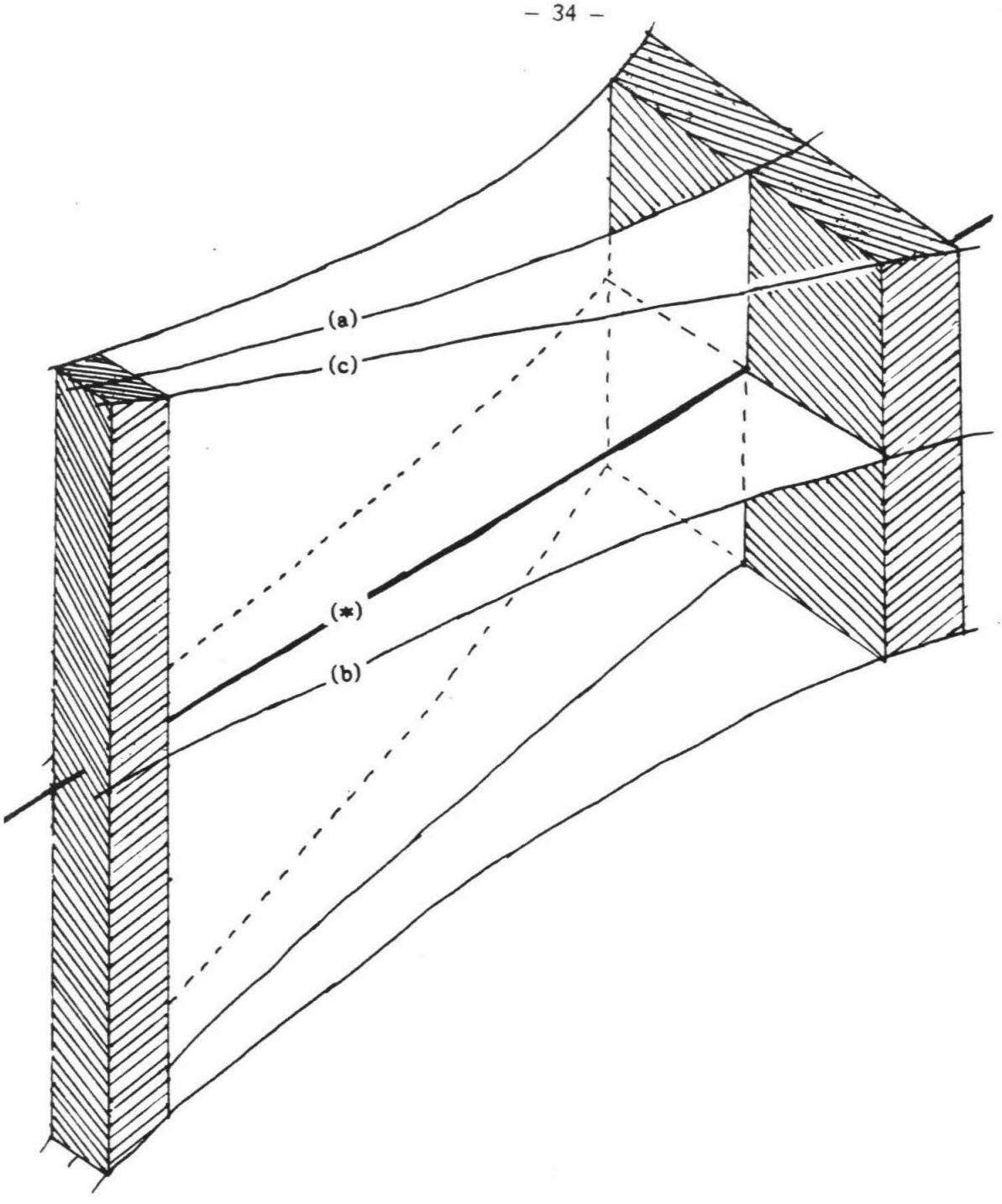


Figure 1.4 Trajectory (a) diverges from the fiducial trajectory $(*)_J$ while trajectory (b) converges toward (*). Since the convergence of (b) is "faster" than the divergence of (a), the volume decreases in time. A typical trajectory, such as (c), contains nonzero components in the directions of (a) and of (b), and ultimately diverges from (*).

CHAPTER TWO

2. OVERVIEW OF DIMENSION ALGORITHMS

The dimension of a set tells 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 is specified by three coordinates. For sets more complicated than lines, surfaces, and volumes, however, this informal definition of dimension needs to be extended.

One way to extend this definition is to determine not how many real numbers but how many bits of information are needed to specify a point to an accuracy of ϵ. On a line segment of unit length, k bits specify the position of a point to within $\epsilon = 2^{-k}$. 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. In general, $S(\epsilon) = -d \log_2 \epsilon$ bits of information specify the position on a unit d dimensional hypercube to an accuracy of ϵ. This leads to a natural definition for the "information dimension" of a set; it is given by the small ϵ limit of $-S(\epsilon)/\log_2 \epsilon$, where $S(\epsilon)$ is the information (in bits) needed to specify a point on the set to an accuracy ϵ.

In this extended definition (and we will discuss other definitions as well) the dimension need not be an integer; some sets can have fractional dimension. Among these are Cantor sets of various types, the fractal sets introduced by Mandlebrot [1], and the strange attractors seen in dynamical systems. It is the last of these that motivates the study in this thesis of practical means for

estimating dimension from a finite sample of points on the attractor.

Hausdorff [2] introduced the first rigorous definition of dimension, though it is a definition that does not immediately lend itself to numerical estimates. Other definitions have been developed that are more straightforwardly adaptable to numerical estimation from a sample of discrete points.

There are two types of dimension algorithms [3], those based on box-counting schemes and those based on distances between pairs of points. The box-counting algorithms are not widely used, having been superseded by the more efficient pairwise distance algorithms.

The notion of box-counting, however, is very useful for introducing the generalized dimensions, D_q. The Hausdorff and standard box-counting algorithms measure D_0, which depends only on the support of the invariant distribution $\rho(x)$; the generalized dimensions take account of the higher moments of $\rho(x)$.

2.1 Hausdorff dimension

We recall the definition of the Hausdorff dimension for the set $\mathcal{A}$. First, let $\mathcal{P}(\epsilon,\mathcal{A})=\{B_0, B_1, \ldots, B_{k-1}\}$ be a finite partition of the set $\mathcal{A}$ into disjoint sets whose diamaters are less than ϵ. That is, $B_i \cap B_j = \varnothing$ and $\bigcup_{i=0}^{k-1} B_i = \mathcal{A}$ and the diameters $\delta_i \equiv \mathrm{diam}(B_i)$ satisfy $\delta_i < \epsilon$. Then we define a partition function

$$m(\mathcal{A},d) \equiv \lim_{\epsilon \to 0} \sup \inf_{\mathcal{P}(\epsilon,\mathcal{A})} \sum_i \delta_i^d \ . \tag{2.1}$$

We find that there exists a value D_H for which $m(\mathcal{A},d)=0$ for $d>D_H$, and $m(\mathcal{A},d)=\infty$ for $d<D_H$. Thus, we define the Hausdorff dimension

$$D_H = \inf \{ d \mid m(\mathcal{A},d)=0 \}. \tag{2.2}$$

2.2 Box-counting algorithm

The Hausdorff dimension is difficult to implement numerically; a simpler and more direct definition is provided by the "capacity" or "box-counting dimension." For most "physically interesting" invariant sets of dynamical systems, the Hausdorff and the box-counting dimension are the same [4].

Break up the embedding space $\mathbb{R}^m$ into a grid of boxes of size ϵ; then count the number of boxes $n(\epsilon)$ that intersect the attractor. The capacity is defined [5]

$$d_0 = \lim_{\epsilon \to 0} \frac{-log\ n(\epsilon)}{log\ \epsilon}. \tag{2.3}$$

Given only a finite sample of N points on the attractor, we estimate $n(\epsilon)$ with $n(N,\epsilon)$, the number of boxes inside of which at least one of the sample points lies. Thus, $n(N,\epsilon) \leq n(\epsilon)$, and

$$n(\epsilon) = \lim_{N \to \infty} n(N,\epsilon). \tag{2.4}$$

Combining these, we have

$$d_0 = \lim_{\epsilon \to 0} \lim_{N \to \infty} \frac{log\ n(N,\epsilon)}{log(1/\epsilon)}. \tag{2.5}$$

The order of the limits is crucial. For finite N, $n(N,\epsilon)$ is bounded by N (there

can be no more nonempty boxes than there are things to put in the boxes), and $\epsilon \to 0$ gives a dimension of zero.

In a formal sense, this is not surprising: a finite set of points *does* have dimension zero. But from a practical point of view, this is meaningless: what we are after is an *estimate* of the dimension of the full set, of which the N points are but a representative sample.

Equation (2.3) implies a scaling of $n(\epsilon)$ with ϵ, namely, that $n(\epsilon) \propto \epsilon^{-d_o}$. However, we note that if $n(\epsilon)$ is of the form $n(\epsilon)=n_o\epsilon^{-d_o}$, then Equation (2.3) converges with logarithmic slowness.

$$\lim_{\epsilon \to 0} \frac{-log\ n_o\epsilon^{-d_o}}{log\ \epsilon} = \lim_{\epsilon \to 0} \frac{d_o\ log\ \epsilon\ -\ log\ n_o}{log\ \epsilon} = d_o - \lim_{\epsilon \to 0} \frac{log\ n_o}{log\ \epsilon} = d_o \qquad (2.6)$$

A more practical and efficient way to observe this scaling, rather than directly applying Equation (2.3) with small ϵ, is to plot $log\ n(\epsilon)$ versus $log\ \epsilon$. The slope of this curve will give $-d_o$ for small ϵ. See Figure 2.1. We are essentially invoking l'Hopital's rule, and taking instead of the limit in Equation (2.3) the limit of the derivatives

$$d_o = \lim_{\epsilon \to 0} - \frac{d[log\ n(\epsilon)]}{d[log\ \epsilon]}. \qquad (2.7)$$

If this limit exists [6], then l'Hopital's rule assures that it has the correct value. Furthermore, the logarithmic slowness in Equation (2.6) is corrected for, and convergence is much quicker.

There is, it turns out, a more serious limitation. For many attractors, the

limit in Equation (2.4) converges very slowly. And although schemes have been devised [7] to accelerate the convergence (essentially to extrapolate the limit from a sequence of $n(N,\epsilon)$ for increasing N), the box-counting method is generally regarded as impractical for numerical computation [8].

2.3 Generalized dimensions

Acknowledging the limited practical value of the box-counting algorithm, we do note its easy adaptability to generalized dimension. The Hausdorff dimension and the standard box-counting dimension are purely geometrical quantities, describing only the attracting set itself (the support, or in a sense, the "zeroth moment" of the invariant measure). They do not account for information in the invariant measure, which tells how often or for how long certain regions of the attractor are visited. The generalized dimension was introduced for nonlinear dynamics by Hentschel and Procaccia [9], though it is based on an entropy formalism developed by Renyi [10], which makes use of the higher moments of the invariant measure.

To see how these generalized dimensions work, again divide the embedding space into boxes of size ϵ. For every box i assign a value p_i associated with the fraction of the attractor in the ith box,

$$p_i = \frac{\mu(\text{box i})}{\mu(\mathcal{A})} = \frac{1}{\mu(\mathcal{A})} \int_{\text{box i}} \rho(X)\, dX. \tag{2.8}$$

Recall that $\mu(\mathcal{A})=1$, so the denominators in Equation (2.8) are not really necessary. Now, define

$$D_q = \frac{1}{q-1} \lim_{\epsilon \to 0} \frac{log \sum p_i^q}{log\ \epsilon} \tag{2.9}$$

and note that $q=0$ leads to the definition of the usual box-counting dimension. q may range from $-\infty$ to $+\infty$, but a dimension of much theoretical interest is the information dimension, which is given by $q=1$.

$$D_1 = \lim_{q \to 1} D_q = \lim_{\epsilon \to 0} \frac{\sum p_i\ log\ p_i}{log\ \epsilon} = \lim_{\epsilon \to 0} \frac{-S(\epsilon)}{log\ \epsilon}, \tag{2.10}$$

where $S(\epsilon)$ is the information content in the partition $\{p_i\}$. Because, as we will later see, it lends itself so well to numerical estimation, we are also interested in D_2, the correlation dimension, which is often denoted ν.

There has been much recent interest in determining D_q for all q, though a reparameterization as a "spectrum of scaling indices," $f(\alpha)$, is usually introduced [11]. Here,

$$\alpha(q) = D_q + (q-1)\frac{d}{dq}D_q \tag{2.11}$$

and

$$f(q) = D_q + q(q-1)\frac{d}{dq}D_q. \tag{2.12}$$

$f(\alpha)$ is usually interpreted as the fractal dimension (D_0) of points on the attractor whose pointwise dimension (§2.4.1) is α [11]. Estimation of an $f(\alpha)$ curve may be done numerically, and has been even for experimental data [12], but this is difficult except at low dimension. Seminumerical results are promising [13] but these involve knowledge of the map that generates the strange

attractor.

2.4 Pairwise distance algorithms

A more successful class of numerical algorithms for estimating dimension is based on the principle of interpoint distances. By computing distances between pairs of points on the attractor, we avoid the deliberate "ϵ-fuzzying" of the box-counting algorithms. We obtain the behavior for the entire range of ϵ at once, instead of having to recalculate $n(N,\epsilon)$ for each new ϵ.

2.4.1 Pointwise dimension

The most straightforward application is to the pointwise dimension. We define the dimension at a point X by

$$d_X = \lim_{r \to 0} \frac{log\ \mu(B_X(r))}{log\ r}, \tag{2.13}$$

where $B_X(r)$ is a "ball" of radius r centered at the point X, and $\mu(B_X(r))$ is its measure. As usual,

$$\mu(B_X(r)) = \int_{B_X(r)} \rho(X)\ dX, \tag{2.14}$$

where $\rho(X)$ is the invariant distribution, or "density." Note that this is a very intuitive definition. For it says the "mass" of a ball of radius r scales like r^{d_X}. For isolated points, straight lines, and plane surfaces, this corresponds to dimensions of 0, 1, and 2, respectively. See Figure 2.2. For a solid, the dimension is clearly and properly 3.

Computing the dimension at a specific point is numerically very straightforward. Simply compute the distance between that point and every other sample point. Then estimate

$$\mu(B_X(r)) \approx C_X(N,r) \equiv \frac{\text{\# of distances less than r}}{\text{total \# of distances}}. \tag{2.15}$$

That is, with a sample of N points $\{X_0, X_1, \ldots, X_{N-1}\}$ with $X_i \in \mathcal{A}$,

$$C_i(N,r) \equiv C_{X_i}(N,r) \equiv \frac{1}{N-1} \sum_{\substack{j=0 \\ j \neq i}}^{N-1} H\left(r - \left\| X_i - X_j \right\|\right). \tag{2.16}$$

From this, the pointwise dimension at X_i is

$$d_i \equiv d_{X_i} \equiv \lim_{r \to 0} \lim_{N \to \infty} \frac{log\ C_i(N,r)}{log\ r}. \tag{2.17}$$

Again, the order of limits is important. N is only as large as the number of points available and r can be only as small as the distance to the nearest point. As in the case of the box-counting dimension, Equation (2.17) is not directly evaluated; instead, a slope is sought in the small r region of a log-log plot of $C_i(N,r)$ versus r.

The pointwise dimension is the same for almost every point on the attractor of a generic dynamical map, and it is the information dimension D_1 [14]. The "average pointwise dimension" is found by averaging the pointwise dimension at each point (or at a small set of N_{ref} reference points) of the sample.

$$D_1 = \langle d_i \rangle \equiv \frac{1}{N} \sum_i d_i \tag{2.18}$$

(or

$$D_1 = \frac{1}{N_{ref}} \sum_j d_j, \tag{2.19}$$

where j indexes the reference points). See §10.3 for further discussion of the average pointwise dimension.

2.4.2 Correlation integral

Finally, we introduce the correlation integral [15,16]. This is obtained by a more direct kind of averaging, and it is statistically very powerful, since the averaging is done *before* the taking of logarithms or limits or slopes. We define

$$C(N,r) \equiv < C_i(N,r) > = \frac{2}{N(N-1)} \sum_{i=0}^{N-1} \sum_{j=0}^{i-1} H\left(r - \|X_i - X_j\|\right). \tag{2.20}$$

Thus, we consider all distances between every pair of points. A slight variation, suggested in Chapter Nine, eliminates from consideration those pairs of points for which $i-j \leq W$ for some $W \geq 1$. That is,

$$C(N,W,r) \equiv \frac{2}{(N-W+1)(N-W)} \sum_{i=0}^{N-1} \sum_{n=W}^{i-1} H\left(r - \|X_{i+n} - X_i\|\right). \tag{2.21}$$

Note that $W=1$ gives the standard correlation integral (2.).

The correlation dimension is given by [17]

$$\nu \equiv D_2 = \lim_{r \to 0} \lim_{N \to \infty} \frac{log\ C(N,r)}{log\ r}, \tag{2.22}$$

though, as in the case of the box-counting definition, what we actually do is plot

log C(N,r) versus *log* r for some large but fixed N, and seek a slope in the small r regime.

Though in principle it is the small r limit that we seek, our interest in the dimension stems from the *scaling* properties of the strange attractor. While we are seeking a slope in the small r regime, what we are really looking for is a wide regime in r over which $C(N,r) \propto r^{\nu}$. Further, the more accurate estimates of ν will be over the wider range.

The power of the correlation integral is that it probes the structure of the attractor down to very small distances, down to the distance between the nearest pair of points. The range over which C(N,r) varies is from 1 to $2/N^2$. This $O(N^2)$ dynamic range is unique to the correlation dimension. In both the average pointwise dimension above, and the Termonia-Alexandrowicz dimension that we will discuss later, the range of distances over which slopes are sought is typically O(N).

2.4.2.1 Takens' maximum likelihood estimate

Rather than trying to find the best-fit slope on a log-log plot of the correlation integral C(N,r) versus r, Takens [18] provides a direct estimate of ν from a list of all the D distances $\{r_0, r_1, r_2, \ldots, r_{D-1}\}$ that are less than some fixed r_o. Using the method of maximum likelihood, and assuming that the distances are independently chosen from a probability density $P(r) \propto r^{\nu-1}$, Takens obtains an estimator for ν

$$\hat{\nu} = \frac{-1}{\frac{1}{D}\sum_{i=0}^{D-1} ln(r_i/r_o)}, \qquad (2.23)$$

which has an expected error

$$\sqrt{<(\nu-\hat{\nu})^2>} = \frac{\hat{\nu}}{\sqrt{D}}. \tag{2.24}$$

Takens further shows that this estimator is optimal in that its expected error is the smallest of all estimators.

We note that the number of distances less than r_o is $D=N^2C(N,r_o)/2$, which for fixed r_o, scales as N^2; so Takens' estimate has an error bar of

$$\sigma = \frac{\hat{\nu}}{N\sqrt{C(N,r_o)/2}}, \tag{2.25}$$

which scales as $1/N$ for fixed r_o.

There are a few notable disadvantages in using Takens' method. What seems at first an advantage is that it provides a direct output, a single numerical estimate (not a bulky log-log plot) for the dimension. This is fine if all is going well, and a single number is all that is desired. But that bulky log-log plot has powerful diagnostic value. It tells the experimentalist whether or not the scaling of C(N,r) with r truly is proportional to r^{ν}. If r_o has been inappropriately chosen, or if there are any second-order effects, these can be seen on the C(N,r) plot.

An apparent disadvantage of Takens' method is computational. It requires evaluating the logarithm of every distance less than r_o. The standard method

merely bins those distances into a histogram. As it turns out, however, we can apply Takens' analysis directly to the correlation integral

$$\hat{\nu} = \frac{-C(N,r_o)}{\int_0^{r_o} \frac{dC(N,r)}{dr}\, ln(r/r_o)} = \frac{C(N,r_o)}{\int_0^{r_o} \frac{C(N,r)}{r}\, dr}. \tag{2.26}$$

A very good approximation to this formula, useful in the case of discretized data, with $r=(k+\frac{1}{2})\epsilon$ and $r_o=K\epsilon$ (see §5.2), is given by

$$\hat{\hat{\nu}} = \frac{\frac{1}{2}\cdot\left[C(N,r_o+\frac{1}{2}\epsilon)+C(N,r_o-\frac{1}{2}\epsilon)\right]}{\left[\sum_{r=\frac{1}{2}\epsilon}^{r_o-\frac{1}{2}\epsilon} \frac{C(N,r)}{r}\right]\epsilon}. \tag{2.27}$$

There are no logarithms in this expression, and more to the point: this is an expression that is evaluated *after* the usual binning algorithm has provided us with a discretized C(N,r). Getting C(N,r) in the first place is the computationally expensive part; the analysis in Equation (2.27) is relatively cheap. Here, the discretization error $\left|\hat{\hat{\nu}}-\hat{\nu}\right|$ will be very small; in fact, it is $O(\epsilon^2)$.

Finally, we note that Takens' estimate can be applied to Equation (2.17) for computation of the pointwise dimension as well.

2.4.3 Nearest neighbors

Having introduced the correlation algorithm, we now review a number of alternative distance algorithms for computing dimension. We begin with an algorithm proposed by Termonia and Alexandrowicz [20]. For a specific point $X \varepsilon \mathcal{A}$, let $r_X(n)$ denote the distance to the nth nearest neighbor of X. Whereas

before we spoke of n scaling as $r_X(n)^d$, with d as the dimension of the set, here we speak of $r_X(n)$ scaling as $n^{1/d}$. Let $\langle r_X(n) \rangle$ be the average taken over X. Then a log-log plot of $\langle r_X(n) \rangle$ versus n can be made, and the slope provides a value for 1/d.

The dimension d that the Termonia-Alexandrowicz algorithm measures appears to be the "fractal" dimension D_0, though Grassberger has provided some caveats [21].

Although it lacks the $O(N^2)$ range that the correlation integral provides, our own numerical experiments suggest that the Termonia-Alexandrowicz algorithm is less sensitive to singularities and to "lacunarity" (see Chapters Six and Seven) than is the correlation integral, and for this reason we consider it a promising candidate for further development .

2.4.4 Recurrence time algorithm

Badii and Politi [22] have generalized the notion of Termonia and Alexandrowicz and consider the statistical properties of the distribution $P(\delta,n)$ of nearest neighbor distances δ among n randomly chosen points. These are related to the generalized dimensions discussed above and their numerical computation is discussed in [22].

2.4.5 Periodic orbits algorithm

Recently, it has been suggested [23] that the fractal invariant measure of a chaotic strange attractor can be systematically approximated by sets of unstable periodic orbits that can be extracted from the time series. The

algorithm permits surprisingly precise estimates of the Hausdorff dimension, and there is speculation that the generalized dimensions may be similarly computed.

2.5 Notes and References

[1] Benoit B. Mandelbrot. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).

[2] Felix Hausdorff. "Dimension und äußeres Maß," Mathematische Annalen 79 (1919) 157.

[3] A third type, which we do not discuss, is based on local linear fitting and is described by H. Froehling, J. P. Crutchfield, D. Farmer, and R. Shaw. "On determining the dimension of chaotic flows," Physica 3D (1981) 605.

[4] An exception is provided by the set of rationals in the interval [0,1]. The Hausdorff dimension of any countable set of points is zero, so $D_H=0$. But the box-counting dimension is easily shown to be $d_O=1$. We speculate that in a dynamical context, the set of rationals is not "physically interesting."

[5] J. Doyne Farmer, Edward Ott, and James A. Yorke. "The dimension of chaotic attractors," Physica 7D (1983) 153.

[6] Examples for which the limit does not exist are discussed in Chapter Seven. Even in these cases, though, an average slope of a log-log plot gives the correct dimension.

[7] Mark J. McGuinness. "A computation of the limit capacity of the Lorenz attractor," Physica 16D (1985) 265.

[8] H. S. Greenside, A. Wolf, J. Swift, and T. Pignataro. "Impracticality of a box-counting algorithm for calculating the dimensionality of strange attractors,"

Phys. Rev. **A** 25 (1982) 3453.

[9] H. G. E. Hentschel and Itamar Procaccia. "The infinite number of generalized dimensions of fractals and strange attractors," Physica 8D (1983) 435.

[10] A. Renyi. *Probability Theory* (North-Holland, Amsterdam, 1970).

[11] Thomas C. Halsey, Mogens H. Jensen, Leo P. Kadanoff, Itamar Procaccia, and Boris I. Shraiman. "Fractal measures and their singularities: the characterization of strange sets," Phys. Rev. **A** 33 (1986) 1141.

[12] Mogens H. Jensen, Leo P. Kadanoff, Albert Libchaber, Itamar Procaccia, and Joel Stavans. "Global universality at the onset of chaos: results of a forced Rayleigh-Bénard experiment," Phys. Rev. Lett. 55 (1985) 2798; E. G. Gwinn and R. M. Westervelt. "Scaling structure of attractors at the transition from quasiperiodicity to chaos in electronic transport in Ge," Phys. Rev. Lett. 59 (1987) 157.

[13] R. Badii and A. Politi. "Renyi dimensions from local expansion rates," Phys. Rev. **A** 35 (1987) 1288.

[14] This has been proven for a very restrictive class of dynamical map by L.-S. Young. "Capacity of attractors," Ergod. Theory Dynam. Syst. 1 (1981) 381. We are unaware of a more general proof.

[15] Peter Grassberger and Itamar Procaccia. "Characterization of strange attractors," Phys. Rev. Lett. 50 (1983) 346.

[16] F. Takens. "Invariants related to dimension and entropy," in *Atas do* 13^{O} (colóqkio brasiliero de matemática, Rio de Janeiro, 1983).

[17] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors," Physica 9D (1983) 189.

[18] F. Takens. "On the numerical determination of the dimension of an

attractor," in *Dynamical Systems and Bifurcations, Groningen*, 1984, Vol. 1125 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1985).

[19] Yves Termonia and Zeev Alexandrowicz. "Fractal dimension of strange attractors from radius versus size of arbitrary clusters," Phys. Rev. Lett. **51** (1983) 1265.

[20] P. Grassberger. "Generalizations of the Hausdorff dimension of fractal measures," Phys. Lett. **107A** (1985) 101.

[21] Remo Badii and Antonio Politi. "Statistical description of chaotic attractors: the dimension function," J. Stat. Phys. **40** (1985) 725.

[22] Ditza Auerbach, Predrag Cvitanović, Jean-Pierre Eckmann, Gemumu Gunaratne, and Itamar Procaccia. "Exploring chaotic motion through periodic orbits," Phys. Rev. Lett. **58** (1987) 2387.

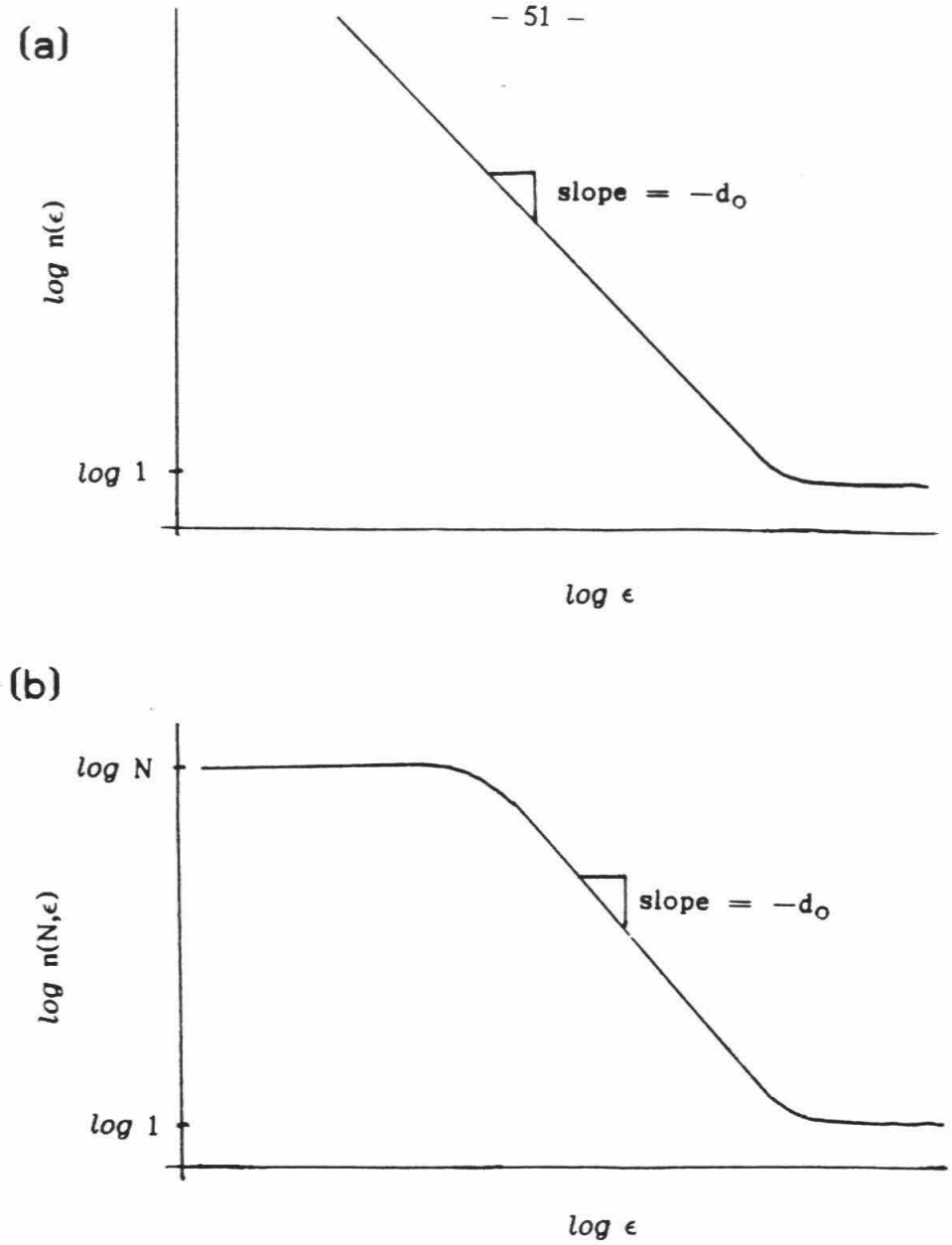


Figure 2.1 (a) On a plot of *log* n(ϵ) versus *log* ϵ, the negative slope is the box-counting dimension d_O. As the box size ϵ approaches the size of the attractor, the number n(ϵ) of boxes needed to cover the attractor approaches one. (b) The effect of finite N is to restrict the scaling range over which a slope of d_O is observed, since $n(N,\epsilon) \to N$ as $\epsilon \to 0$.

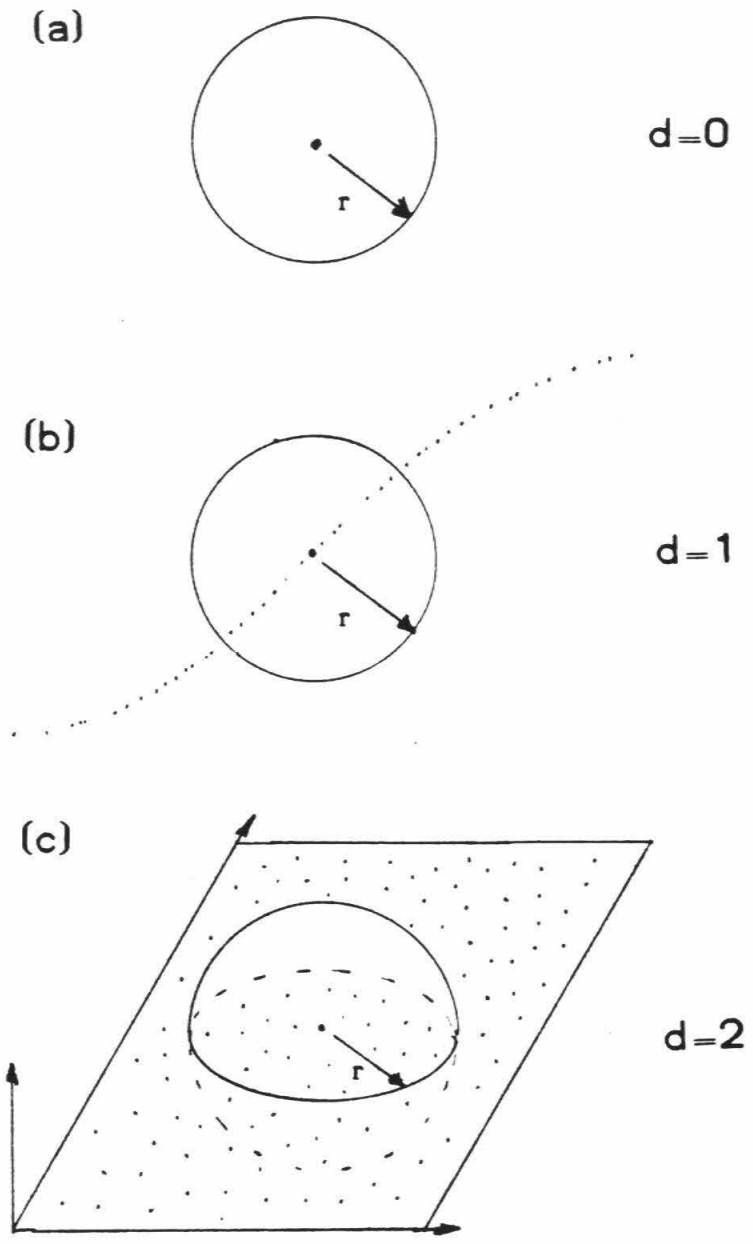


Figure 2.2 Dimension as a scaling of "mass" with radius. Here, mass is taken as the number of points inside the circle or sphere of radius r. (a) d=0 is a point; mass is independent of radius; (b) d=1 is a line; mass varies linearly with radius; (c) d=2 is a plane; mass varies quadratically with radius.

CHAPTER THREE

3. GEOMETRICAL PROPERTIES OF THE CORRELATION INTEGRAL

The correlation integral is the prime object of study in the remaining chapters. Before we begin the detailed analysis of the limitations and potential pitfalls in the use of the correlation integral, we relate in this chapter three crucial properties of the correlation integral and the correlation dimension.

We classify the properties in this chapter as "geometrical," since they are concerned only with the attractor and its invariant measure, not with the dynamical process through which points on the attractor are sampled. In Chapter Four, we will discuss the statistical issues involved with sampling an invariant measure with a discrete set of points.

3.1 Correlation dimension is norm-independent

Finding the correlation dimension from a set of discrete points in $\mathbb{R}^m$ requires computing distances between pairs of points. There are a variety of ways to define what we mean by distance. Mathematically, the distance function $d:\mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ need satisfy only these minimal requirements:

$$d(x,y)=0 \Leftrightarrow x=y; \tag{3.1}$$

$$d(x,y) = d(y,x); \tag{3.2}$$

$$d(x,y)+d(y,z) \geq d(x,z). \tag{3.3}$$

However, we will restrict our attention to a small class of distance functions

called L_p norms. For $x \varepsilon \mathbb{R}^m$, write in coordinate notation $x = (x_0, x_1, \ldots, x_{m-1})$. Then the distance function, $d_p(x,y)$, for the L_p norm is given by

$$d_p(x,y) \equiv \left[\sum_{i=0}^{m-1} \left| x_i - y_i \right|^p \right]^{1/p}. \tag{3.4}$$

Three specific cases will be considered, the "taxicab" norm, the usual Euclidean norm, and the maximum norm, corresponding to $p=1$, 2, and ∞, respectively.

In the L_1 or "taxicab" norm, the distance between two points in $\mathbb{R}^m$ is given by the sum of the distances between each of the coordinates. For a taxicab to go three blocks north and four blocks west requires seven blocks of travel, as $7=3+4$.

L_2, the Euclidean norm, is the distance "as the crow flies." Three blocks north and four blocks west is a distance of only five blocks, as $5=\sqrt{3^2+4^2}$ [1].

For the L_∞ norm, we have

$$d_\infty(x,y) = \lim_{p \to \infty} \left[\sum_{i=0}^{m-1} \left| x_i - y_i \right|^p \right]^{1/p} = \max_{0 \le i < m} \left| x_i - y_i \right|. \tag{3.5}$$

Three blocks north and four blocks west amounts in this norm to a distance of four blocks, as $4=\max(3,4)$.

There are a number of advantages to using the L_∞ norm. For one, it is efficient to compute, certainly more efficient than the Euclidean norm, which requires squares and square roots. Also, the size of the attractor does not grow as the embedding dimension is increased. Another advantage of L_∞ has to do with computation of a diagnostic called the K_2 entropy [2]. The primary

advantage of the L_2 norm is that it does not require any preferred directions; distances are invariant to changes in the orientation of the axes.

Despite the various practical advantages and disadvantages of the different norms, it is an important fact that the correlation dimension itself is norm-independent. We can see this by looking at interiors of "circles" in the various norms. A circle of radius r is defined as the locus of points whose distance from the center is r. Figure 3.1 shows what these circles look like in the L_1, L_2, and L_∞ norms. We have a diamond, a circle, and a square, respectively. Recalling the definition of C(r) as the fraction of distances less than r, we can see immediately that

$$C_{L1}(r) \leq C_{L2}(r) \leq C_{L\infty}(r). \tag{3.6}$$

In two dimensions, it is easy to see that a diamond ("L1 circle") of radius r will completely enclose a square ("L∞ circle") of radius r/2; in m dimensions the square should be of radius r/m to be just enclosed by the diamond of radius r. That is,

$$C_{L\infty}(r/m) \leq C_{L1}(r) \leq C_{L\infty}(r). \tag{3.7}$$

From which,

$$\lim_{r\to 0} \frac{\log C_{L\infty}(r/m)}{\log r} \leq \lim_{r\to 0} \frac{\log C_{L1}(r)}{\log r} \leq \lim_{r\to 0} \frac{\log C_{L\infty}(r)}{\log r}. \tag{3.8}$$

But

$$\lim_{r\to 0} \frac{\log C_{L\infty}(r/m)}{\log r} = \lim_{r'\to 0} \frac{\log C_{L\infty}(r')}{\log mr'} = \lim_{r'\to 0} \frac{\log C_{L\infty}(r')}{\log m + \log r'}$$

$$= \lim_{r'\to 0} \frac{\log C_{L\infty}(r')}{\log r'} = \lim_{r\to 0} \frac{\log C_{L\infty}(r)}{\log r}, \tag{3.9}$$

so the inequality in Equation (3.7) is in fact a strict equality, and in particular, $\nu_{L1}=\nu_{L\infty}$. Using

$$C_{L2}(r/\sqrt{m}) \leq C_{L1}(r) \leq C_{L2}(r), \tag{3.10}$$

we similarly have $\nu_{L1}=\nu_{L2}$. Thus, we can write

$$\nu_{L1} = \nu_{L2} = \nu_{L\infty}. \tag{3.11}$$

Dimension is norm-independent.

3.2 Dimensions of direct products add

We will show that this is true for an aligned direct product in the L_∞ norm, and will use the result above to assert that it is true as well for the L_1 and L_2 norms.

Let $\mathcal{A}$ be an attractor in $\mathbb{R}^{m_a}$, and let $\mathcal{B}$ be an attractor in $\mathbb{R}^{m_b}$.

Consider the direct product $\mathcal{A}\times\mathcal{B} \subset \mathbb{R}^{m_a+m_b}$. A point x in the set $\mathcal{A}\times\mathcal{B}$ corresponds to a point $x_a \in \mathcal{A}$, and $x_b \in \mathcal{B}$, where we usually write $x=(x_a,x_b)$. The distance between the pair x and y is given in the L_∞ norm, by

$$\|x-y\| = \max (\|x_a-y_a\|, \|x_b-y_b\|). \tag{3.12}$$

Thus, $\|x-y\| \leq r$ implies that $\|x_a-y_a\| \leq r$ and $\|x_b-y_b\| \leq r$. For points chosen at random, x_a and x_b are entirely independent of each other; thus,

$$P(\|x-y\| \leq r) = P(\|x_a-y_a\| \leq r) \times P(\|x_b-y_b\| \leq r). \tag{3.13}$$

That is,

$$C_{ab}(r) = C_a(r)C_b(r) \tag{3.14}$$

and

$$\nu_{ab} \equiv \lim_{r\to 0} \frac{\log C_{ab}(r)}{\log r} = \lim_{r\to 0} \frac{\log C_a(r)C_b(r)}{\log r} = \lim_{r\to 0} \frac{\log C_a(r) + \log C_b(r)}{\log r}$$

$$= \nu_a + \nu_b. \tag{3.15}$$

We have shown that dimensions add in the case of aligned direct products in the L_∞ norm. We have seen above, however, that dimension is independent of norm, so we can say $\nu_{ab}=\nu_a+\nu_b$ in the L_1 and L_2 norms as well. Further, since L_2 is a rotationally invariant norm, we see that $\nu_{ab}=\nu_a+\nu_b$ does not depend on the alignment of the axes. (Indeed, that dimension should not depend on alignment of the axes satisfies a very primal intuition.)

3.3 A geometric definition of the correlation integral

In [3], the correlation integral C(r) is defined as the $N\to\infty$ limit of the

function

$$C(N,r) = \frac{2}{N(N-1)} \sum_{i=0}^{N-1} \sum_{j=0}^{i-1} H\left(r - \left\|X_i - X_j\right\|\right), \tag{3.16}$$

where $X_0, X_1, \ldots$ is an orbit that is dense over the attractor. In this section, we will provide an expression for C(r) directly in terms of the invariant measure μ of the attractor.

First, rewrite Equation (3.16) as

$$C(r) = \lim_{N\to\infty} \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} H\left(r - \left\|X_i - X_j\right\|\right), \tag{3.17}$$

which is clearly equivalent for $N\to\infty$. Next use the ergodic property, see Equation (1.17),

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} H\left(r - \left\|X_i - X_j\right\|\right) = \int_{\mathcal{A}} H\left(r - \left\|X_i - X\right\|\right) \rho(X)\, dX \tag{3.18}$$

to give

$$C(r) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \int_{\mathcal{A}} H\left(r - \left\|X_i - X\right\|\right) \rho(X)\, dX. \tag{3.19}$$

We can again invoke the ergodic property,

$$C(r) = \int_{\mathcal{A}}\int_{\mathcal{A}} H\left(r - \left\|Y - X\right\|\right) \rho(X)\, \rho(Y)\, dY\, dX. \tag{3.20}$$

An equivalent form "measures" the subset of $\mathcal{A}\times\mathcal{A}$ for which the condition

$\|X-Y\| \leq r$ holds.

$$C(r) = \int_{\substack{\mathcal{A}\times\mathcal{A} \\ \|X-Y\|\leq r}} \rho(X)\ \rho(Y)\ dX\ dY \tag{3.21}$$

We will find the previous two expressions useful in modeling the geometric effects that are discussed in Chapters Five, Six, and Seven.

3.4 Notes and References

[1] The geometrical significance of this arithmetic identity is credited to Pythagoras, d. 497 B.C.

[2] Peter Grassberger and Itamar Procaccia. "Estimation of the Kolmogorov entropy from a chaotic signal," Phys. Rev. **A** 28 (1983) 2591; also, Aviad Cohen and Itamar Procaccia. "Computing the Kolmogorov entropy from time series of dissipative and conservative dynamical systems," Phys. Rev. **A** 31 (1985) 1872.

[3] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors," Physica 9D (1983) 189.

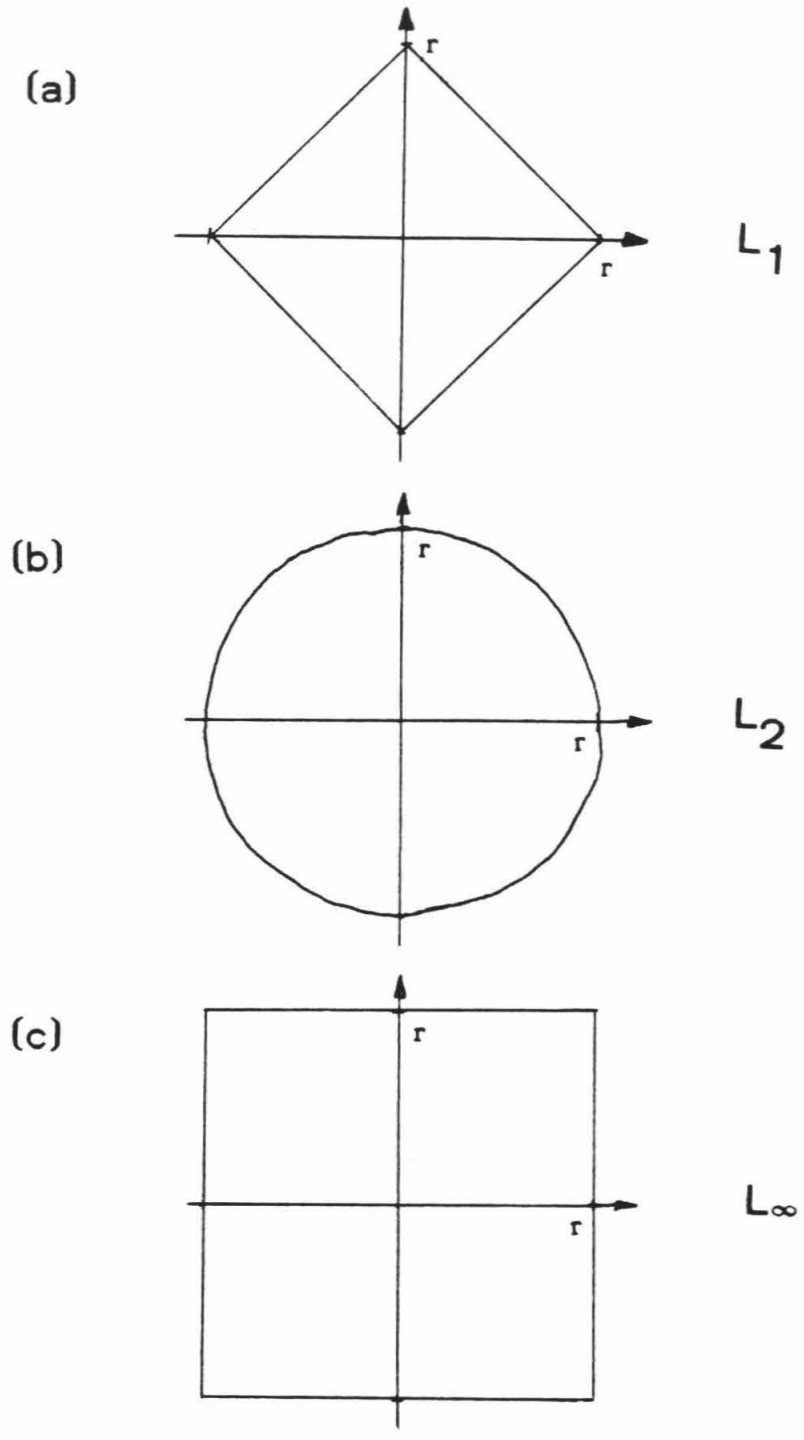


Figure 3.1 "Circles" of radius r in various L_p norms. (a) L_1, a diamond; (b) L_2, a circle; (c) L_∞, a box.

CHAPTER FOUR

4. STATISTICAL ANALYSIS OF CHATOIC MOTION

We find the study of nonlinear systems amenable to statistical analysis in two distinct ways [1]. First, from an ensemble (E_0) of dynamical orbits on the attractor, we can evaluate the performance of algorithms that compute dimension from a finite set of points. Although the dimension of an attractor is defined in terms of a limit as $N \to \infty$, our algorithms estimate the dimension from a finite set of N points on the attractor, and the value of a given estimate will depend on which N points are taken. We are interested in the accuracy of these finite N estimates (that is, how near a typical finite N estimate is to the $N \to \infty$ limit) and in their precision (that is, how near finite N estimates are to each other).

Second, we can approximate chaotic orbits by stochastic orbits. This second approach is equivalent to replacing the dynamical ensemble E_0 by larger ensembles (E_1 and E_2), which still capture the essential (geometric) features of the attractor. The advantage of these approximate ensembles is that they allow us to make *theoretical* predictions about an algorithm's behavior in certain model situations.

4.1 Honest ensemble: E_0

In practice, a computation is based on a sample of N points on the attractor, and the N points are provided by an experiment (be it physical or numerical). A repeat of the experiment (with different initial conditions) will provide a different sample of N points, and — it follows — a different

computational result.

The first ensemble, E_0, is the "honest" ensemble and is defined in terms of the true dynamics on the attractor. We begin with an *ensemble of initial conditions*, points X_o on the attractor $\mathcal{A}$ weighted by the natural invariant measure μ. To each X_o, we can assign an orbit of length N: X_o, $f(X_o)$, $f^2(X_o)$, ..., $f^{N-1}(X_o)$. This orbit is a typical member of the ensemble $E_0(N)$.

In practice, we are given not an ensemble but a single finite orbit (or at most a few orbits, from each repetition of the experiment). From this finite orbit, we attempt to estimate quantities (*e.g.*, dimension) that characterize the attractor. Suppose F is such a quantity, and $\mathfrak{F}$ is the function (or algorithm) that estimates F from the finite time series. Then

$$\hat{F}(N;X_o) \equiv \mathfrak{F}(\, N;\, X_o,\, f(X_o),\, f^2(X_o),\, \ldots,\, f^{N-1}(X_o)\,) \tag{4.1}$$

defines a particular estimate. It is often the case that we do not have X, the position in the original phase space, but $v=\pi_m(X)$, its projection into the reconstructed space $\mathbb{R}^m$ (see §1.4.1). Then,

$$\hat{F}(N;X_o) \equiv \mathfrak{F}(\, N;\, \pi_m(X_o),\, \pi_m(f(X_o)),\, \pi_m(f^2(X_o)),\, \ldots,\, \pi_m(f^{N-1}(X_o))\,) \tag{4.2}$$

$$\equiv \mathfrak{F}(\, N;\, v_0,\, v_1, \ldots,\, v_{N-1}). \tag{4.3}$$

We define the *accuracy* of our algorithm $\mathfrak{F}$ by $|F(N)-F|$, where F(N) is the (ensemble) average finite N estimate

$$F(N) \equiv \langle\, \hat{F}(N;X_o)\,\rangle \equiv \int_{\mathcal{A}} \rho(X)\, \hat{F}(N;X)\, dX, \tag{4.4}$$

and F is the quantity being estimated; of course, we expect

$$\lim_{N\to\infty} F(N) = F. \tag{4.5}$$

The precision of the algorithm is defined by the spread in individual finite N estimates.

$$\sigma(N) \equiv \sqrt{\langle\, (\hat{F}(N;X_o)-F(N))^2 \,\rangle} = \sqrt{\langle\, \hat{F}^2(N;X_o) \,\rangle - F^2(N)} \tag{4.6}$$

We want the precision to improve as the number of points is increased, and we express this with the condition

$$\lim_{N\to\infty} \sigma(N) = 0. \tag{4.7}$$

This is an important condition because it tells us that in the $N\to\infty$ limit, the algorithm almost always yields the correct value; that is,

$$\lim_{N\to\infty} \hat{F}(N;X_o) = F \tag{4.8}$$

for all but a measure zero set of X_o [2]. We note that Equation (4.8) is stronger than the usual ergodic property, which holds only for averages of a quantity $\mathcal{F}(X)$; that is, for functions $\mathfrak{F}$ of the form

$$\mathfrak{F}(N;X_0,X_1,\ldots,X_{N-1}) = \frac{\mathcal{F}(X_0)+\cdots+\mathcal{F}(X_{N-1})}{N}. \tag{4.9}$$

However, we view Equation (4.8) not as a property of the motion's ergodicity, but instead as a property of the estimator algorithm $\mathfrak{F}$.

Our particular interest is with algorithms for estimating attractor dimension from a finite time series. The estimator for the correlation integral is (see §2.4.2)

$$\hat{C}(N,r;\, X_o) \equiv \frac{2}{N(N-1)} \sum_{0 \le i < j < N} H\left(r - \left\| v_i - v_j \right\|\right). \tag{4.10}$$

And Takens' estimator [3] is given by

$$\hat{\nu}(N,r_o;\, X_o) \equiv \frac{\displaystyle\sum_{0 \le i < j < N} H\left(r_o - \left\| v_i - v_j \right\|\right)}{\displaystyle\sum_{0 \le i < j < N} H\left(r_o - \left\| v_i - v_j \right\|\right) \ln\left[\frac{r_o}{\left\| v_i - v_j \right\|}\right]}. \tag{4.11}$$

4.2 Approximate ensembles: E_1, E_2

As an approximation to the honest ensemble E_0, we introduce E_1, which is based on the natural invariant measure but does not use the dynamical information that describes the motion of an orbit on the attractor. A typical member of $E_1(N)$ is the orbit $X_0, X_1, \ldots X_{N-1}$, where now each X_i is chosen independently from the ensemble of initial conditions. The constraint $X_{i+1} = f(X_i)$ that was present in E_0 has been lifted. Indeed, we can think of E_1 as the honest ensemble associated with the stochastic dynamical system f^* where f^* is defined as the stochastic map which sends state X to any state in the original ensemble of initial conditions (according to the usual weighting by the invariant measure μ), independently of X. For highly chaotic systems

especially, the approximation of f by f^* can be very useful.

From N points on the attractor, the correlation integral considers the distances between each of the $N(N-1)/2$ pairs. For instance, if N=4, the distances associated with a typical member of $E_1(4)$ is the set $\{r_{01}, r_{02}, r_{03}, r_{12}, r_{13}, r_{23}\}$, where $r_{ij} \equiv \|X_i - X_j\|$, with $\{X_0, X_1, X_2, X_3\} \varepsilon E_1(4)$. We note that the individual distances associated with a member of E_1 are *not* independent. We have, for instance, that they are constrained by the triangle inequality $r_{ij} \leq r_{ik} + r_{kj}$.

Incorporating these constraints into the statistics may be difficult, so we further approximate E_1 with another ensemble (E_2) in which these constraints are neglected.

E_2 is something of a specialized ensemble; it is applicable only to estimator functions that are based on pairwise distances

$$\hat{F}(N; X_0, \ldots, X_{N-1}) = \mathfrak{F}(N; \|X_0 - X_1\|, \ldots, \|X_i - X_j\|, \ldots, \|X_{N-2} - X_{N-1}\|), \qquad (4.12)$$

such as correlation dimension algorithms; see Equations (4.10, 4.11).

Conceptually, $E_2(N)$ is constructed by taking $N(N-1)/2$ *pairs* of points, all points independent of each other, and dropping them onto the attractor. For each pair a distance is measured. For example, with N=4, six pairs or a total of twelve points are dropped (where only four points would be dropped in the E_1 ensemble). Distances in the E_2 ensemble are independent of each other, and a typical member of the ensemble $E_2(4)$ produces distances $\{r_0, r_1, r_2, r_3, r_4, r_5\}$, where $r_i \equiv \|X_{ia} - X_{ib}\|$ with X_{ia} and X_{ib} the two points in the ith pair.

A typical member of $E_0(N)$ involves a single choice of initial condition. Choosing the initial condition, we have chosen the entire orbit, and the entire set of N(N-1)/2 distances $r_{ij} = \|X_i - X_j\|$. A typical member of $E_1(N)$, on the other hand, involves N choices $X_0, \ldots, X_{N-1}$; and from these the $N(N-1)/2$ distances follow. There are $N(N-1)/2$ distances between all the pairs of points in E_0 and E_1; but those distances are constrained by various triangle inequalities among the distances. In $E_2(N)$, we create an ensemble of $N(N-1)/2$ *independent distances*. Each distance is chosen by taking a pair of points X_{ia} and X_{ib} independently from the ensemble of initial conditions and computing $r_i = \|X_{ia} - X_{ib}\|$; thus, there are $N(N-1)$ choices in each member of $E_2(N)$. We have $E_0(N) \subset E_1(N) \subset E_2(N)$.

Although E_0 is the ensemble that correctly describes the statistics on the attractor, calculations based on E_0 are usually intractable. Approximating E_0 by E_1 or E_2 often simplifies the calculations and permits theoretical treatment of algorithm performance.

Replacing deterministic orbits by stochastic orbits is on its face invalid; points follow one another over the attractor of a dynamical system in an absolutely deterministic way. Yet for systems that are very "chaotic" (bounded systems with very sensitive dependence on initial conditions), this process of dropping points independently (E_1) or pairs independently (E_2) nonetheless leads to reasonable predictions of (*e.g.*,) the correlation integral behavior.

For systems that are not chaotic, or are chaotic but still have an appreciable time autocorrelation, the approximation of E_0 by E_1 can lead to difficulties in estimating attractor dimension. We will discuss these effects in

Chapters Eight and Nine.

The further approximation of E_1 by E_2, on the other hand, is hardly ever a problem, as the rest of this chapter will attempt to demonstrate.

4.3 Probability distribution

We define the probability distribution of distances for an ensemble with $P(r_0,r_1,r_2,r_3,r_4,r_5)$, where $P(r_0,r_1,r_2,r_3,r_4,r_5)\, dr_0dr_1dr_2dr_3dr_4dr_5$ denotes the probability that the ith distance is between r_i and r_i+dr_i, for i=0,1,...,5. Were we to write out the probability distributions for the ensembles E_0 and E_1 and E_2, we would see different expressions. In E_0 and E_1, the expressions are liable to be unwieldy in the extreme for all but the simplest cases, whereas in E_2 the property of independence allows us to write

$$P(r_0,r_1,r_2,r_3,r_4,r_5) = P(r_0)P(r_1)P(r_2)P(r_3)P(r_4)P(r_5). \tag{4.13}$$

4.4 Distinguishing E_1 and E_2

For the rest of this chapter, we will discuss the approximation of E_1 by E_2 for a specific model. We will say nothing about the dynamics of the system, and in fact we will specify the attractor only by its invariant measure.

Our real concern is with "measurable" statistical properties, specifically with the expectation values of relevant random variables. We will find, for instance, that the random variable

$$C(N,r) = \frac{2}{N(N-1)} \sum_i H(r-r_i), \tag{4.14}$$

corresponding to the correlation integral, has the same expected value in ensembles E_1 and E_2 (also the same variance, though the third moments do not agree). This is a very useful fact; it justifies using the second ensemble (which is easier to use) in place of the first ensemble (which is the geometrically correct one) for analysis of the correlation integral.

On the other hand, the random variable

$$r_{min} = \text{minimum of } (r_0, r_1, r_2, \cdots, r_{N(N-1)/2-1}) \tag{4.15}$$

has a different expectation value in E_1 and E_2 for finite N; however, they both have the same $N \to \infty$ limit. It is conjectured that all "interesting" [4] random variables have this asymptotic equality in the two ensembles.

4.4.1 The "wraparound" metric

We will now prove the assertions above for a specific model. Our attractor is the interval [0,1) with the "wraparound" metric: this is the same as a circle of unit circumference with distance measured along the arc between the two points on the circumference. In this metric, the distance between x and y is given by

$$d(x,y) = \min(|x-y|, 1-|x-y|). \tag{4.16}$$

For pairs of points dropped randomly on the interval, distances are uniformly distributed between 0 and $\frac{1}{2}$. This metric is also discussed in §6.1 and §8.1.

4.4.2 Ensemble E_1

The E_1 ensemble incorporates the physical restrictions imposed by the geometry of distances (such as the various triangle inequalities). On the other hand, the calculations in the E_1 ensemble are never as straightforward as in the E_2 ensemble, and in many cases some cleverness (that is, nonrigorous heuristic reasoning) is necessary to work through to a solution.

First, we write down an expression for $P(r_{01},\cdots,r_{(N-2)(N-1)})$ for the $N=3$ case of the wraparound model above. We find

$$\begin{aligned} P(r_{01},r_{02},r_{12}) = \; & 2\delta(\; r_{01}+r_{12}-r_{02}\;) + \\ & 2\delta(\; r_{01}-r_{12}+r_{02}\;) + \\ & 2\delta(-r_{01}+r_{12}+r_{02}\;) + \\ & 2\delta(\; r_{01}+r_{12}+r_{02}-1). \end{aligned} \tag{4.17}$$

We do not have independent distances (in the sense of Equation (4.13) for the E_2 ensemble), but we do have a symmetry in the arguments (that we could not, in general, expect for the E_0 ensemble). Here [5], we have

$$\begin{aligned} P(r_{01},r_{02},r_{12}) &= P(r_{01},r_{12},r_{02}) = P(r_{02},r_{01},r_{12}) = P(r_{02},r_{12},r_{01}) \\ &= P(r_{12},r_{01},r_{02}) = P(r_{12},r_{02},r_{01}). \end{aligned} \tag{4.18}$$

4.4.2.1 E_1: Correlation integral

To do this calculation, we require the probability density given in Equation (4.17); this necessarily limits us to the $N=3$ case. The symmetry in

the arguments of Equation (4.18) allows us to write

$$\langle \sum_i H(r-r_i) \rangle = \frac{N(N-1)}{2} \langle H(r-r_{01}) \rangle = 3 \langle H(r-r_{01}) \rangle. \tag{4.19}$$

Hence,

$$\langle C(N,r) \rangle \equiv \langle \frac{2}{N(N-1)} \sum_i H(r-r_i) \rangle = \langle H(r-r_{01}) \rangle \tag{4.20}$$

$$= \int_0^{1/2} dr_{01}\, H(r-r_{01}) \int_0^{1/2} dr_{02} \int_0^{1/2} dr_{12}\, P(r_{01},r_{02},r_{12}). \tag{4.21}$$

Let us do the inner integral first, letting r_{12} range over $[0,\frac{1}{2}]$. We have, for instance, that

$$\int_0^{1/2} dr_{12}\, \delta(r_{01} + r_{12} - r_{02}) = \begin{cases} 1 & \text{if } 0 \le (r_{02}-r_{01}) \le \frac{1}{2} \\ 0 & \text{otherwise,} \end{cases} \tag{4.22}$$

whereas

$$\int_0^{1/2} dr_{12}\, \delta(-r_{01} + r_{12} + r_{02}) = \begin{cases} 1 & \text{if } 0 \le (r_{01}-r_{02}) \le \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{4.23}$$

The two integrals are exactly complementary, and their sum is identically 1. The other two delta functions can also be shown to be complementary. The inner integral of Equation (4.21) is therefore given by

$$\int_0^{1/2} dr_{12}\, P(r_{01},r_{02},r_{12}) = 2(1+1) = 4. \tag{4.24}$$

Thus,

$$\int_0^{1/2} dr_{02} \int_0^{1/2} dr_{12}\, P(r_{01},r_{02},r_{12}) = \int_0^{1/2} 4\, dr_{02} = 2, \tag{4.25}$$

and finally

$$\langle C(N=3,r)\rangle = \int_0^{1/2} dr_{01}\, 2\, H(r-r_{01}) = 2r. \tag{4.26}$$

We remark that the expected value of the correlation integral varies linearly with r. This is just what we expect for a one-dimensional system.

The calculation of $\langle C(N=3,r)^2\rangle$ follows similar lines.

$$\langle C(N=3,r)^2\rangle = \frac{4}{N^2(N-1)^2} \left\langle \left[\sum_i H(r-r_i)\right]^2 \right\rangle \tag{4.27}$$

$$= \frac{4}{N^2(N-1)^2} \left\langle \sum_{i,j} H(r-r_i)\, H(r-r_j) \right\rangle. \tag{4.28}$$

We note that $H(r-r_i)H(r-r_i) = H(r-r_i)$. This, and the symmetry of the indices allow us to reduce the equation to

$$\langle C(N=3,r)^2\rangle = \frac{4}{N^2(N-1)^2} \langle D\cdot H(r-r_{01}) + D(D-1)\cdot H(r-r_{02})\cdot H(r-r_{12})\rangle, \tag{4.29}$$

where $D \equiv N(N-1)/2 = 3$ is the total number of distances. Thus,

$$\langle C(N=3,r)^2\rangle = \frac{2}{N(N-1)} \left\langle H(r-r_{01}) + \left(\frac{N(N-1)}{2}-1\right)\cdot H(r-r_{01})\cdot H(r-r_{02}) \right\rangle, \tag{4.30}$$

where now,

$$\langle H(r-r_{01})\cdot H(r-r_{02})\rangle = \int_0^{1/2} dr_{01} H(r-r_{01}) \int_0^{1/2} dr_{02} H(r-r_{02}) \int_0^{1/2} dr_{12} P(r_{01},r_{02},r_{12}). \tag{4.31}$$

But we have seen that the inner integral is equal to the constant 4, leaving two integrals over r_{01} and r_{02}, which can be evaluated *independently*. The result is

$$\langle H(r-r_{01})\cdot H(r-r_{02})\rangle = 4r^2. \tag{4.32}$$

We have for $\langle C(N=3,r)^2\rangle$, finally,

$$\langle C(N=3,r)^2\rangle = \frac{2}{N(N-1)}\left[2r + (\frac{N(N-1)}{2}-1)\cdot 4r^2\right] \tag{4.33}$$

$$= \frac{4r}{N(N-1)} + (1-\frac{2}{N(N-1)})\cdot 4r^2 \tag{4.34}$$

$$= 4r^2 + \frac{4}{N(N-1)}(r-2r^2). \tag{4.35}$$

We note that this leads to an "error bar" for C(N,r) of

$$\sigma(N=3,r) \equiv \sqrt{\langle C(N,r)^2\rangle - \langle C(N,r)\rangle^2} = 2\sqrt{\frac{r(1-2r)}{N(N-1)}} = \sqrt{\frac{2r(1-2r)}{3}}. \tag{4.36}$$

We caution that this derivation is valid only for N=3. Even at N=4, the algebra (all those delta functions!) becomes difficult to manage. Nonetheless, there is still a high order of symmetry in the distances; as long as all but the

outer two integrals (in the N>3 equivalent of Equations (4.21, 4.31)) can be shown to be constants independent of r, then the results for $\langle C(N,r)\rangle$ and $\langle C(N,r)^2\rangle$ should hold for all N. We will do the general N calculation in the E_2 ensemble.

4.4.2.2 E_1: Shortest distance

We will attempt to calculate $\langle r_{min}\rangle$ for arbitrary N in the E_1 ensemble, apologizing in advance for any obscurity in the derivation. The fact is, E_1 is a difficult ensemble under which to perform these calculations.

Begin by considering the probability $P(r_{min}>r)$ [6]. Drop the first point anywhere on the interval. Since we are using the wraparound metric, we can assume without loss of generality that the first point specifies the origin 0. The remaining N−1 points must be dropped so that the N intervals between neighboring points are all greater than r. This leaves a target area of 1−Nr that each of the N−1 points must land on [7]. Thus,

$$P(r_{min}>r) = (1-Nr)^{N-1}, \tag{4.37}$$

from which $F_{min}(r) \equiv P(r_{min}\leq r) = 1 - (1-Nr)^{N-1}$, and the probability density is

$$P_{min}(r) = \frac{dF_{min}}{dr} = N(N-1)\,[1-Nr]^{N-2}, \tag{4.38}$$

and

$$\langle r_{min}\rangle \equiv \int r P_{min}(r)dr = N(N-1)\int_0^{1/N} r\cdot[1-Nr]^{N-2}\,dr = \frac{1}{N^2}. \tag{4.39}$$

That $\langle r_{min}\rangle = O(1/N^2)$ is an important property. It is this that allows the

correlation algorithm to "see" to such small scales. Since $\langle r_{max} \rangle = O(1)$, we have a range in distances of order N^2. We emphasize that this $O(N^2)$ result depends crucially on our assumption that we are in the E_1 ensemble. We still observe $\langle r_{min} \rangle = O(1/N^2)$ with numerical simulations of chaotic attractors (in the honest ensemble E_0), though usually with a different (larger) coefficient. For nonchaotic motion, however, as we will see in Chapter Eight, we have $\langle r_{min} \rangle = O(1/N)$ and the algorithm does not probe so deeply to such small scales.

4.4.3 Ensemble E_2

Calculating probabilities is much easier with the E_2 ensemble; in fact, we will be able to do the calculations in this section for arbitrary N. We will estimate $\langle C(N,r) \rangle$, $\langle [C(N,r)]^2 \rangle$, and $\langle r_{min} \rangle$. Since the $D=N(N-1)/2$ distances are independent, we can write

$$P(r_0 r_1,\ldots,r_{D-1}) = P(r_0)P(r_1)\cdots P(r_{D-1}). \tag{4.40}$$

4.4.3.1 E_2: Correlation integral

We compute $\langle C(N,r) \rangle$ and $\langle C(N,r)^2 \rangle$ in the E_2 ensemble. We write

$$\langle C(N,r) \rangle = \int \left[\frac{2}{N(N-1)} \sum_i H(r - r_i) \right] P(r_0,\ldots,r_{D-1})\, dr_0 \cdots dr_{D-1} \tag{4.41}$$

$$= \frac{2}{N(N-1)} \sum_i \int H(r - r_i)\, P(r_0,\ldots,r_{D-1})\, dr_0 \cdots dr_{D-1}. \tag{4.42}$$

Symmetry of the arguments $r_0,\ldots r_{D-1}$ implies that each term of the summation is equal; thus,

$$\langle C(N,r)\rangle = \frac{2D}{N(N-1)} \int H(r-r_0)\, P(r_0,\ldots,r_{D-1})\, dr_0 \ldots dr_{D-1} \tag{4.43}$$

$$= \int H(r-r_0)\, P(r_0,\ldots,r_{D-1})\, dr_0 \ldots dr_{D-1}. \tag{4.44}$$

Here, we use independence of the probability distribution to write

$$\langle C(N,r)\rangle = \int_0^{1/2} dr_0\, H(r-r_0)\, P(r_0) \int_0^{1/2} dr_1\, P(r_1) \cdots \int_0^{1/2} dr_{D-1}\, P(r_{D-1}) \tag{4.45}$$

$$= \int_0^{1/2} dr_0\, H(r-r_0)\, P(r_0), \tag{4.46}$$

using $\int dr_i\, P(r_i) \equiv 1$ for $i \geq 1$. Finally, we use $P(r_0)=2$, and get

$$\langle C(N,r)\rangle = 2r \tag{4.47}$$

for $r \leq \frac{1}{2}$, and $C(N,r)=1$ for $r \geq \frac{1}{2}$. We remark that the ensemble average of $C(N,r)$ is independent of N. Thus,

$$\langle C(N,r)\rangle = \lim_{N\to\infty} C(N,r) \equiv C(r),$$

and we say that $\hat{C}(N,r;X_0)$ is an "accurate" estimate (see Equations (4.4, 4.5)) in the E_2 ensemble.

To calculate the *precision* of the estimate, we first calculate $\langle C(N,r)^2\rangle$. Write

$$\langle C(N,r)^2 \rangle = \frac{2}{N(N-1)} \langle H(r-r_{01}) + (\frac{N(N-1)}{2}-1)\cdot H(r-r_{01})\cdot H(r-r_{02}) \rangle. \quad (4.48)$$

We have from Equations (4.46, 4.47) above that

$$\langle H(r-r_0) \rangle = 2r, \quad (4.49)$$

and independence of the distances allows us to write

$$\langle H(r-r_0)\cdot H(r-r_1) \rangle = \langle H(r-r_0) \rangle \cdot \langle H(r-r_1) \rangle = 4r^2. \quad (4.50)$$

Substituting these into Equation (4.48),

$$\langle C(N,r)^2 \rangle = \frac{2}{N(N-1)} \left[2r + (\frac{N(N-1)}{2}-1)\cdot 4r^2 \right], \quad (4.51)$$

so, following Equations (4.33-4.36), we have

$$\sigma(N,r) = 2\sqrt{\frac{r(1-2r)}{N(N-1)}}, \quad (4.52)$$

which is the same result as in the E_1 ensemble (*cf.* Equation (4.36)). We note that for fixed r, $\sigma(N,r) \to 0$ as $N \to \infty$, and in particular that the precision is $O(1/N)$.

As an aside we note that this agreement between E_1 and E_2 holds only for the first and second moments of the random variable for correlation integral. It can be shown that

$$\langle H(r-r_0)\cdot H(r-r_1)\cdot H(r-r_2)\rangle = 8r^3 \qquad \text{for all N in the } E_2 \text{ ensemble} \tag{4.53}$$

and

$$\langle H(r-r_{01})\cdot H(r-r_{02})\cdot H(r-r_{12})\rangle = 3r^2 \qquad \text{for N=3 in the } E_1 \text{ ensemble.} \tag{4.54}$$

There is a noticeable disagreement in expectation values of the third moment for the N=3 case. We do not know if this discrepancy holds for larger N; the algebra even at N=4 becomes very unwieldy. Our conjecture, though, is that the discrepancy vanishes in the $N\to\infty$ limit.

4.4.3.2 E_2: Shortest distance

Let $F_i(r)$ denote the probability that $r_i \le r$; for $r \le \frac{1}{2}$, this probability is given by $F_i(r)=2r$. Further, we have $P(r) = \frac{dF}{dr}$, so $P_i(r)=2$ and

$$P(r_0 r_1,\cdots,r_{D-1}) = 2^{D-1}. \tag{4.55}$$

Now, let $F_{min}(r)$ denote the probability that $r_{min} \le r$. We note that

$$\begin{aligned} F_{min}(r) &= P(r_{min}>r) && (4.56)\\ &= 1 - P(r_{min}>r) && (4.57)\\ &= 1 - P(r_0>r \textit{ and } r_1>r \textit{ and } \cdots \textit{ and } r_{D-1}>r) && (4.58)\\ &= 1 - \left[P(r_i>r)\right]^D && (4.59)\\ &= 1 - \left[1-P(r_i\le r)\right]^D && (4.60)\\ &= 1 - [1-2r]^D. && (4.61) \end{aligned}$$

Thus,

$$P_{min}(r) = \frac{dF_{min}}{dr} = 2D\,[1-2r]^{D-1}, \tag{4.62}$$

and finally,

$$\langle r_{min} \rangle \equiv \int r P_{min}(r) dr = 2D \int_0^{1/2} r[1-2r]^{D-1} \tag{4.63}$$

$$= \frac{1}{2(D+1)} = \frac{1}{N^2-N+2}, \tag{4.64}$$

a result in disagreement with Equation (4.39) for the E_1 ensemble. We note, however, that the leading $1/N^2$ behavior agrees. For large N, the ensembles act almost as though they were identical.

Intuition and example support the general conjecture that the asymptotic $N \to \infty$ behavior of relevant random variables will be the same in the E_1 and E_2 ensembles. Most of the analysis in the chapters that follow will be based on the E_2 ensemble.

4.5 Notes and References

[1] A third way is suggested in §13.6, in which a "purely nonlinear" component of the motion is replaced by white noise of the same amplitude.

[2] For example, if $X_0 \varepsilon \mathcal{A}$ is an unstable fixed point of the motion, then the orbit $X_0, f(X_0), f^2(X_0), \ldots$ will fail to sample the full attractor and it is likely that $F(X_0) \neq F$.

[3] Floris Takens. "On the numerical determination of the dimension of an attractor," in *Dynamical Systems and Bifurcations, Groningen*, 1984, Vol. 1125 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1985).

[4] It *is* possible to contrive exceptions: one that comes to mind is the expression $N^3 r_{min} - N$, which has, in the $N \to \infty$ limit, an expectation value of one in $E_1(N)$ and zero in $E_2(N)$.

[5] The symmetry over *all* permutations of the arguments is valid only for $N=3$. In general, we expect a symmetry that allows N! equal rearrangements of the distances (corresponding to the N! allowable orderings of the N points); however, there are $N(N-1)/2$ distances or $[N(N-1)/2]!$ *possible* permutations of the distance arguments for general N. Thus, it is not generally true that all permutations of the arguments are allowed for $N>3$.

[6] The notation $P(\mathbf{X})$ is the probability that statement $\mathbf{X}$ is true; we hope that this is not confused with the notation P(r), which is the probability *density* defined in §4.3.

[7] This is just the kind of "nonrigorous heuristic reasoning" we spoke of. We do point out that the solution $\langle r_{min} \rangle = 1/N^2$ was verified numerically to an accuracy of a few percent for N=3,4,5,6, and 7.

CHAPTER FIVE

5. NOISE AND DISCRETIZATION

Geometric effects arise from the lack of exact self-similarity on all scales. If an attractor is bounded, for instance, then the self-similarity must fail at large scales (these effects are discussed in Chapter Six). This chapter discusses the smallest scales and how they are affected by noise in the signal or by the digital measurement of the signal (so that a discrete and not a continuous variable is measured).

Unlike the other geometric effects that we will discuss in subsequent chapters, the limitations at small scales are fundamental in that we cannot take the $r \to 0$ limit, no matter how much data we have.

5.1 Noise

The effect of noise on the correlation integral is of obvious concern to experimentalists and has been well studied [1] in the literature. In this section, we will consider a very simple model of a low-dimensional system into which Gaussian noise has been introduced.

The effect of noise, we argue, is to convolve the attractor with a Gaussian of width σ, where σ is the amplitude of the noise. Specifically, if $\rho_{\text{quiet}}(\vec{x})$ is the natural invariant distribution of the attractor. then for the noisy case,

$$\rho_{\text{noisy}}(\vec{x}) = \int \rho_{\text{quiet}}(\vec{y})\, e^{-|\vec{x}-\vec{y}|^2/2\sigma^2} d\vec{y}. \tag{5.1}$$

We will consider a d dimensional attractor embedded in m dimensional space, but we will consider the very special case of the attractor occupying only d of the m dimensions. Adding noise to the attractor will add a size σ to each of the m directions in the embedding space. The effect on the d dimensions over which the attractor already extends will be negligible, but the effect in the m—d directions will be to increase the "thickness" from zero to σ. See Figure 5.1 for a picture of the d=1, m=2 case: the fuzzy line.

Indeed, we can think of the noisy attractor as a direct product of the d dimensional noiseless (quiet) attractor and an m—d dimensional Gaussian of width σ.

Thus, we have

$$C_{quiet}(r) = (r/R)^d \tag{5.2}$$

for small $r \ll R$, where R is the "size" of the attractor, and for Gaussian noise (see §6.2.3 and §9.2),

$$C_{noisy}(r) = [erf(r/2\sigma)]^{m-d}. \tag{5.3}$$

In the L_∞ norm, the correlation integral for the direct product of two sets is just the product of the individual correlation integrals (see Equation (3.14)).

$$C(r) = (r/R)^d \, [erf(r/2\sigma)]^{m-d} \tag{5.4}$$

As we see in Figure 5.2, the behavior of this correlation integral breaks up into distinct regimes. For $\sigma \ll r \ll R$, we have a slope of $d=\nu$ in the log-log plot of C(r) versus r; that is, $C(r) \propto r^{\nu}$. However, for $r \lesssim \sigma$, the slope gets steeper and approaches m as $r \to 0$.

The effect of noise, then, is to fill out the embedding space. As long as it is low amplitude noise, however, it leaves a fairly recognizable signature: a steepening only at small r. And there is a fairly straightforward remedy: just be sure to take the slope in the scaling regime, with $r > \sigma$.

5.2 Discretization

Typically, values of x_i measured in a time series $\{x_i\}$ are known only to finite precision, often because they have been measured digitally. And even if they *are* known to high precision, we often deliberately discretize the values in the time-series in order to improve the computational performance of our dimension algorithm [2]. Discretized time series are of the form $x_i = k_i \epsilon$, where k_i is an integer and ϵ is a discretization level. Thus, distances between pairs of discretized points will themselves be discrete multiples ϵ.

It seems likely that this built-in discretization might bias the dimension calculated from a *log* C(r) versus *log* r plot. In particular, the discretized points have a finite probability of producing $r_{ij} \equiv \| x_i - x_j \| = 0$ distances; that is, $C(0) \neq 0$.

Fortunately, there is a quick first-order correction, which unbiases the dimension estimate. Actually, we may think of it not as a correction but as the resolution of an ambiguity, for C(r) plotted as a continuous function of r

displays "steps," as shown in Figure 5.3, and it is not immediately obvious where to plot the best fit. We will show, below, a model that suggests that the appropriate plot to make is of *log* C(r) versus *log* $(r+\frac{1}{2}\epsilon)$. This recipe is independent of embedding dimension and of norm, and is very easy to implement.

5.2.1 Lattice Model

We consider points dropped randomly onto an m dimensional lattice with node spacing ϵ. We assume that ϵ is small compared to the scale of variation of the invariant density ρ. Choosing one of the nodes as an origin, and we count the number of nearby nodes including the origin whose distance to the origin is less than or equal to $r=k\epsilon$ as a function of r. Let f(m,k) denote this count. Then $C(r) \propto f(m, r/\epsilon)$. We will do this for the L_∞ and the L_1 norms; we do not know a good argument that can be used for the L_2 norm (except in the trivial case m=1), but we conjecture that the basic result, $r \mapsto r+\frac{1}{2}\epsilon$, should apply there as well.

5.2.1.1 L_∞ norm

The distance to a point $\mathbf{x} = (x_0, x_1, \ldots, x_{m-1})$ from the origin is given by $r = \max_i |x_i|$. That $|\mathbf{x}| \leq k\epsilon$ requires that $|x_i| \leq k\epsilon$ for all i. There are $2k+1$ choices of x_i satsifying $|x_i| \leq k\epsilon$, so there are $(2k+1)^m$ choices of $\mathbf{x}$ satisfying $|\mathbf{x}| \leq k\epsilon$. Thus,

$$f(m,k) = (2k+1)^m = 2^m(k+\tfrac{1}{2})^m \tag{5.5}$$

for the L_∞ norm.

5.2.1.2 L_1 norm

Here, the distance to a point $\mathbf{x} = (x_0, x_1, \ldots, x_{m-1})$ from the origin is given by

$$r = |\mathbf{x}| = |x_0| + |x_1| + |x_2| + \cdots + |x_{m-1}|. \tag{5.6}$$

In counting the number of lattice points satisfying $|\mathbf{x}| \leq k\epsilon$, we find at once that

$$f(1,k) = 2k+1 = 2(k+\tfrac{1}{2}) + \tfrac{1}{2}, \tag{5.7}$$
$$f(2,k) = 2k^2+2k+1 = 2(k+\tfrac{1}{2})^2 + \tfrac{1}{2}, \tag{5.8}$$
$$f(m,0) = 1, \tag{5.9}$$
$$f(m,1) = 1 + 2m. \tag{5.10}$$

Furthermore, to count the number of lattice points satisfying Equation (5.6), we note that fixing $x_0 = k_0\epsilon$ still leaves $m-1$ terms, which must add up to $k - |k_0|$. Thus, we can write

$$f(m,k) = \sum_{k_0=-k}^{k} f(m-1, k-|k_0|) \tag{5.11}$$

or equivalently,

$$f(m+1,k) = f(m,k) + 2\sum_{j=1}^{k} f(m,k-j) = f(m,k) + 2\sum_{i=0}^{k-1} f(m,i). \tag{5.12}$$

We will prove, given this recursion relation, the following:

$$\underline{\text{Claim}}: \quad f(m,k) = \frac{2^m}{m!}(k+\tfrac{1}{2})^m + O(k^{m-2}). \tag{5.13}$$

Proof is by induction. The claim is true for m=1 and m=2; assume inductively that it is true for m. Then for m+1,

$$f(m+1,k) = f(m,k) + 2\sum_{i=0}^{k-1} f(m,i) \tag{5.14}$$

$$= \frac{2^m}{m!}(k+\tfrac{1}{2})^m + 2\sum_{i=0}^{k-1}\frac{2^m}{m!}(i+\tfrac{1}{2})^m + O(k^{m-2}) \tag{5.15}$$

$$= \frac{2^m}{m!}k^m + 2\sum_{r=\frac{1}{2}}^{k-\frac{1}{2}}\frac{2^m}{m!}r^m + O(k^{m-1}). \tag{5.16}$$

Here we invoke the midpoint approximation for integrals:

$$\int_0^k r^m dr = \sum_{r=\frac{1}{2}}^{k-\frac{1}{2}} r^m + O\left(\int_0^k r^{m-2}dr\right) = \sum_{r=\frac{1}{2}}^{k-\frac{1}{2}} r^m + O(k^{m-1}), \tag{5.17}$$

from which,

$$f(m+1,k) = \frac{2^m}{m!}k^m + 2\cdot\frac{2^m}{m!}\int_0^k r^m dr + O(k^{m-1}) \tag{5.18}$$

$$= \frac{2^m}{m!}k^m + 2\cdot\frac{2^m}{m!}\cdot\frac{k^{m+1}}{m+1} + O(k^{m-1}) \tag{5.19}$$

$$= \frac{2^{m+1}}{(m+1)!}\left(k^{m+1} + \frac{m}{2}\cdot k^m\right) + O(k^{m-1}) \tag{5.20}$$

$$= \frac{2^{m+1}}{(m+1)!}(k+\tfrac{1}{2})^{m+1} + O(k^{m-1}) \qquad \square. \tag{5.21}$$

5.2.1.3 Implications for C(r)

In general, in the limit $\epsilon \to 0$ of no discretization, we know that $C(r) \propto r^m$ for random data. But for both the L_1 and the L_∞ norms, we have

$$f(m,k) \propto (k+\tfrac{1}{2})^m\Big[\ 1\ +\ O(k^{-2})\Big], \tag{5.22}$$

so

$$C(r) \propto (r+\tfrac{1}{2}\epsilon)^m\Big[\ 1\ +\ O(\epsilon^2)\Big]. \tag{5.23}$$

To first order in ϵ, then, we have that $C(r) \propto (r+\frac{1}{2}\epsilon)^m$, so the appropriate plot is *log* $C(r)$ versus *log* $(r+\frac{1}{2}\epsilon)$. As a final comment, we note that from our definition of correlation dimension, we have for the discrete r case that $C(r)$ is constant over any range $k\epsilon \leq r < (k+1)\epsilon$. Thus, it is an equivalent recipe to suggest that the appropriate plot is of *log* $C(r+\frac{1}{2}\epsilon)$ versus *log* $(r+\frac{1}{2}\epsilon)$, where the only points that are actually plotted are those for which $r=k\epsilon$ with integer k.

5.3 Notes and References

[1] The reader is referred to the review article by J.-P. Eckmann and D. Ruelle. "Ergodic theory of chaos and strange attractors," Rev. Mod. Phys. 57 (1985) 617.

[2] We will not say much about this strategy of discretization in this thesis, but the idea is that since the only operations performed in computing $C(r)$ are subtraction, addition, and counting, it is more efficient for most machines to do these with fixed (rather than floating) point precision. Further computational issues are discussed in Chapter Twelve.

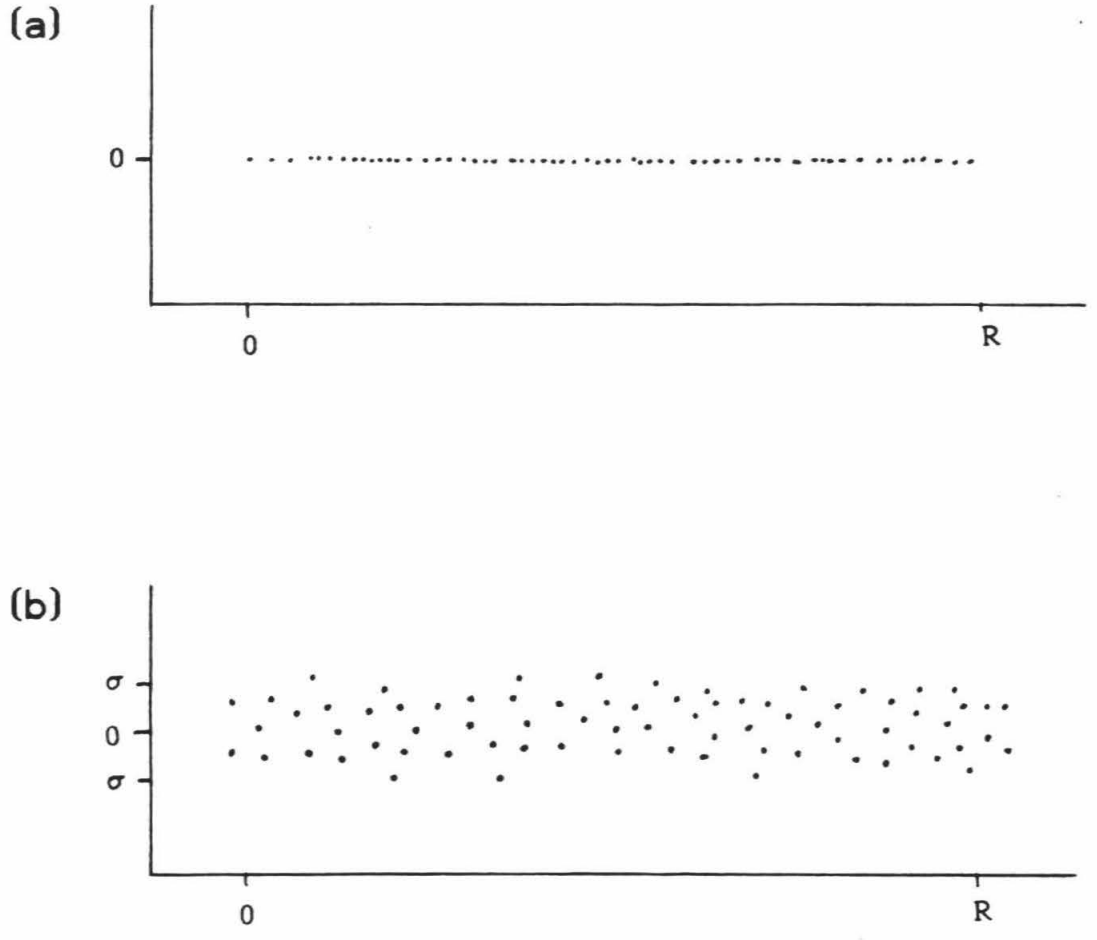


Figure 5.1 The "fuzzy" line: $m=2$, $d=1$. (a) Noiseless line segment. (b) The effect of noise is to widen the line by an amount σ in all $m-2$ directions. On scales $r \lesssim \sigma$, the fuzzy line appears two-dimensional; for $r \gg \sigma$, it appears line-like, that is, one-dimensional.

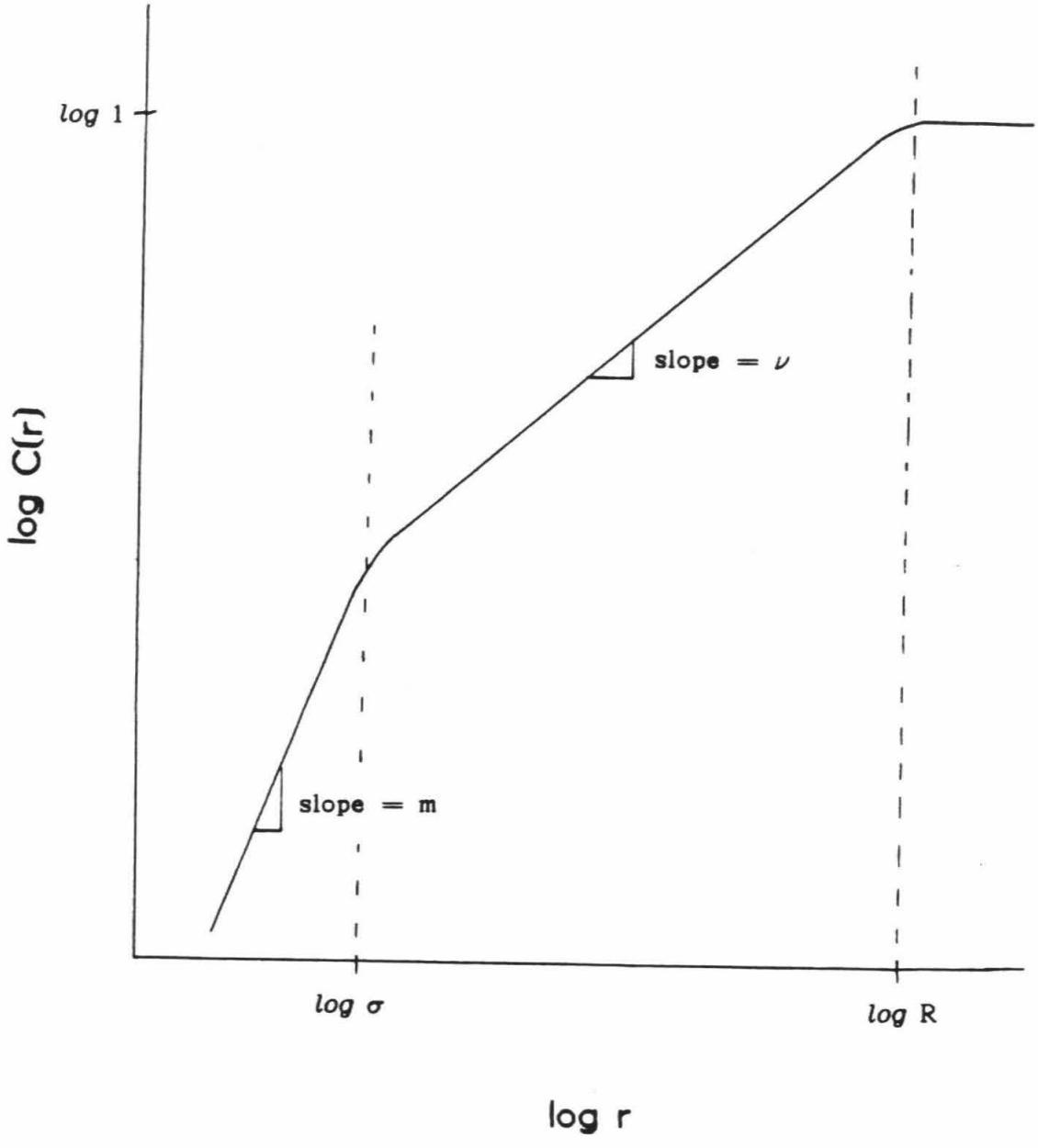


Figure 5.2 Distinct regimes in the correlation integral C(r) for noisy data. For $r \lesssim \sigma$, where σ is the amplitude of the noise, the slope will be the embedding dimension m. For $\sigma \lesssim r \lesssim R$, where R is the size of the attractor, the slope will be the attractor dimension ν. For $r \gtrsim R$, the correlation integral saturates $C(r) \to 1$, and the slope goes to zero.

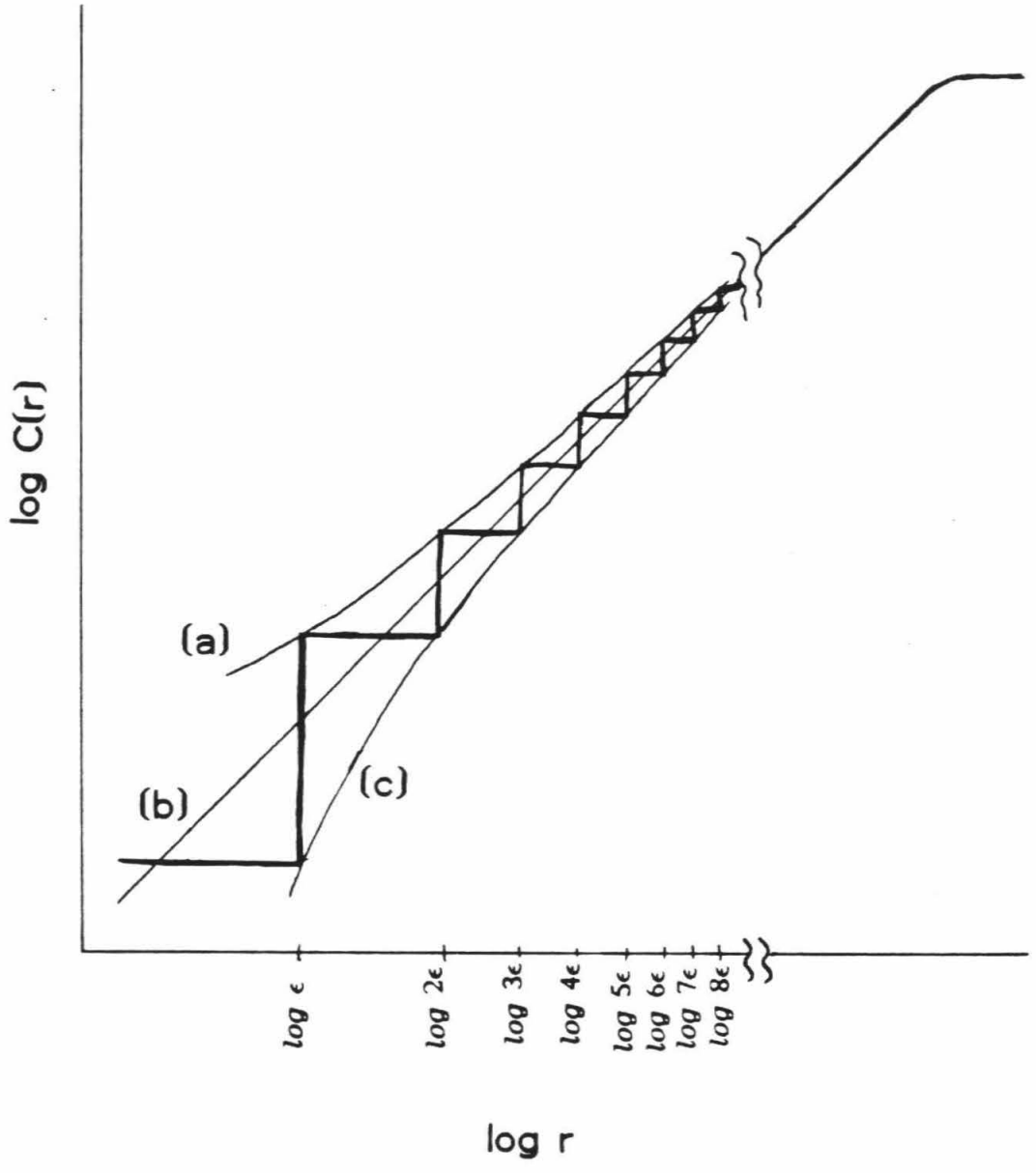


Figure 5.3 The correlation integral C(r) for discretized data. Since distances are discrete multiples of some ϵ, the actual correlation integral C(r) is a stair-step function. The smooth curves are fits through the data according to the following schemes: (a) $C(k\epsilon)$ versus $k\epsilon$; (b) $C((k+\frac{1}{2})\epsilon)$ versus $k\epsilon$; (c) $C((k+1)\epsilon)$ versus $k\epsilon$. It is pointed out in the text that (b) is equivalent to C(r) versus r for $r=(k+\frac{1}{2})\epsilon$.

CHAPTER SIX

6. EDGES, SHARP EDGES, AND SINGULARITIES

Because it is bounded, an attractor necessarily fails to be self-similar at large scales. This is inevitable, but in itself the "finite-size" effect is not so much of a problem. As long the effect is confined to large scales, $r > r_o$ for some r_o, then accurate estimates of dimension can still be obtained from the slope of a *log* C(r) versus *log* r plot in the $r \leq r_o$ regime.

The real problem stems from the "edges" that finite-sized objects in $\mathbb{R}^m$ all have. The neighborhoods around points near an edge are not similar to neighborhoods of points farther into the interior. This is an effect that occurs on *all* scales, though to a lesser degree at smaller scales. In fact, the effect vanishes as the scale goes to zero. In other words, the proportionality $C(r) \propto r^{\nu}$ becomes more accurate in the limit as $r \to 0$.

In this chapter, we will discuss the *rate* of convergence to the correct dimension as $r \to 0$. We will find that the sharper the edge, the slower the convergence, and the more noticeable the effect. We will further see how "very sharp" edges, or singularities, can drastically reduce the rate of convergence and how in some cases these edges can alter the limit itself.

For some fractal attractors, edges and singularities appear over the whole range of scales; the effect of these on the correlation integral is addressed in the next chapter.

6.1 Wraparound model

We begin with an example of a finite-sized set for which there is a "finite size" effect but no "edge" effect. However, to do this we introduce a space that is topologically distinct from $\mathbb{R}^m$. We consider the interval [0,1) with the "wraparound" metric. We identify the interval with a circle of unit circumference; distance in the wraparound metric is measured as the shortest path along the circle. Specifically,

$$d(x,y) = \min(|x-y|, 1-|x-y|). \tag{6.1}$$

On this space and in this metric (see also §4.4.1 and §8.1), distances are distributed uniformly over $[0,\frac{1}{2}]$. That is, if two points are chosen at random, the random variable corresponding to the distance between them will have a uniform distribution over the interval $[0,\frac{1}{2}]$. If not for the wraparound metric, as we will see in §6.2.1, distances could range from 0 to 1, with the small distances preferred. We can compute the correlation integral for this wraparound model; we find

$$C(r) = \begin{cases} 2r & r \leq \frac{1}{2} \\ 1 & r \geq \frac{1}{2}. \end{cases} \tag{6.2}$$

We do have saturation of $C(r) \to 1$ for large r, but the proportionality $C(r) \propto r^1$ is exact for all $r \leq \frac{1}{2}$, so the dimension will be correctly found to be one. This is the best we could expect; the models below are more generally applicable and they will show that the saturation of $C(r) \to 1$ affects the proportionality $C(r) \propto r^{\nu}$ for all nonzero r.

6.2 One-dimensional models

As a simple model to demonstrate the edge effect, consider "dropping" points on the real line $(-\infty,\infty)$ with probability density $\rho(x)$. Of course, a real dynamical process drops points according to its own deterministic rules, but the result is that the line is populated according to a "natural invariant distribution" $\rho(x)$. Other chapters will consider the dynamical means by which an invariant measure is populated or the statistical problems associated with finite samples of the invariant measure. All we will concern ourselves with here is the information that is contained in the density function $\rho(x)$. In particular, drop two points and compute the probability that the points are separated by less than r.

$$C(r) \equiv P(|x_1 - x_2| \leq r) \tag{6.3}$$

$$= \int_{-\infty}^{\infty} dx_1 \rho(x_1) \int_{-\infty}^{\infty} dx_2 \rho(x_2)\, H\,(r - |x_1 - x_2|), \tag{6.4}$$

where H(x) is the Heaviside function. (See §3.3 for a more rigorous justification of this result.) Thus,

$$C(r) = \int_{-\infty}^{\infty} dx_1 \rho(x_1) \int_{x_1 - r}^{x_1 + r} dx_2 \rho(x_2). \tag{6.5}$$

Alternatively, we can "measure" the set **S**.

$$\mathbf{S} = \left\{(x_1,x_2) \in \mathbb{R}^2 \text{ such that } |x_1 - x_2| \leq r\right\} \tag{6.6}$$

and

$$\mu(S) = \iint_S dx_1 dx_2 \; \rho(x_1)\rho(x_2). \tag{6.7}$$

This geometric interpretation is, of course, equivalent to the integral in Equation (6.5), but it is sometimes easier to visualize. We use this alternative in the next section and in §7.4.1.

6.2.1 Uniform distribution

Let us consider, as a specific $\rho(x)$, uniform distribution over the unit interval. That is, $\rho(x)=1$ for $0 \le x \le 1$, and $\rho(x)=0$ otherwise. We can follow the recipe in Equation (6.5),

$$P(\,|x_1 - x_2| \le r\,) = \int_{-\infty}^{\infty} dx_1 \rho(x_1) \int_{x_1 - r}^{x_1 + r} dx_2 \rho(x_2). \tag{6.8}$$

$$= \int_0^1 dx_1 \int_{\max(0, x_1 - r)}^{\min(x_1 + r, 1)} dx_2 \cdot \tag{6.9}$$

$$= 2r - r^2. \tag{6.10}$$

In this case, the geometric approach provides a quick result. We measure (the area of) the set defined by

$$S = \{(x_1, x_2) \in [0,1] \times [0,1] \text{ such that } |x_1 - x_2| \le r\}. \tag{6.11}$$

This is just the shaded region of Figure 6.1, and its area is clearly $1 - 2 \cdot \frac{1}{2}(1-r)^2 = 2r - r^2$. Here, $C(r) \propto r$ only if the quadratic term is negligible, $r^2 \ll 2r$, or r is

very small. If we compute the dimension by taking the slope of a *log* C(r) versus *log* r curve at a specific r, we find

$$\nu(r) = \frac{d[log\ C(r)]}{d[log\ r]} = \frac{r}{C(r)} \cdot \frac{dC}{dr} \tag{6.12}$$

$$= \frac{r}{2r-r^2}(2-2r) = \frac{1-r}{1-\frac{1}{2}r}, \tag{6.13}$$

or $\nu(r) \approx 1-\frac{1}{2}r$, for small r. Sure enough, this approaches $\nu=1$ in the limit as $r \to 0$, but more specifically it indicates how fast the calculated dimension approaches the actual dimension.

In this case, we find the approach linear in r. As this model and the one in the next section suggest, a linear approach is symptomatic of a "sharp" edge, such as this uniform $\rho(x)$ displays at its endpoints 0 and 1.

6.2.2 "Butterfly" distribution

Let us modify the above example by introducing the following density whose "edgeness" is parameterized by a.

$$\rho(x) = \begin{cases} (1-a) - (1-2a)\,|x| & \text{for } -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{6.14}$$

First of all, note that $a=\frac{1}{2}$ is just the uniform distribution studied above. For $a=0$, we have a density function that is triangular and essentially "edgeless." For $a=1$ we have a butterfly-shaped density function with maximum-sized edges. See Figure 6.2. The integration is straightforward to show that

$$C(r) = r - \frac{a}{2}r^2 - \tfrac{1}{2}(1-2a)r^3, \tag{6.15}$$

from which

$$\nu(r) = 1 - \frac{a}{2}r - (1-2a-\frac{a^2}{4})r^2 + O(r^3). \tag{6.16}$$

We see that the coefficient of the leading term of $\nu(r)-1$ is directly proportional to the "edgeness" parameter, a. In particular, $a=0$ (no edge) eliminates the linear term and provides a *quadratic* approach for $\nu(r)\to 1$ as $r\to 0$. Conversely, $a=2$ (with the most edge) has the slowest approach.

6.2.3 Gaussian distribution

Consider as a final example the Gaussian density:

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}. \tag{6.17}$$

There is still an edge (in the sense that $<x^2>$ is bounded), but it is "softer" than the edge for the uniform density. We find in this case

$$C(r) = \int_{-\infty}^{\infty} dx_1\rho(x_1) \int_{x_1-r}^{x_1+r} dx_2\rho(x_2) \tag{6.18}$$

$$= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} dx_1 e^{-x_1^2/2\sigma^2} \int_{x_1-r}^{x_1+r} dx_2 e^{-x_2^2/2\sigma^2} \tag{6.19}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} dx_1 \int_{-r/2\sigma}^{r/2\sigma} dx_2\ e^{-x_1^2-(x_1+x_2)^2} \tag{6.20}$$

$$= \frac{1}{2\pi} \int_{-r/2\sigma}^{r/2\sigma} dx_2 \int_{-\infty}^{\infty} dx_1\ e^{-2(x_1+\frac{1}{2}x_2)^2-\frac{1}{2}x_2^2} \tag{6.21}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-r/2\sigma}^{r/2\sigma} dx_2 \; e^{x_2^2/2} \tag{6.22}$$

$$= erf(r/2\sigma), \tag{6.23}$$

where erf is the error function. $erf(x)$ is linear with x for small x, but saturates to 1 as x is large. So again, we have the appropriate levelling off of C(r) to 1 as r gets large, and further,

$$\nu(r) = \frac{r}{erf(r/2\sigma)} \cdot \frac{1}{2\sigma} e^{-r^2/2\sigma^2} \approx 1 - \frac{5}{3}(r/2\sigma)^2, \tag{6.24}$$

so the approach of $\nu(r) \to 1$ as $r \to 0$ is quadratic in r.

6.2.4 General one-dimensional distribution

In general, if $\rho(x)$ is a smooth function of x, then $\nu - 1 = O(r^2)$ for small r. In particular, we can show that

$$C(r) = 2r \cdot \int_{-\infty}^{\infty} [\rho(x)]^2 \, dx - \frac{r^3}{3} \cdot \int_{-\infty}^{\infty} [\rho'(x)]^2 \, dx + O(r^4) \tag{6.25}$$

and

$$\nu(r) = 1 - \frac{r^2}{3} \cdot \frac{\int [\rho'(x)]^2 dx}{\int [\rho(x)]^2 dx} + O(r^3) \tag{6.26}$$

so that as long as $\rho'(x) \equiv \frac{d\rho}{dx}$ is bounded, we will have quadratic convergence of $\nu(r) \to \nu$ as $r \to 0$. Also, in this one-dimensional case, we have $\nu(r) < \nu$ strictly for nonzero r.

6.3 A two-dimensional model

We have seen in one dimension that the effect of the edge is usually to underestimate the dimension. We present an example here of a one-dimensional set embedded in $\mathbb{R}^2$ for which the effect of the edge is to cause the dimension to be overestimated.

Consider a uniformly dense circle of radius R. We will take the distance between two points on the circle to be given by the two-dimensional Euclidean norm [1]; that is,

$$d((x_1,y_1), (x_2,y_2)) = \sqrt{(x_1-x_2)^2 + (y_1-y_2)^2}. \tag{6.27}$$

We find

$$C(r) = P(\|\vec{x}_1-\vec{x}_2\|\leq r) = \frac{2}{\pi} \arcsin(r/2R), \tag{6.28}$$

for which

$$\nu \approx 1 + \frac{\pi^2}{12}(r/2R)^2 \tag{6.29}$$

for $r \ll R$. Again, we see quadratic convergence, indicative of a "soft" edge, but ν is actually overestimated. See Figure 6.3.

6.4 A higher dimensional model

Here, we consider the very particular case of a $\nu = m$ dimensional set embedded in $\mathbb{R}^m$ and formed from the direct product of m identical one-dimensional sets. That is, we consider the invariant distribution

$$\rho_m(x_0,x_1,\ldots,x_{m-1}) = \rho_1(x_0)\rho_1(x_1)\cdots\rho_1(x_{m-1}), \tag{6.30}$$

where $\rho_1(x)$ is the one-dimensional invariant distribution. We know in this case (see §3.2) that the correlation integral $C^m(r)$ is just the product of the m one-dimensional correlation integrals. Thus,

$$C^m(r) = [C(r)]^m, \tag{6.31}$$

and the estimated dimension $\nu_m(r)$ is given by

$$\nu_m(r) = \frac{r}{C^m(r)}\frac{dC^m(r)}{dr} = \frac{r}{[C(r)]^m}\frac{d[C(r)]^m}{dr} = \frac{r}{[C(r)]^m}\, m[C(r)]^{m-1}\frac{dC(r)}{dr} \tag{6.32}$$

$$= m\frac{r}{C(r)}\frac{dC(r)}{dr} = m\nu_1(r), \tag{6.33}$$

so

$$\frac{\nu_m(r)}{m} = \nu_1(r) \tag{6.34}$$

or

$$\frac{\nu_m(r)-m}{m} = \nu_1(r)-1. \tag{6.35}$$

What this tells us, somewhat surprisingly, is that the relative error in the approximation of ν by $\nu(r)$ depends only on r and is independent of the dimension ν (=m in this case). We caution that this does *not* imply that higher dimensions are in practice as easy to obtain accurately as lower dimensions; on the contrary, at higher dimensions it is more difficult to obtain the small r behavior of the correlation integral (as we discuss in Chapter Eleven), so that at higher

dimensions the edge effect is even more noticeable.

6.5 Singularities

Again, we consider one-dimensional density functions $\rho(x)$, but with edges that are *very* sharp. We've seen that the "sharper" the edge, the poorer the approximation for ν. Here, we investigate singular edges, whose effect should be even worse.

6.5.1 Logistic map

In [2], Grassberger and Procaccia point out that the natural invariant distribution of the logistic map $x_{n+1}=4x_n(1-x_n)$ is given by the density

$$\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \tag{6.36}$$

which has singular edges at $x=0$ and $x=1$. The correlation integral for this density distribution can (after a little bit of work) be shown to be

$$C(r) \approx \frac{4}{\pi^2}\, r \left[\ln(\tfrac{4}{r}) + 1\right] + O(r^2), \tag{6.37}$$

which leads to a *logarithmically* slow convergence of $\nu(r)$ to 1.

$$\nu(r) \approx 1 - \frac{1}{\ln(\frac{4}{r})+1} \tag{6.38}$$

In their paper, Grassberger and Procaccia describe a "remedy" for this effect, which involves embedding the invariant set into $\mathbb{R}^m$ with $m>1$. The remedy

does reduce the coefficient of the logarithmic term, but the logarithm remains. The singularity acts as a very sharp edge, slowing the convergence of ν to 1.

6.5.2 Power law singularities

The above example had a singularity at the edges proportional to $x^{-1/2}$; we say that it has a singularity of index $-\frac{1}{2}$. We consider now the density function

$$\rho(x) = \begin{cases} \frac{x^{\alpha} e^{-x}}{\Gamma(\alpha+1)} & \text{for } x>0 \\ 0 & \text{for } x\leq 0, \end{cases} \tag{6.39}$$

which has a singularity at the origin of index α. The factor of $\frac{e^{-x}}{\Gamma(\alpha+1)}$ is just to ensure that $\int \rho(x)dx = 1$. Computing the correlation integral, we find for small r that

$$C(r) \propto r^{2\alpha+2} \qquad \text{for } -1<\alpha<-\tfrac{1}{2} \tag{6.40}$$

$$\propto r \log r \qquad \text{for } \alpha=-\tfrac{1}{2} \tag{6.41}$$

$$\propto r + O(r^2) \qquad \text{for } -\tfrac{1}{2}<\alpha<\tfrac{1}{2} \tag{6.42}$$

$$\propto r + O(r^3) \qquad \text{for } \tfrac{1}{2}<\alpha. \tag{6.43}$$

For $\alpha\leq -1$, $\Gamma(\alpha+1)$ does not exist and $\int \rho(x)dx$ cannot be bounded. What is notable is that too sharp an edge actually not only affects the approach of ν to its limit as $r\to 0$, but affects the limit as well. Even the *dimension* of the set is affected by very singular edges. This may seem counterintuitive at first, but remember that it is the correlation dimension D_2 that is being measured, and that depends on the invariant density, whereas the capacity D_0 or Hausdorff

dimension D_H depends only on the underlying set (the "support" of the invariant measure), and these, in fact, *are* insensitive to singularities in the invariant measure. In [3], it is shown that the generalized dimension, D_q, of a power law singularity α is given by

$$D_q = \min\left(\frac{q(\alpha+1)}{q-1}, 1 \right). \tag{6.44}$$

For the correlation dimension $q=2$, and substitution into this equation repeats our result. See Figure 6.4.

6.6 Notes and References

[1] We find the same result with other norms, but the derivation isn't as clear.

[2] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors," Physica 9D (1983) 189.

[3] Thomas C. Halsey, Mogens H. Jensen, Leo P. Kadanoff, Itamar Procaccia, and Boris I. Shraiman. "Fractal measures and their singularities: the characterization of strange sets," Phys. Rev. A 33 (1986) 1141.

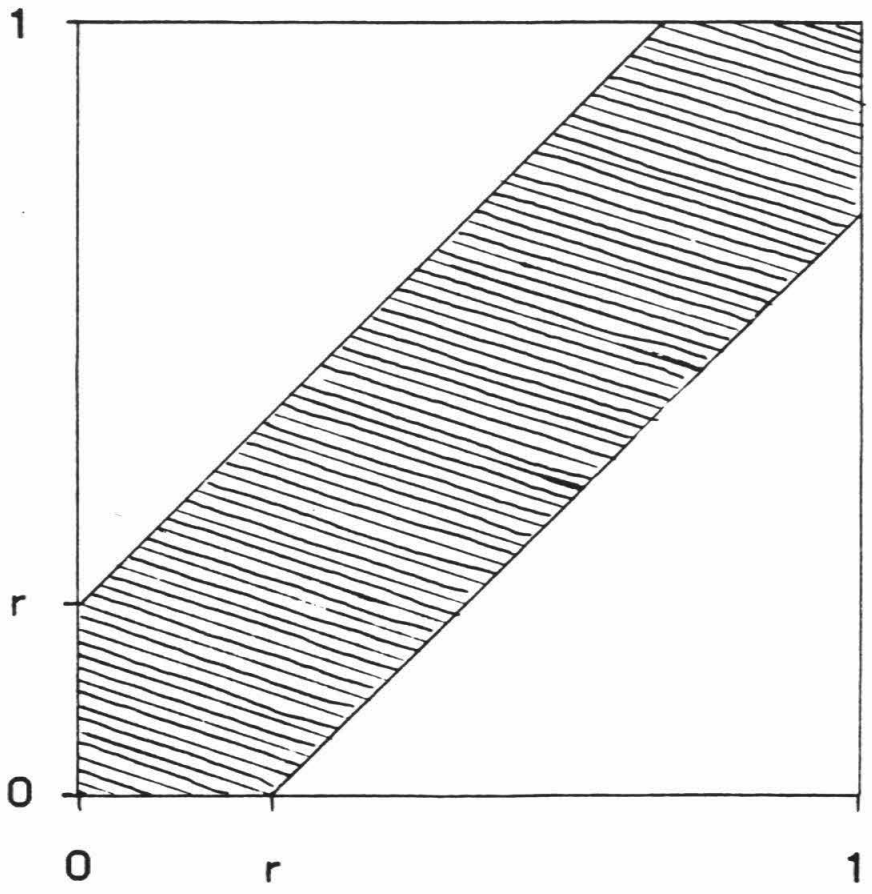


Figure 6.1 The set $S=\{(x_1,x_2)\in[0,1]\times[0,1]$ such that $|x_1-x_2|<r\}$ is the shaded region. Note that the area is $1-2[\frac{1}{2}(1-r)^2] = 2r-r^2$.

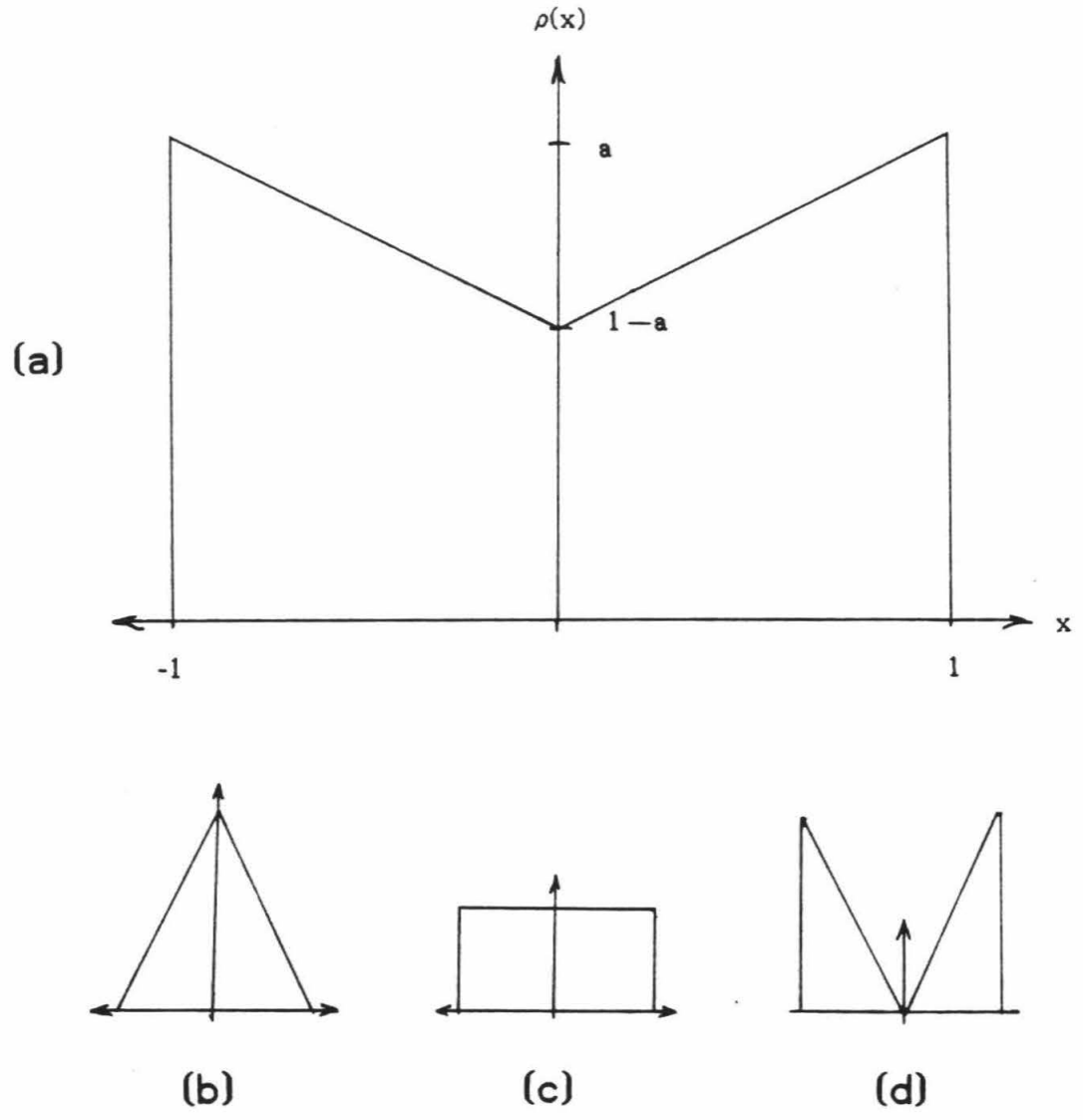


Figure 6.2 "Butterfly" density $\rho(x)$ on interval $[-1,1]$. (a) for arbitrary edge parameter a. (b) for $a=0$, or no edge; (c) for $a=\frac{1}{2}$, or uniform distribution; (d) for $a=1$, or maximum edge.

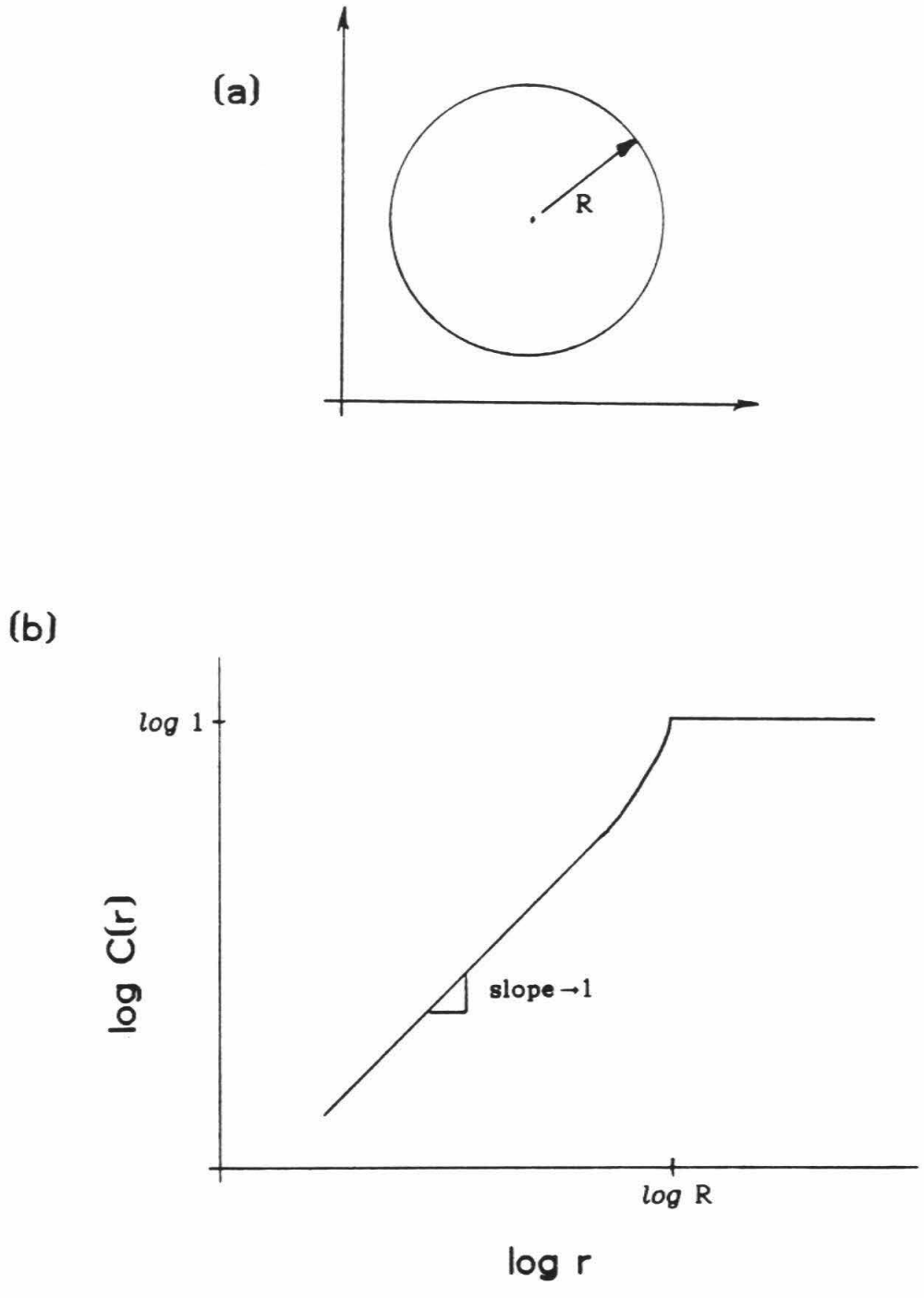


Figure 6.3 (a) Uniformly dense circle in L_2. (b) Correlation integral C(r).

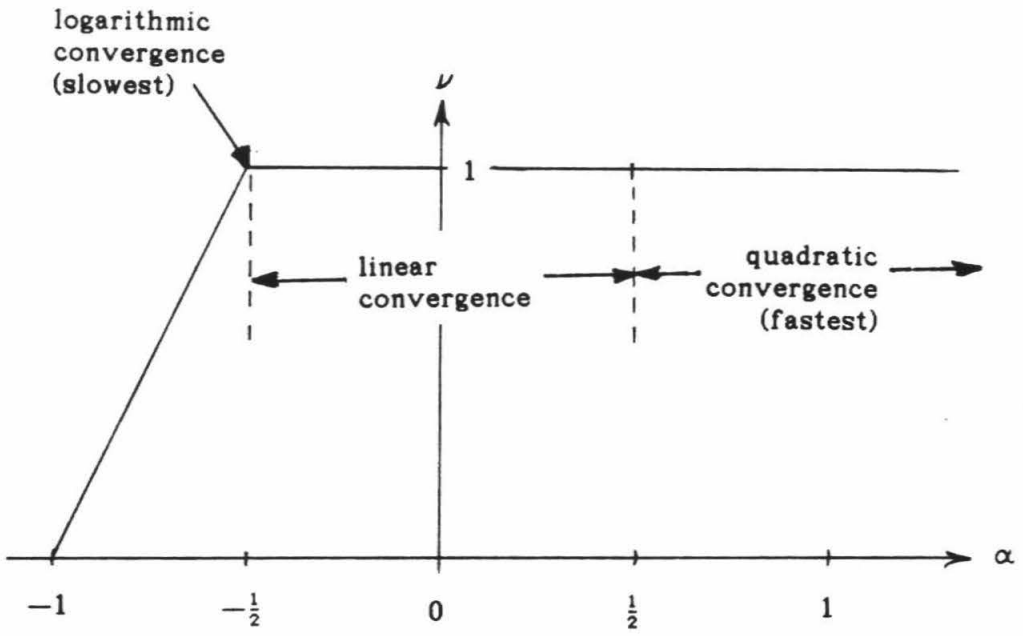


Figure 6.4 Convergence to correlation dimension ν for singularity index α.

CHAPTER SEVEN

7. LACUNARITY IN FRACTAL SETS

While points, lines, and planes are *continuously* self-similar, in that they look identical under any magnification, sets with fractional dimension usually display a "texture" or "lacunarity" that leads to self-similiarity only at particular magnifications. For instance, the fractal set in Figure 7.1 looks like itself only when magnified by a power of two. The strange attractors of many dissipative dynamical systems display such a textured fractal structure.

In [1,2], it is shown empirically that this property of lacunarity can lead to an oscillation in the correlation integral that degrades the accuracy of the dimension estimate. In this chapter, we introduce a Cantor set model that exhibits lacunarity, and we find for this model an implicit analytical expression for the oscillations in the correlation integral [3]. Finally, we show examples of dynamical maps that display these oscillations.

7.1 Cantor sets

7.1.1 Standard Cantor set

We begin with a definition of what we will call the "standard" Cantor set. This set is usually constructed in stages by removing the middle third from the segments remaining in the previous stage. Thus, $C_0=[0,1]$ becomes $C_1=\left[0,\frac{1}{3}\right]\cup\left[\frac{2}{3},1\right]$ in the first stage. The segments of the first stage are sliced to give $C_2=\left[0,\frac{1}{9}\right]\cup\left[\frac{2}{9},\frac{1}{3}\right]\cup\left[\frac{2}{3},\frac{7}{9}\right]\cup\left[\frac{8}{9},1\right]$ in the second stage, and so on. The Cantor set itself is just the limit

$$C = \bigcap_{i=0}^{\infty} C_i. \tag{7.1}$$

7.1.2 Weighted Cantor sets

Next, we generalize this definition of the Cantor set, and at the same time, impose a measure μ upon it. To construct a weighted Cantor set, we first specify a nonnegative partition of unity $\mathbb{P}=(p_0,p_1,p_2,\cdots,p_{n-1})$, with $p_0+p_1+p_2+\cdots+p_{n-1}=1$ and $p_s\geq 0$ for all s. For instance, the choice $\mathbb{P}=(\frac{1}{2},0,\frac{1}{2})$ leads to the standard Cantor set defined in the previous section.

Divide the unit interval into n equal segments $\left[0, \frac{1}{n}\right]$, $\left[\frac{1}{n}, \frac{2}{n}\right]$, ..., $\left[\frac{n-1}{n}, 1\right]$. Define the measure of these segments in terms of the "probabilities" p_s.

$$\mu\left(\left[\frac{s}{n}, \frac{s+1}{n}\right]\right) = p_s \tag{7.2}$$

At the next stage, we divide each of these intervals into n subintervals, each of length $1/n^2$, and assign them measures according, again, to the probabilities p_s. Here, taking the s_2 subinterval of the s_1 interval (see Figure 7.2),

$$\mu\left(\left[\frac{ns_1+s_2}{n^2}, \frac{ns_1+s_2+1}{n^2}\right]\right) = p_{s_1}p_{s_2}. \tag{7.3}$$

The procedure is carried out for all intervals of the form $\left[\frac{k}{n^m}, \frac{k+1}{n^m}\right]$. If we write the base n expansion for k/n^m,

$$\frac{k}{n^m} = \text{“}0.s_1s_2\cdots s_m\text{”} \equiv \sum_{j=1}^{m} s_j n^{-j}, \tag{7.4}$$

then

$$\mu\left(\left[\frac{k}{n^m}, \frac{k+1}{n^m}\right]\right) = \prod_{j=1}^{m} p_{s_j} = p_{s_1} p_{s_2} \cdots p_{s_m}. \tag{7.5}$$

Having defined the measure on these basis intervals, we can find the measure of any interval (since any interval can be expressed as a countable disjoint union of the basis intervals) and therefore the measure of any conventionally measurable subset of the unit interval.

This measure defines the weighted Cantor set.

The "support" of the measure μ is the set composed of points whose neighborhoods have nonzero measure; alternatively, the support may be defined as the complement of the union of all open sets $\mathcal{O}$ for which $\mu(\mathcal{O})=0$. We note that if all of the probabilities are nonzero, then every interval has a positive measure, and the "support" of the set is the full unit interval. On the other hand, if any of the middle probabilities $(p_1,\ldots,p_{n-2})$ are zero, the support will be a fractal Cantor set.

Having defined a measure on the unit interval, we can associate a density $\rho(x)$, which satisfies

$$\mu([a,b]) = \int_a^b \rho(x)\, dx. \tag{7.6}$$

We note that $\rho(x)$ is a "distribution," not a proper function (in the same sense that the Dirac "delta function" is not a proper function).

At each stage, we have divided intervals into n *equal* segments. Weighted Cantor sets can be defined just as easily in terms of inequal segment lengths, but we have avoided this generalization because it makes more difficult the

expressions for oscillations in the correlation integral.

We note finally that the special case of these weighted Cantor sets in which all nonzero probabilities are equal — *e.g.*, $\mathbb{P}=(\frac{1}{5},0,0,\frac{1}{5},0,\frac{1}{5},\frac{1}{5},0,\frac{1}{5})$ — describe the class of Cantor sets considered in [1].

7.2 Generating Sample points

Having defined the weighted Cantor measure μ, we may ask if there is a dynamical system for which μ is the natural invariant measure. First of all, just knowing that there *is* a such a dynamical system, we are more justified in studying the invariant measure. It lends credence to the notion that the properties this measure exhibits are relevant to the natural invariant measures of common dynamical systems. Secondly, given the dynamical system, we have a direct procedure for generating a series of points that sample the interval with a distribution $\rho(x)$. This is useful for numerical studies.

7.2.1 Iterated Function System (IFS)

The dynamical system defined by the following map of the unit interval into itself has a natural invariant measure that corresponds to the μ defined by the weighted Cantor set.

$$f: x \mapsto \frac{1}{n}x+\mathfrak{K}, \tag{7.7}$$

where

$$\mathfrak{K} = \begin{cases} 0 & \text{with probability } p_0 \\ \frac{1}{n} & \text{with probability } p_1 \\ & \cdots \\ \frac{n-1}{n} & \text{with probability } p_{n-1}. \end{cases} \tag{7.8}$$

This is an example of what Barnsley and Demko [4] call an Iterated Function System or an IFS. The IFS has many interesting properties in its own right; among other uses, it is currently being applied to image compression and computer graphics [5]. It provides an efficient way to populate the unit interval with a sample of points corresponding to the measure of the weighted Cantor set discussed above (also it provides the nomenclature we use for referring to the partition of unity as a set of "probabilities").

We can create a time series, a list of numbers $\{x_j\}$ that populate the unit interval, with iterates of the stochastic map $x_{j+1}=f(x_j)$ [6]. Note that an iteration of the map contracts volume, so an attractor is expected. In fact, after the initial transients die out, the limiting set is a weighted Cantor set with the above measure.

It is clear, for instance, that the probability for a point x to be mapped into the segment $f(x)\varepsilon\left[\frac{s}{n}, \frac{s+1}{n}\right]$ is p_s. Furthermore, for two iterates of a point x to be mapped into the narrower segment $f^2(x)\varepsilon\left[\frac{ns_1+s_2}{n^2}, \frac{ns_1+s_2+1}{n^2}\right]$ the first must have been mapped to the segment $f(x)\varepsilon\left[\frac{s_2}{n}, \frac{s_2+1}{n}\right]$ and the second iterate from this segment into $f(f(x))\varepsilon\left[\frac{s_1}{n}, \frac{s_1+1}{n}\right]$. The probability for both to occur is the product $p_{s_1}p_{s_2}$, which is just the measure of the interval. In general, the probability after "many" iterates for a point to be mapped into any interval is given by the measure of that interval. If the interval is specified to a precision of $1/n^m$, then m iterates are sufficiently many.

Because computing the value of the next x_j, given the history $x_{j-1}, x_{j-2}, \cdots$,

involves the choice among n objects with probabilities $p_0,p_1,\ldots,p_{n-1}$, we can write down the entropy per step of the sequence:

$$S = -\sum_{k=0}^{n-1} p_k \log p_k. \tag{7.9}$$

That this is a finite entropy tells us that successive values of x are not independent.

7.2.2 Independent sample points

Another method for generating points to sample the weighted Cantor measure is the following. Choose the sequence $s_1,s_2,s_3,\ldots$ randomly and independently, each from the set $\{0,1,\ldots,n-1\}$ with a probability for $s_j=s$ given by p_s. Then let

$$x = \sum_{j=1}^{\infty} s_j n^{-j} \tag{7.10}$$

be the point. To get more points just repeat the process.

In practice, of course, only a finite sequence $s_1,s_2,\ldots,s_m$ is taken for each point, and the points x sample the measure to a precision of n^{-m}. Still, this is a lot of choosing; there are m choices for each point; the IFS above made only one choice per point. In fact, to that same precision of n^{-m}, we can express this method of generating points with the map f^m, where f is the IFS. The entropy per step of this method is

$$S = -m \sum_{k=0}^{n-1} p_k \log p_k. \tag{7.11}$$

For large m, the entropy per step is high. Successive values of x by this method are essentially independent. For the IFS, successive values are correlated, but values well separated in time become increasingly less correlated.

7.3 Pointwise dimension at the origin

Recall the definition (§2.4.1) of pointwise dimension at x.

$$d_x \equiv \lim_{r \to 0} \frac{\log C_x(r)}{\log r}, \tag{7.12}$$

where $C_x(r) \equiv \mu(B_x(r))$ is the measure of a ball of radius r centered at x. At the origin of the unit interval [0,1] with the weighted Cantor measure

$$d_o = \lim_{r \to 0} \frac{\log \mu([0,r])}{\log r}. \tag{7.13}$$

7.3.1 The Fraction function

We define a function F(r), which we call the "fraction" of the weighted Cantor set, by

$$F(r) \equiv C_{x=0}(r) = \mu([0,r]), \tag{7.14}$$

for $0 \leq r \leq 1$. F(r) denotes the fraction of the Cantor set within r of the origin. For completeness, and what will turn out to be ease of notation, we further define

$$F(r) = 0, \text{ for } r \leq 0, \qquad F(r) = 1, \text{ for } r \geq 1. \tag{7.15}$$

7.3.2 Standard Cantor Set

We'll first compute F(r) for the standard Cantor set, and then generalize. What we do is break up the Cantor set into the union of two smaller Cantor sets, each of which is precisely similar to the original Cantor set.

Let $C[a,b]$ denote the intersection of the Cantor set C with the interval [a,b], that is $C[a,b] \equiv C \cap [a,b]$. Then our union looks like

$$C[0,1] = C\left[0,\tfrac{1}{3}\right] \cup C\left[\tfrac{2}{3},1\right], \tag{7.16}$$

and the similarity maps are defined

$$f_1\colon C[0,1] \to C\left[0,\tfrac{1}{3}\right] \qquad \text{by } f_1(x)=\tfrac{1}{3}x\ ; \tag{7.17}$$

$$f_2\colon C[0,1] \to C\left[\tfrac{2}{3},1\right] \qquad \text{by } f_2(x)=\tfrac{1}{3}x + \tfrac{2}{3}\ . \tag{7.18}$$

Now let

$F_1(r) \equiv$ fraction of points on Cantor subset $C\left[0,\tfrac{1}{3}\right]$ within r of the origin 0;
$F_2(r) \equiv$ fraction of points on Cantor subset $C\left[\tfrac{2}{3},1\right]$ within r of the point $\tfrac{2}{3}$.

In this case, since $C\left[\tfrac{2}{3},1\right]$ is just a translation of $C\left[0,\tfrac{1}{3}\right]$, we have $F_1(r)=F_2(r)$. Also, by the definition of "fraction" we have endpoints: $F(0)=0$, $F(1)=1$; and $F_1(0)=F_2(0)=0$, and $F_1(\tfrac{1}{3})=F_2(\tfrac{1}{3})=1$. The similarity relations allow us to express

$F_1(r)$ and $F_2(r)$ completely in terms of $F(r)$.

$$F_1(r) = F_2(r) = F(3r) \tag{7.19}$$

The geometry of the placement of $C\left[0,\frac{1}{3}\right]$ and $C\left[\frac{2}{3},1\right]$ provides an expression for $F(r)$ in terms of $F_1(r)$ and $F_2(r)$.

$$F(r) = \begin{cases} \frac{1}{2}F_1(r) & \text{for } 0 \le r \le \frac{1}{3} \\ \frac{1}{2} & \text{for } \frac{1}{3} \le r \le \frac{2}{3} \\ \frac{1}{2} + \frac{1}{2}F_2(r-\frac{2}{3}) & \text{for } \frac{2}{3} \le r \le 1 \end{cases} \tag{7.20}$$

We can combine these with Equation (7.19) into a single implicit equation for $F(r)$

$$F(r) = \frac{1}{2}\left[F(3r) + F(3r-2)\right], \tag{7.21}$$

where we have made use of Equation (7.15) to simplify the expression.

This may be "solved" for $F(r)$ by taking a sequence of functions $F^{(0)}(r)$, $F^{(1)}(r)$, etc., defined by

$$F^{(j+1)}(r) = \frac{1}{2}\left[F^{(j)}(3r) + F^{(j)}(3r-2)\right], \tag{7.22}$$

where $F^{(0)}$ is any function that satisfies the boundary conditions $F^{(0)}(0)=0$, $F^{(0)}(1)=1$; for example, $F^{(0)}(r)=r$. Then,

$$F(r) = \lim_{j\to\infty} F^{(j)}(r). \tag{7.23}$$

More practically, the exact value of F(r) may be obtained at a discrete set of points with only a finite number of iterations. Begin with F(0)=0, F(1)=1, follow by evaluation at $r=\frac{1}{3}$ and $r=\frac{2}{3}$, then at $r=\frac{1}{9}, \frac{2}{9}, \frac{4}{9}, \frac{5}{9}, \frac{7}{9}$, and $\frac{8}{9}$, and so on. In each case, evaluation at $r=\frac{k}{3^n}$ can be expressed in terms of values already computed at r's of the form $\frac{k'}{3^{n-1}}$; *e.g.*,

$$\begin{aligned} F(\tfrac{20}{27}) &= \tfrac{1}{2} + \tfrac{1}{2}F(3\cdot\tfrac{20}{27} - 2) && = \tfrac{1}{2} + \tfrac{1}{2}F(\tfrac{2}{9}) \\ &= \tfrac{1}{2} + \tfrac{1}{2}\left[\tfrac{1}{2}F(3\cdot\tfrac{2}{9})\right] && = \tfrac{1}{2} + \tfrac{1}{2}\left[\tfrac{1}{2}F(\tfrac{2}{3})\right] \\ &= \tfrac{1}{2} + \tfrac{1}{2}\left[\tfrac{1}{2}\left[\tfrac{1}{2}+F(3\cdot\tfrac{2}{3}-2)\right]\right] && = \tfrac{1}{2} + \tfrac{1}{2}\left[\tfrac{1}{2}\left[\tfrac{1}{2}+F(0)\right]\right] \\ &= \tfrac{5}{8}. \end{aligned} \tag{7.24}$$

This provides a way to very quickly get a plot F(r) versus r, such as is shown in Figure 7.3. This is a graph of the "Devil's staircase." It is a staircase because the function increases monotonically, and the slope $\frac{dF}{dr}$ is almost everywhere zero. But the function, amazingly enough, is continuous — the Devil need never lift his foot!

From the recursion in Equation (7.21), we can determine F(r) well enough to find the pointwise dimension of the Cantor set at the origin.

$$d_o \equiv \lim_{r\to 0} \frac{log\ F(r)}{log\ r} \tag{7.25}$$

For $r<\frac{1}{3}$, we have $F(r) = \frac{1}{2}F(3r)$, and for $r<\frac{1}{3^k}$, we have $F(r) = \frac{1}{2^k}F(3^k r)$, so

$$\frac{log\ F(r)}{log\ r} = \frac{log\left[\frac{1}{2^k}F(3^k r)\right]}{log\ r} = \frac{log\ F(3^k r) - k\ log\ 2}{log\ 3^k r - k\ log\ 3}. \tag{7.26}$$

If for a given r we take k so that $\frac{1}{3} \leq 3^k r < 1$, then

$$log\ \tfrac{1}{3} \leq log\ 3^k r < log\ 1 \tag{7.27}$$

and

$$log\ F(\tfrac{1}{3}) \leq log\ F(3^k r) < log\ F(1), \tag{7.28}$$

so

$$\frac{log\ F(\frac{1}{3}) - k\ log\ 2}{log\ 1 - k\ log\ 3} \leq \frac{log\ F(r)}{log\ r} < \frac{log\ F(1) - k\ log\ 2}{log\ \frac{1}{3} - k\ log\ 3}. \tag{7.29}$$

Finally, since $r \to 0$ implies $k \to \infty$, we have rigorously that

$$d_o \equiv \lim_{r \to 0} \frac{log\ F(r)}{log\ r} = \frac{log\ 2}{log\ 3} \approx 0.6309. \tag{7.30}$$

It is worth noting that the following "alternate" definition of d_o involving the slope of a *log* F(r) vs. *log* r plot leads to a limit that does not exist.

$$d_o \stackrel{?}{=} \lim_{r \to 0} \frac{d[log\ F(r)]}{d[log\ r]} = \lim_{r \to 0} \frac{r}{F(r)} \frac{dF}{dr} \tag{7.31}$$

The problem is caused by lacunarity in the Cantor set. For if the limit exists, then L'Hopital's rule assures us that the limit is $d_o = (log\ 2)/(log\ 3)$. But if we consider the set of points 1/2, 1/6, ..., $1/2 \cdot 3^k$, all of which are in voids of the Cantor set, we have that $\frac{dF}{dr} = 0$ at each of those points, implying that the limit is $d_o = 0$, a contradiction.

Looking at a log-log plot of F(r) vs. r, Figure 7.4, we can see why this limit fails. There is a periodic undulation to the curve whose "average slope" is well defined as $d_o = (log\ 2/log\ 3)$. This effect provides another argument in

favor of a wide scaling regime in r; to numerically determine the average slope, we want to see as many "periods" of the fraction curve as possible.

7.3.3 Weighted Cantor Set

At this point we can generalize. Compute the fraction F(r) for the weighted Cantor set specified by the partition $\mathbb{P}=(p_0,p_1,\ldots,p_{n-1})$. We can write down the implicit expression for F(r) in terms of F(nr−k). Following Equation (7.21),

$$F(r) = p_0F(nr) + p_1F(nr-1) + \ldots + p_{n-1}F(nr-(n-1)) \tag{7.32}$$

$$= F(\tfrac{s}{n}) + p_sF(nr-s), \quad \text{for } \tfrac{s}{n}\le r\le\tfrac{s+1}{n}. \tag{7.33}$$

And again, we can compute F(r), starting at the endpoints F(0)=0, F(1)=1, and then at $r=\frac{1}{n}, \frac{2}{n},\ldots$, then $r=\frac{1}{n^2}, \frac{2}{n^2},\ldots$, etc.

Equation (7.32) is sufficient to define F(r) at arbitrary r. We can, however, give a direct series expression for F(r), where r is expressed in base n notation,

$$r = 0.s_1s_2\ldots s_j\ldots = \sum_{j=1}^{\infty} s_jn^{-j}, \tag{7.34}$$

with $0\le s_j<n$ as usual. Note that $s_1n^{-1} \le r < (s_1+1)n^{-1}$, so

$$F(r) = F(s_1) + p_{s_1}F(nr-s_1). \tag{7.35}$$

where, in general,

$$F(s_k) = P_{s_k} \equiv p_1 + p_2 + \cdots + p_{s_k-1}. \tag{7.36}$$

Also, $s_2 n^{-1} \leq nr - s_1 < (s_2+1)n^{-1}$, so

$$F(r) = P_{s_1} + p_{s_1}\left[P_{s_2} + p_{s_2}F(n^2 r - ns_1 - s_2)\right] \tag{7.37}$$

and so on;

$$F(r) = P_{s_1} + p_{s_1}P_{s_2} + \cdots + p_{s_1}p_{s_2}\cdots p_{s_k}P_{s_{k+1}} + \cdots, \tag{7.38}$$

an expression which, along with Equation (7.36), gives F(r) in terms of $p_0, p_1, \cdots, p_{n-1}$.

Finally, we are able to find the pointwise dimension at the origin from $F(r) = p_0 F(nr)$ for $r < \frac{1}{n}$. It is

$$d_o = \frac{log\ p_0}{log(1/n)}. \tag{7.39}$$

7.3.4 Pointwise dimension at other points

By symmetry (thinking of $r=1$ as the "origin" on the other side), we have that the pointwise dimension at $r=1$ is

$$d_{r=1} = \frac{log\ p_{n-1}}{log\ (1/n)}. \tag{7.40}$$

In general, the pointwise dimension depends on the point at which it is taken. At

an arbitrary point, $r_o=0.s_1s_2s_3...$, we have that the pointwise dimension is defined by

$$d_{r_o} = \lim_{r\to 0} \frac{log\ \mu(B_{r_o}(r))}{log\ r}, \tag{7.41}$$

where $\mu(B_{r_o}(r))=F(r_o+r)-F(r_o-r)$ is the fraction of the Cantor set within r of the point r_o. By choosing intervals

$$\left[\frac{s_1}{n},\frac{s_1+1}{n}\right], \left[\frac{ns_1+s_2}{n^2},\frac{ns_1+s_2+1}{n^2}\right], \left[\frac{n^2s_1+ns_2+s_3}{n^3},\frac{n^2s_1+ns_2+s_3+1}{n^3}\right], \ldots \tag{7.42}$$

we have at the kth step an interval of length n^{-k} and of measure $p_{s_0}p_{s_1}p_{s_2}\cdots p_{s_k}$. It follows that the dimension is given by

$$d_{r_o} = \frac{\sum_s p'_s\, log\ p_s}{log(1/n)}, \tag{7.43}$$

where p'_s is the fraction of the s_j's in the base n expansion of r_o that are equal to s. For instance, at the point $r_o=0.sssssss...$, the dimension is $d = \frac{log\ p_s}{log(1/n)}$. If $p_s<\frac{1}{n}$, we get the nonintuitive result that the pointwise dimension can be greater than one, that is, ***greater*** than the dimension of the embedding space!

For a typical r_o (and all but a measure zero set of r_o are "typical"), we have $p'_s=p_s$ for all s and

$$d_{r_o} = \frac{\sum_s p_s\, log\ p_s}{log(1/n)}, \tag{7.44}$$

or $d=S/S_{max}$, where $S=-\sum_s p_s log(p_s)$ is the entropy of the partition $\mathbb{P}$, and

$S_{max}=log\ n$ is the entropy of the equidistributed partition $(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$. Of course, $S \leq S_{max}$, so the pointwise dimension at a typical point is never greater than the embedding dimension.

In §10.3, we speak of the "average pointwise dimension" of the attractor, and we show for this model that the "average" and the "typical" pointwise dimensions are the same.

7.3.5 Pointwise dimension on an attractor

We consider the pointwise dimension at the fixed point of the Hénon map [8],

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1 - ax_n^2 + y_n \\ bx_n \end{bmatrix}. \tag{7.45}$$

For the canonical values a=1.4, b=0.3, we have a saddle fixed point at $(x_o,y_o)=$ (0.6313, 0.1894). The linearization of the map about this point

$$\begin{bmatrix} x_{n+1}-x_o \\ y_{n+1}-y_o \end{bmatrix} \approx \begin{bmatrix} 2ax_o & 1 \\ b & 0 \end{bmatrix} \begin{bmatrix} x_n-x_o \\ y_n-y_o \end{bmatrix} \tag{7.46}$$

has eigenvalues 0.1559 and −1.9243. The stable ($|\lambda_<|<1$) eigenvalue defines the self-similarity ratio (the same role as 1/n in the Cantor model), and the unstable ($|\lambda_>|>1$) eigenvalue tells how fast points leave a neighborhood of the fixed point, so $1/|\lambda_>|$ corresponds to the probability, p_o, that an iterate stays in the neighborhood. The pointwise dimension of the "Cantor slice" in the fractal direction is therefore

$$d_{cantor} = \frac{log(1/|\lambda_>|)}{log\ |\lambda_<|}. \tag{7.47}$$

Perpendicular to this slice, the attractor is linelike, so we have that the pointwise dimension at (x_o,y_o) is

$$d = 1 + \left|\frac{log\ 1.9243}{log\ 0.1559}\right| = 1.338. \tag{7.48}$$

Figure 7.5 shows the self-similarity (and lacunarity) of the Hénon attractor at the fixed point and Figure 7.6 is a log-log plot of the fraction $F(r)=C_{x_o,y_o}(r)$ of points within r of the fixed point; here the oscillations are quite noticeable.

7.4 The correlation integral

We have seen how lacunarity can lead to oscillation of the "fraction" F(r); we now discuss how lacunarity can cause similar oscillations in the correlation integral.

7.4.1 Standard Cantor set

As with the fraction, we start by computing the correlation integral for the standard Cantor set. We are looking for the proportion of the distances between pairs of points that are less than r. Choosing $x_1 \varepsilon C$ and $x_2 \varepsilon C$, we measure the fraction of the set $C\times C$ for which $|x_1-x_2|\leq r$.

$$C(r) = \mu\ (\ \{(x_1,x_2)\varepsilon C\times C \text{ such that } |x_1-x_2|\leq r\}\ ) \tag{7.49}$$

By symmetry, we need only consider the "upper left half" of $C\times C$, where

$x_1 \leq x_2$. If we project this set along the $x_2 - x_1 = \text{constant}$ axis, we obtain a set whose fraction corresponds to the correlation integral C(r) of the Cantor set. Using the same tools we used to express F(r) recursively, we can analytically express the correlation integral for the Cantor set.

In Figure 7.7, each of the shaded triangles is similar to the full triangle, and each of unshaded triangles is empty. For $r \leq \frac{1}{3}$, we can write C(r), the fraction of the set below r,

$$C(r) = 2 \cdot \frac{1}{4} \cdot C(3r), \tag{7.50}$$

where "2" is the number of triangles, "$\frac{1}{4}$" is the mass of a small triangle relative to the full triangle, and "3" is the length of a large triangle relative to the small one. For $r \geq \frac{2}{3}$, we can write,

$$C(r) = \frac{3}{4} + 1 \cdot \frac{1}{4} \cdot C(3(r - \frac{2}{3})), \tag{7.51}$$

where $\frac{3}{4}$ is the mass of the set below $r = \frac{2}{3}$. For $\frac{1}{3} \leq r \leq \frac{2}{3}$, it's a bit tricky because the triangle is upside down, but it's not too hard to see that

$$C(r) = \frac{1}{2} + 1 \cdot \frac{1}{4} \cdot \left[1 - C(3(\frac{2}{3} - r))\right]. \tag{7.52}$$

Combining these into a single expression for C(r),

$$C(r) = \begin{cases} \frac{1}{2}C(3r) & \text{for } 0 \leq r \leq \frac{1}{3} \\ \frac{3}{4} - \frac{1}{4}C(2-3r) & \text{for } \frac{1}{3} \leq r \leq \frac{2}{3} \\ \frac{3}{4} + \frac{1}{4}C(3r-2) & \text{for } \frac{2}{3} \leq r \leq 1. \end{cases} \tag{7.53}$$

We make several observations. Because $C(r) = \frac{1}{2}C(3r)$ for $r \leq \frac{1}{3}$, we have *strict* self-similarity of the correlation integral C(r). Whatever lacunarity we see in C(r) for $\frac{1}{3} \leq r \leq 1$ occurs in precisely similar detail over every range $\frac{1}{3^k} \leq r \leq \frac{1}{3^{k-1}}$.

The correlation dimension is

$$\nu = \lim_{r \to 0} \frac{log\ C(r)}{log\ r} = \frac{log\ \frac{1}{2}C(3r)}{log\ \frac{1}{3} \cdot 3r} = \frac{log\ 2}{log\ 3}, \tag{7.54}$$

which is the same as the pointwise dimension at the origin that we calculated earlier. On the other hand, $\frac{dC}{dr} = 0$ for $r = \frac{1}{3^k}$, so we have in particular that the limit

$$\nu = \lim_{r \to 0} \frac{d[log\ C(r)]}{d[log\ r]} = \lim_{r \to 0} \frac{r}{C(r)} \frac{dC(r)}{dr} \tag{7.55}$$

does *not* exist. In Figure 7.8 we show a log-log plot of the correlation integral of the Cantor set.

7.4.2 Weighted Cantor set

Now we compute C(r) for a more general member of our class of Cantor sets. We can write the correlation integral in terms of the contributions from all the smaller triangles.

$$C(r) = \sum_{j=0}^{n-1} \sum_{k=j}^{n-1} \left[C_{\triangle jk}(r) \right] + \sum_{j=0}^{n-1} \sum_{k=j+1}^{n-1} \left[C_{\triangledown jk}(r) \right], \tag{7.56}$$

where

$$C_{\triangle jk}(r) = \begin{cases} 0 & \text{for } 0 \le r \le \frac{k-j}{n} \\ p_j p_k C(nr-(k-j)) & \text{for } \frac{k-j}{n} \le r \le \frac{k-j+1}{n} \\ p_j p_k & \text{for } \frac{k-j+1}{n} \le r \le 1 \end{cases} \tag{7.57}$$

and

$$C_{\triangledown jk}(r) = \begin{cases} 0 & \text{for } 0 \le r \le \frac{k-j-1}{n} \\ p_j p_k[1-C(k-j+1-nr)] & \text{for } \frac{k-j-1}{n} \le r \le \frac{k-j}{n} \\ p_j p_k & \text{for } \frac{k-j}{n} \le r \le 1. \end{cases} \tag{7.58}$$

Adding the contributions from all the triangles gives

$$\begin{aligned} C(r) = C(\tfrac{s}{n}) &+ (p_0 p_s + \cdots + p_{n-1-s} p_{n-1}) C(nr-s) \\ &+ (p_0 p_{s+1} + \cdots + p_{n-2-s} p_{n-1})[1-C(1+s-nr)] \end{aligned} \tag{7.59}$$

for $\frac{s}{n} \le r \le \frac{s+1}{n}$. This allows computation of C(r) exactly at rational points k/n^m in terms of values at k'/n^{m-1}. And although it is in principle possible to express C(r) as a series in terms of the base n expansion of r, as we did for F(r) above, it is an unwieldy series of limited value.

7.4.3 Strict and asymptotic self-similarity

What is noteworthy about this function is that it is *not* in general strictly self-similar for small r. There is no equation of the form C(r) = constant×C(nr) for $r \le \frac{1}{n}$. Instead, that relation is "contaminated" by a term with coefficient

$$p_0p_1 + p_1p_2 + \cdots + p_{n-2}p_{n-1}, \tag{7.60}$$

which vanishes only if every term p_kp_{k+1} vanishes. This occurs only when there are *no* adjacent nonempty segments.

In this case of nonadjacent nonempty segments, we can easily write down the correlation dimension:

$$\nu = \lim_{r\to 0} \frac{log\ C(r)}{log\ r} = \frac{log\left[\sum_s p_s^2\ C(nr)\right]}{log\ \frac{1}{n}\cdot nr} = \frac{log\left[\sum_s p_s^2\right]}{log\ \frac{1}{n}}\ . \tag{7.61}$$

We note that this agrees with the formula for generalized dimension [9] of a Cantor set.

$$D_q = \frac{1}{(q-1)}\cdot \frac{log\left[\sum_s p_s^q\right]}{log\ \frac{1}{n}}, \tag{7.62}$$

recalling that $q=2$ corresponds to the correlation dimension.

Equation (7.62) makes no assumptions about strictness of self-similarity, and in fact, we can show that self-similarity is not necessary for Equation (7.61). We begin with the statement that

$$C(1-1/n^k) = 1 - (p_0p_{n-1})^k, \tag{7.63}$$

which follows from Equation (7.59) with $s=n-1$. This enables us to write

$$C(1/n^{k+1}) = (p_0^2+\ldots+p_{n-1}^2)\cdot C(\frac{1}{n^k}) + (p_0p_1+\ldots+p_{n-2}p_{n-1})\cdot(p_0p_{n-1})^k. \tag{7.64}$$

Write

$$\alpha \equiv p_0^2 + \cdots + p_{n-1}^2 \tag{7.65}$$

$$\beta \equiv p_0 p_1 + \cdots + p_{n-2} p_{n-1} \tag{7.66}$$

$$\gamma \equiv p_0 p_{n-1}, \tag{7.67}$$

so

$$C(1/n^{k+1}) = \alpha C(1/n^k) + \beta\gamma^k \tag{7.68}$$

$$= \alpha^{k+1} + \alpha^k\beta + \alpha^{k-1}\beta\gamma + \alpha^{k-2}\beta\gamma^2 + \cdots + \beta\gamma^k \tag{7.69}$$

$$= \alpha^{k+1}\Big[\, 1 + \alpha^{-1}\beta\Big[\, 1 + \alpha^{-1}\gamma + \alpha^{-2}\gamma^2 + \cdots + \alpha^{-k}\gamma^k\Big]\Big]; \tag{7.70}$$

we remark that $p_0 p_{n-1} < p_0^2 + \cdots + p_{n-1}^2 \Rightarrow \alpha^{-1}\gamma < 1 \Rightarrow$ the series is bounded $\Rightarrow$

$$C(1/n^{k+1}) \approx \alpha^{k+1}\left[\, 1 + \frac{\alpha^{-1}\beta}{1-\alpha^{-1}\gamma}\right], \quad \text{for large } k. \tag{7.71}$$

Thus,

$$\nu = \lim_{r\to 0}\frac{log\ C(r)}{log\ r} = \frac{(k+1)\ log\ \alpha}{(k+1)\ log(1/n)} = \frac{log\ \alpha}{log(1/n)} = \frac{log\left[\sum_s p_s^2\right]}{log\ \frac{1}{n}}, \tag{7.72}$$

in agreement with the fully self-similar case.

7.4.4 Correlation integral on an attractor

The reader is referred to [10] in which the Zaslavskii [11] map

$$x_{n+1} = [\, x_n + \nu(1+\mu y_n) + \epsilon\nu\mu\cos(2\pi x_n)\,] \pmod 1,$$
$$y_{n+1} = e^{-\Gamma}(y_n + \epsilon\cos(2\pi x_n)), \tag{7.73}$$

with the parameters $\mu=(1-e^{-\Gamma})/\Gamma$, $\Gamma=3.0$, $\nu=400/3$, and $\epsilon=0.3$, is studied. The authors find a "kink" in their correlation integral for this map. A look at the

attractor itself explains the source of the kink. There is a very-long-period oscillation in the correlation integral (the "period" here has nothing to do with time, but is the ratio of length scales for which self-similarity is observed). There is an infinite train of kinks, but only the first kink is observed.

There is an obvious danger in trying to compute the dimension of such long-period attractors; the value can be skewed if slopes are taken over incomplete periods.

7.5 Conclusion

Periodic undulations in correlation integrals and fraction curves are common for weighted Cantor sets and are seen as well in generic dynamical attractors. These undulations demand that extra care be taken when estimating the dimension from the average slope of these curves.

In [1], it is observed numerically that the correlation integral C(r) is of the form $C(r)=r^d\psi(log\ r)$, where $\psi(x)$ is periodic function in x. What we have shown here is that ψ is *strictly* periodic for the standard Cantor set, and for any weighted Cantor set whose nonempty segments are not adjacent. Also, we have provided an implicit formula for C(r) from which ψ can be more directly computed.

7.6 Notes and References

[1] L. A. Smith, J.-D. Fournier, and E. A. Spiegel. "Lacunarity and intermittency in fluid turbulence," Physics Letters **114A** (1986) 465.

[2] R. Badii and A. Politi. "Intrinsic oscillations in measuring the fractal dimension," Phys. Lett. **104A** (1984) 303.

[3] Another analytic treatment of oscillations in the correlation integral is given by D. Bessis, J.-D. Fournier, G. Servizi, G. Turchetti, and S. Vaienti. "Mellin transforms of correlation integrals and generalized dimension of strange sets," Phys. Rev. **A 36** (1987) 920.

[4] M. F. Barnsley and S. Demko. "Iterated function systems and the global construction of fractals," Proc. Roy. Soc. London **A 399** (1985) 243.

[5] Stephen Demko, Laurie Hodges, and Bruce Naylor. "Construction of fractal objects with Iterated Function Systems," Computer Graphics **19** (1985) 271.

[6] As an aside, we note that we can use a two-dimensional deterministic map

$$y_{i+1} = \lambda y_i \pmod 1$$

$$x_{j+1} = \frac{1}{n} x_j + \mathcal{K}(y_j)$$

with $\lambda > 1$ and where $\mathcal{K}(y) = \frac{k}{n} \Leftrightarrow p_0 + p_1 + \ldots + p_{k-1} \le y < p_0 + p_1 + \ldots + p_k$. This is much like the generalized Baker's transform introduced by J. Doyne Farmer, Edward Ott, and James A. Yorke. "The dimension of chaotic attractors," Physica **7D** (1983) 153. In fact, the stochastic map may be viewed as the limiting case of this map as $\lambda \to \infty$.

[7] M. Hénon. "A two-dimensional mapping with a strange attractor," Comm. Math. Phys. **50** (1976) 69.

[8] H.G.E. Hentschel and Itamar Procaccia. "The infinite number of generalized dimensions of fractals and strange attractors," Physica **8D** (1983) 435. See in particular Equation (3.13) of this reference. For further discussion of generalized dimension in Cantor sets, see Thomas C. Halsey, Mogens H. Jensen, Leo P. Kadanoff, Itamar Procaccia, and Boris I. Shraiman. "Fractal measures and

their singularities: the characterization of strange sets," Phys. Rev. A 33 (1986) 1141.

[9] Peter Grassberger and Itamar Procaccia, "Measuring the strangeness of strange attractors," Physica 9D (1983) 189.

[10] G. M. Zaslavskii, Phys. Lett. 69A (1978) 145; quoted in Reference 9.

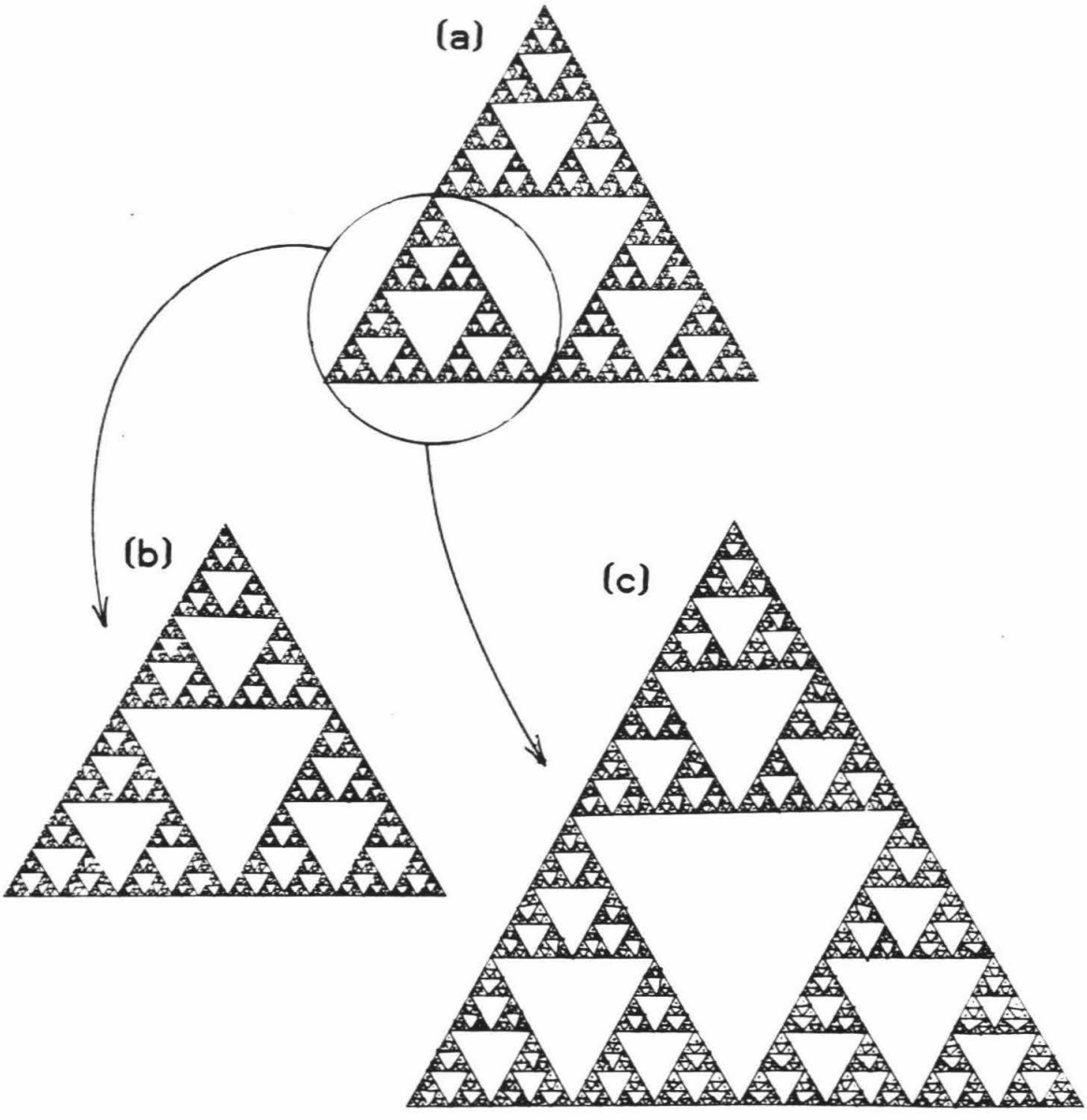


Figure 7.1 Discrete self-similarity of a fractal. (a) The Sierpinski gasket. (b) Magnification by a factor of two yields a figure that is congruent to the original. (c) Magnification by an arbitrary factor that is not a power of two yields a figure *not* congruent to the original.

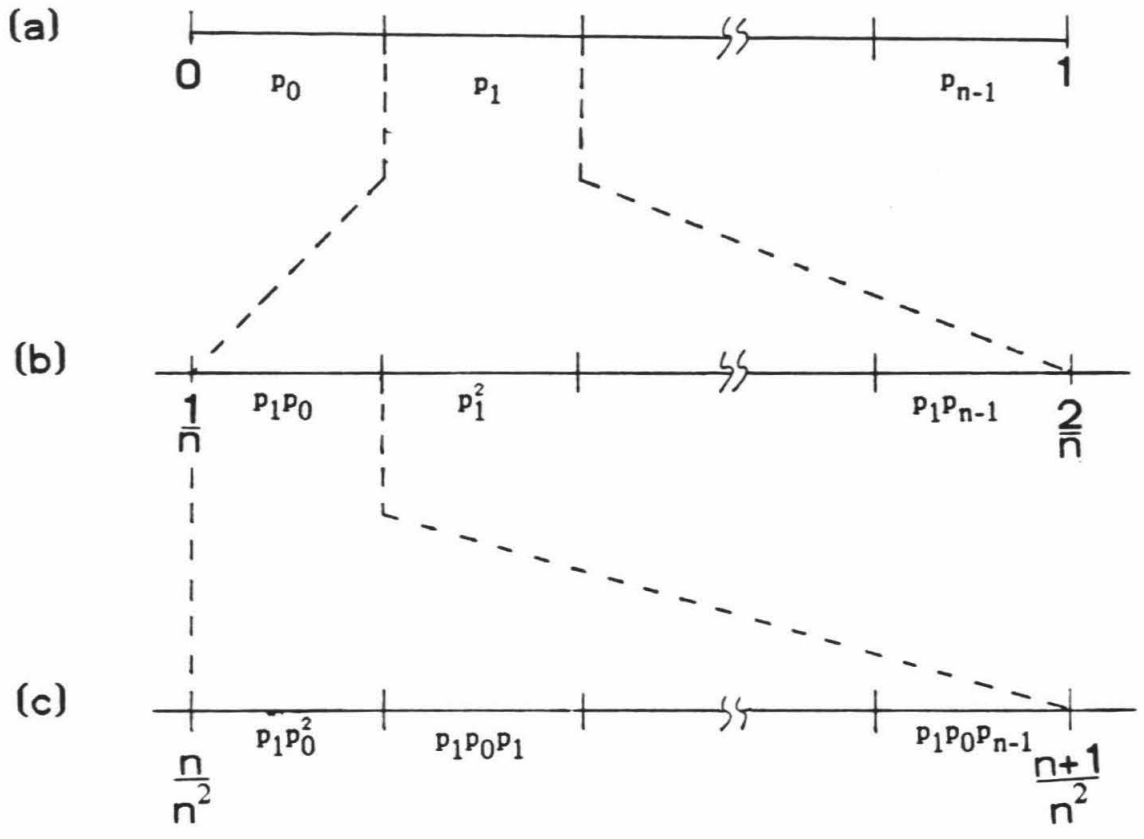


Figure 7.2 The weighted Cantor measure. (a) The interval [0,1] is divided into subintervals of size 1/n, and each is assigned a measure $\mu([k/n,(k+1)/n])=p_k$. We have $p_0+p_1+\cdots+p_{n-1}=1$, so $\mu([0,1])=1$. (b) The interval [1/n,2/n] is further divided into n subintervals, each of length $1/n^2$, and each assigned a measure p_1p_k. (c) The process continues until intervals of arbitrarily small length are assigned measures.

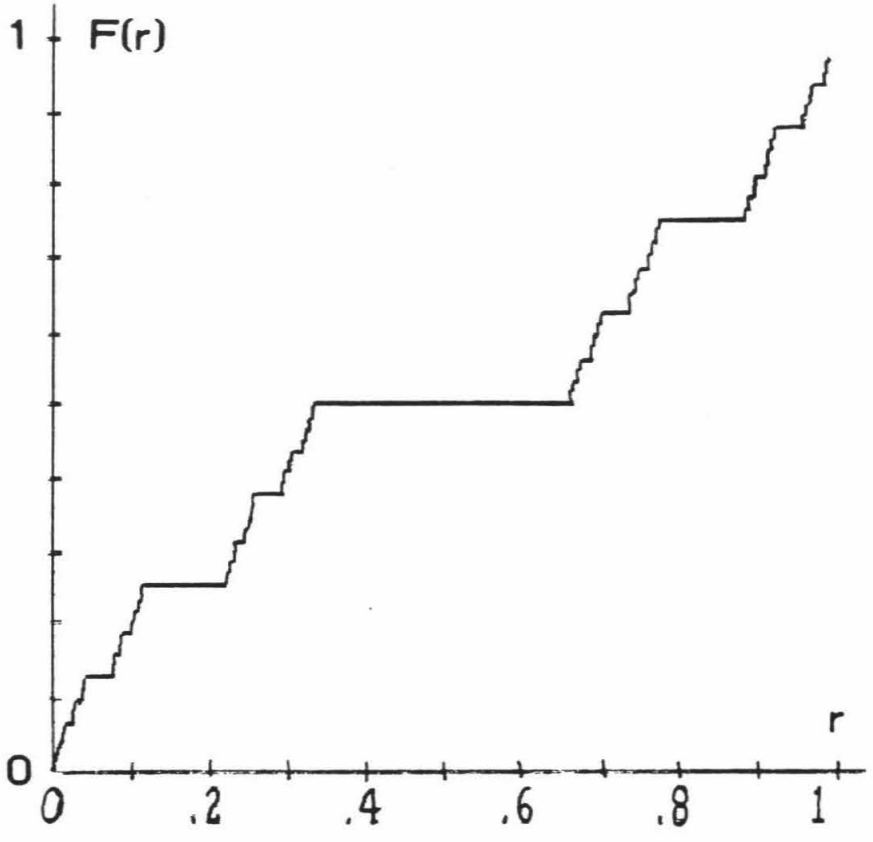


Figure 7.3 Devil's Staircase: the fraction F(r) for the standard Cantor set.

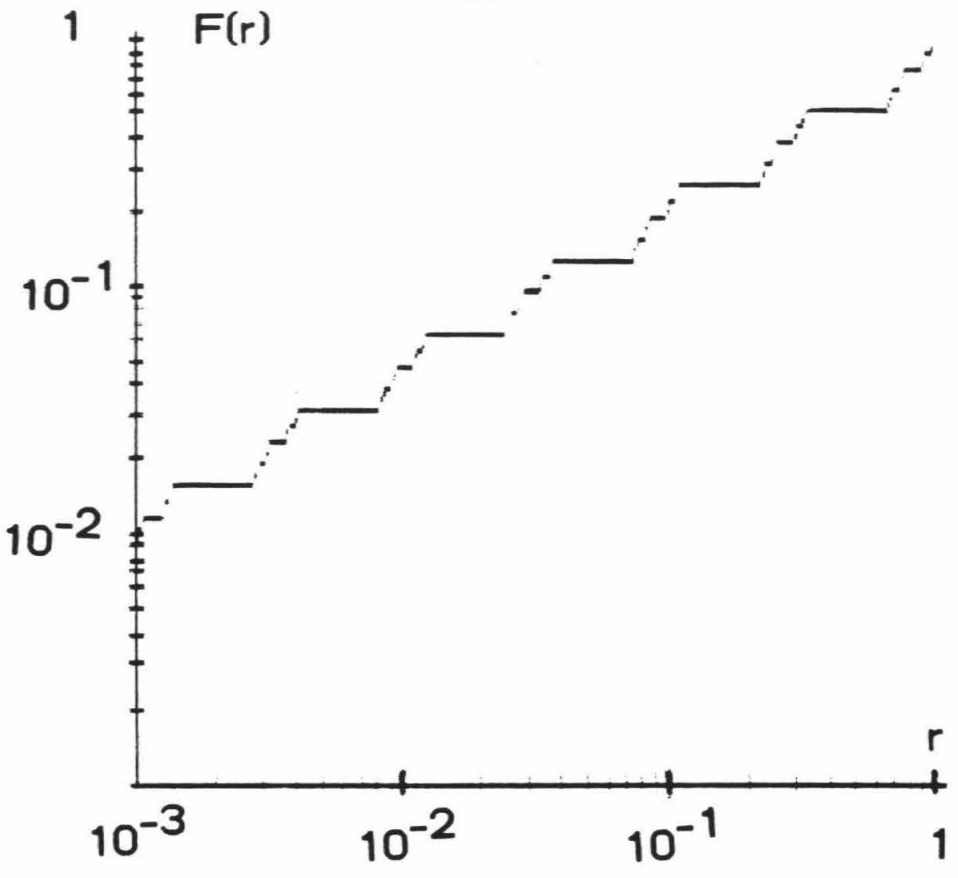


Figure 7.4 Periodic undulations in a log-log plot of the fraction F(r). Although the slope at almost every r is zero, the overall slope is $log\ 2/log\ 3 \approx 0.6309$.

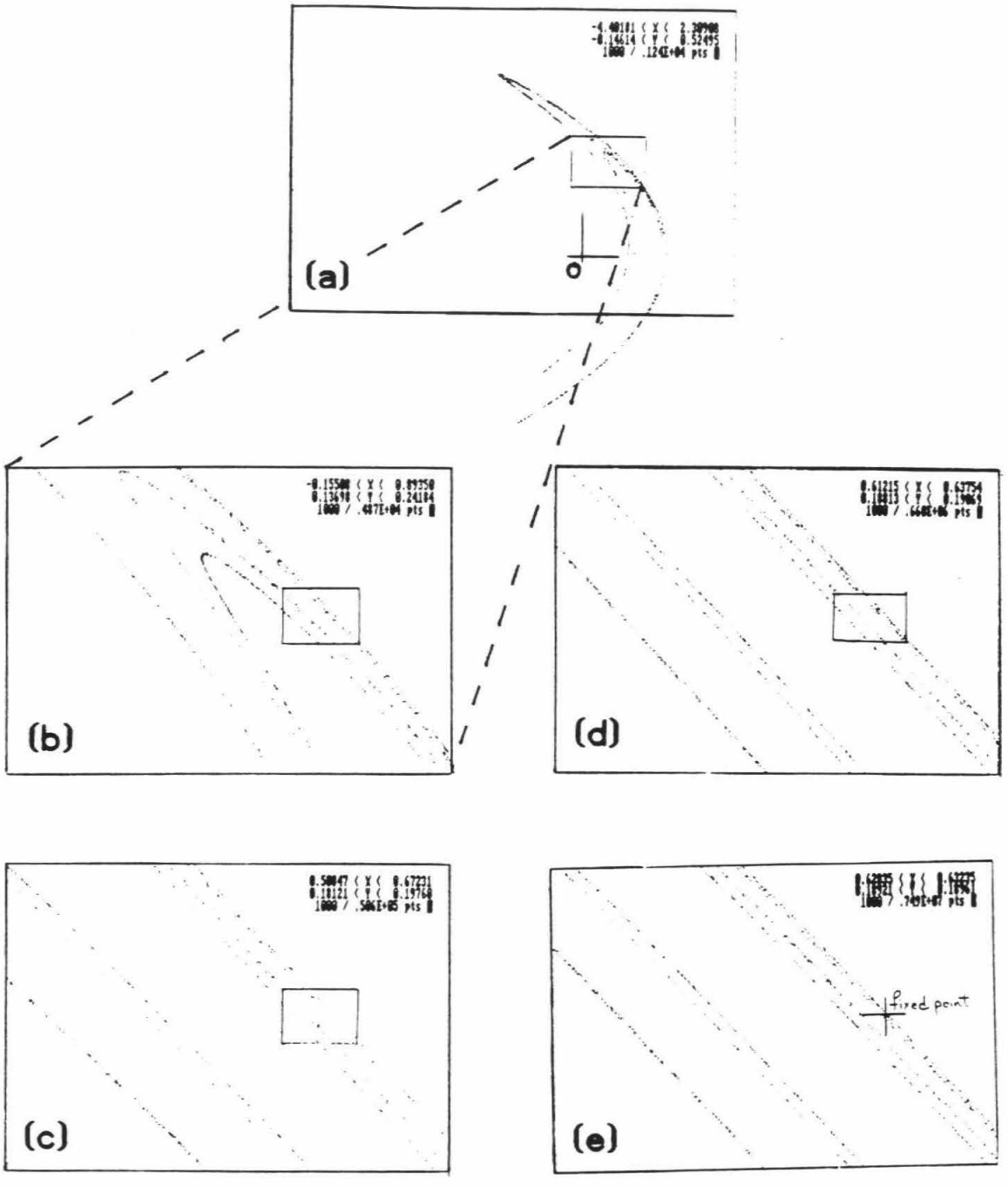


Figure 7.5 Self-similarity of Hénon attractor at fixed point (x_o,y_o)=(0.6313, 0.1894). Figures (a) (b) (c) (d) and (e) represent successive magnifications by $1/\lambda_< = 6.414$.

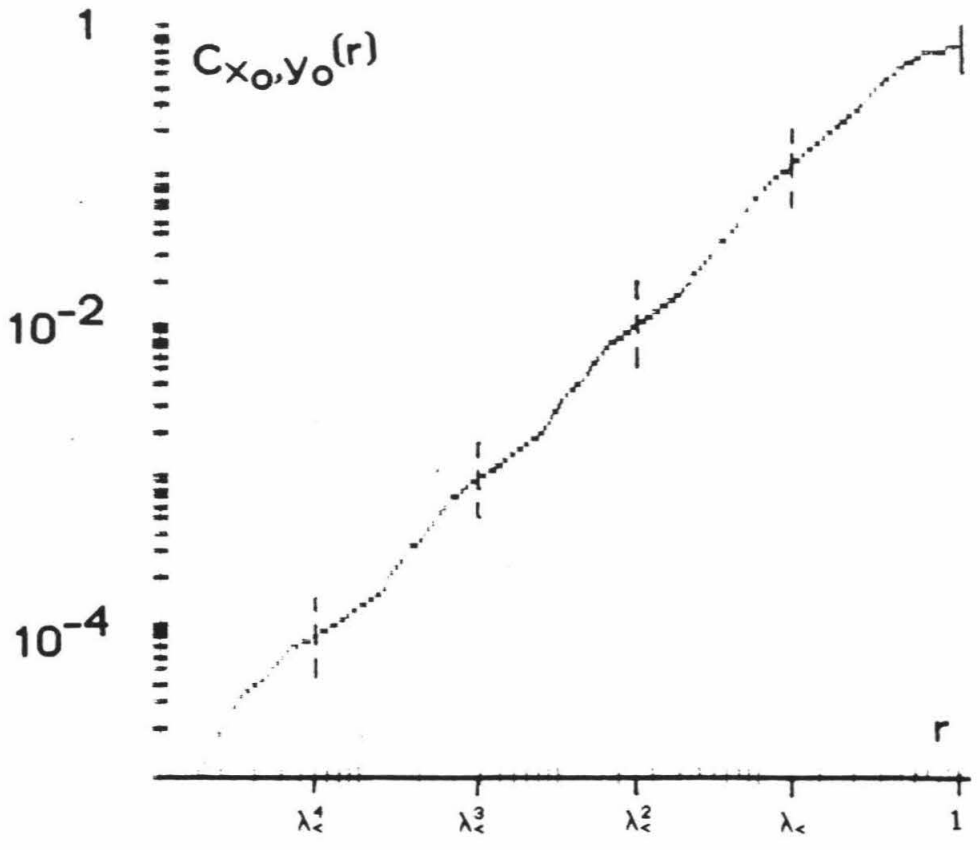


Figure 7.6 *log* $C_{x_o,y_o}(r)$ **versus** *log* **r for Hénon attractor at fixed point. Here, we see oscillations with a period of** $\lambda_<=0.1559$.

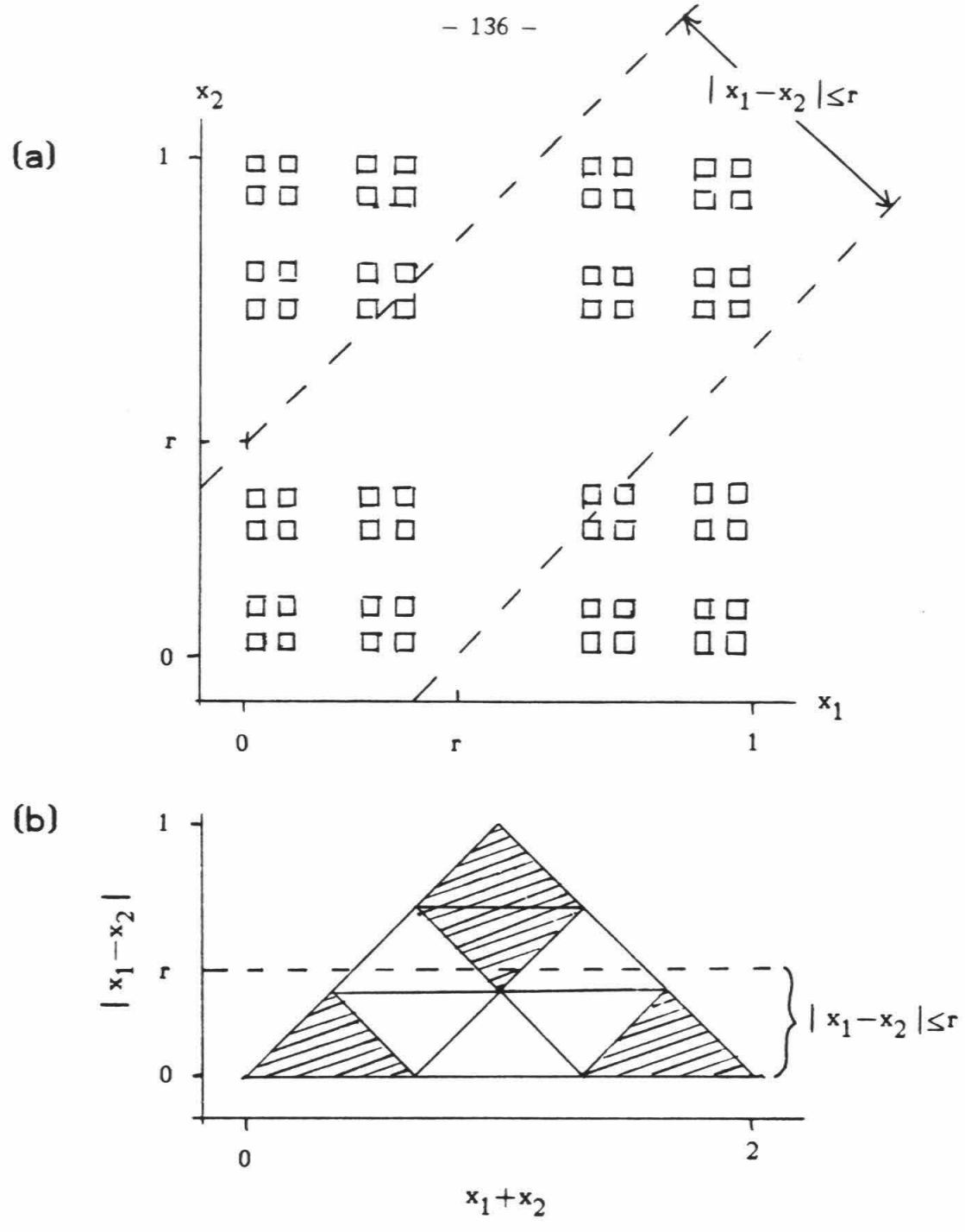
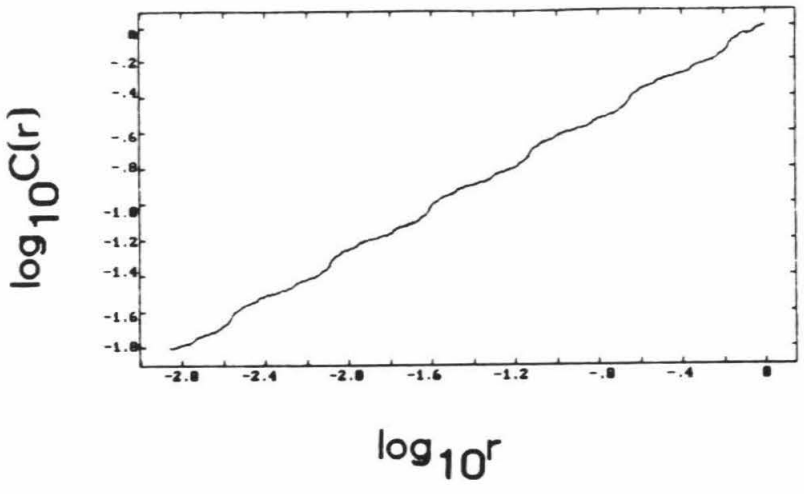


Figure 7.7 (a) The set C×C, where C is the standard Cantor set. (b) C×C on rotated axes. Each shaded triangle is similar to the large triangle. The correlation integral C(r) measures the fraction of the set below r.

(a)



(b)

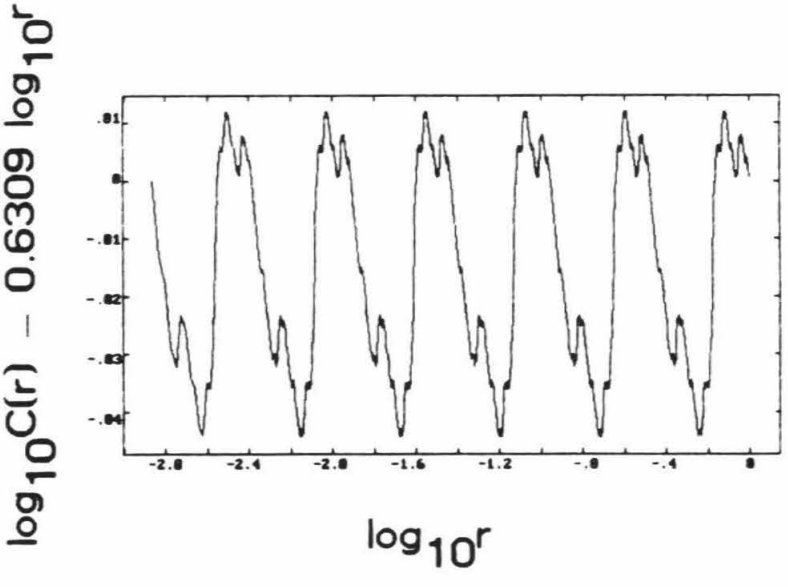


Figure 7.8 Oscillations in the correlation integral for standard Cantor set. (a) *log* C(r) versus *log* r; note the average slope is *log* 2/*log* 3 ≈ 0.6309. (b) Difference plot: [*log* C(r)—0.6309 *log* r] versus *log* r. Here, the strict periodicity in the correlation integral is evident.

CHAPTER EIGHT

8. NONCHAOTIC ATTRACTORS: QUASIPERIODICITY

Though dimension is a geometric property of an attractor, we find that the accuracy and efficiency of dimension estimates from a finite sample of N points depend as well on the dynamics of the system. In this and the next chapter, we are concerned not only with the attractor's natural invariant measure, but with the dynamical process by which points are placed on the attractor.

Deterministic dynamical systems can be classified into two categories, chaotic and nonchaotic. (As an aside, we remark that these two categories are *not at all* related to the classification into conservative and nonconservative dynamical systems.) Chaotic systems are characterized by their sensitive dependence on initial conditions.

We argue that the *best* way to populate an attractor $\mathcal{A}$ with a finite set of points is to drop the points randomly and independently over $\mathcal{A}$ with a weight given by the invariant measure. Although a dynamical system places points on the attractor in a manner that is strictly deterministic, we find that chaotic dynamical systems are *qualitatively better* at populating their attractors than nonchaotic systems.

In Chapter Four, we show that points dropped randomly and independently over an attractor lead to a scaling range in C(N,r) of order N^2. That is, $C(N,r) \propto r^{\nu}$ over a range in which C(N,r) varies from $2/N^2$ to 1. Experience with chaotic attractors leads us to expect this same $O(N^2)$ scaling range in C(N,r). For

nonchaotic attractors, on the other hand, we find an O(N) scaling range.

This chapter discusses nonchaotic attractors, particularly those with periodic or quasiperiodic motion. (Quasiperiodic motion can be characterized by functions of angular variables

$$X(t) = Q(\theta_1,\theta_2,\ldots), \tag{8.1}$$

where

$$\theta_i = \omega_i t \pmod{2\pi}, \tag{8.2}$$

and the frequencies $\omega_1,\omega_2,\ldots$ are incommensurate. Since periodic motion is trivially quasiperiodic, we will henceforth use the term "quasiperiodic" to refer to both kinds of motion.) Limit cycles and limit tori are the usual sources of quasiperiodic motion; we will consider a simplified limit cycle defined by a one-dimensional twist map with a single winding number ϕ. Before analyzing the twist map model, however, we will do two "control experiments," in which the attractor is populated according to very specific strategies that caricature the distribution of points as laid down by nonchaotic and by chaotic dynamical systems.

In the next chapter, we will address an effect that occurs with systems that are chaotic but that have time autocorrelations.

8.1 Wraparound metric

We again (see §4.4.1 and §6.1) introduce as an embedding space the unit interval with the wraparound metric. The distance between two points $x,y \varepsilon [0,1)$

is given by

$$d(x,y) = \min(|x-y|, 1-|x-y|). \tag{8.3}$$

This is equivalent to the arcwise distance between two points on a circle of unit circumference. The advantage of this metric is that it eliminates edge and curvature effects (see §6.1 and §6.3).

8.2 Two control experiments

To illustrate the two qualitatively different behaviors that the C(N,r) curve can display, we perform two experiments in which we populate the interval [0,1) in the wraparound metric with points according to two strategies; the first models the essential features of a quasiperiodic system, and the second models a highly chaotic system.

The first control experiment is to populate the circle uniformly, with N points, dropped so that every point is situated at the midpoint between its nearest neighbors on either side of it. This very organized distribution simulates the kind of behavior that would be expected from a nonchaotic system. In other words, if the first point defines $x_0=0$, then $x_1=1/N$, ..., $x_k=k/N$, etc. In this case, we see that the shortest distance is just $r_{min}=1/N$. In fact, there will be N of these distances, so $C(N,r=1/N) = \frac{2}{N^2}N = 2/N$. Further, all distances will be multiples of 1/N, and it is not too hard to see that $C(N,r=k/N) = 2k/N$. The correlation integral is plotted on a log-log graph in Figure 8.1. Though the curve jumps in big steps, the overall slope is one — as expected, our points "sample" a one-dimensional space. The range over which this slope of one

persists is O(N). This seems a bit wasteful, since the algorithm computed $O(N^2)$ distances. We see that this is a worst case example, however, and that for any sampling of N points, we will always have $r_{min} \leq 1/N$.

The second control experiment is to populate the circle randomly with N points, all dropped independently of each other. This is not unlike how a highly chaotic system would distribute points. In this case (see §4.2.2 and §4.3.2), we find that the nearest pair of points is $r_{min}=O(1/N^2)$. A typical correlation integral is plotted in Figure 8.2. Here we get the full $O(N^2)$ range from our correlation integral; this is twice as many orders of magnitude as the O(N) range observed for the more organized distribution of points. This enables a much more accurate estimate of the slope, which again is one.

8.3 Twist map

Now, to model the effect of quasiperiodicity, we consider a simple twist map

$$x_{n+1} = x_n + \phi \quad \text{(modulo 1)}. \tag{8.4}$$

Though this particular model is a conservative system, that is not an essential feature; we could have designed a two-dimensional model [1] for which this would be the "limit cycle."

8.3.1 Rational ϕ

If ϕ is rational, then the trajectory is periodic with period $\mathcal{D}$, where $\phi=\frac{\mathcal{N}}{\mathcal{D}}$

is the fraction expressed in reduced form. The motion in this case is not ergodic; a single trajectory is not dense on the attractor. The attractor will appear as a finite set of points, and will be seen to have dimension zero.

8.3.2 Irrational ϕ: the golden mean

For ϕ irrational, the trajectory is aperiodic and dense on the circle. Consider an orbit of length N, $\{x_0,x_1,\ldots,x_{N-1}\} \subset [0,1)$. We will show that the smallest distance r_{min} varies with N in a way that is reminiscent of the uniform control experiment above; that is $r_{min}=O(1/N)$. It is for this reason that the correlation integral performs poorly for quasiperiodic systems.

Suppose that the smallest distance r_{min} is between the two points x_a and x_b, with $a<b$ for definiteness. Because of the nature of the twist map, the distance between x_a and x_b is the same as the distance between x_0 and x_{b-a}. (Indeed, this implies that there are at most N distinct distances, our first hint that the smallest of them is only of order 1/N.) Taking k=b-a, we can write

$$x_k = x_0 + k\phi \quad \text{(modulo 1)}, \tag{8.5}$$

or

$$x_k - x_0 = k\phi \quad \text{(modulo 1)}, \tag{8.6}$$

which implies

$$|x_k - x_0| = |k\phi - [k\phi]|, \tag{8.7}$$

where $[k\phi]$ is the nearest integer to $k\phi$. Thus,

$$r_{min} = \min_{1\leq k<N} |k\phi - [k\phi]|. \tag{8.8}$$

We introduce

$$\epsilon_k(\phi) = \left| \phi - \frac{[k\phi]}{k} \right| \tag{8.9}$$

as the error that a fraction with denominator k makes in estimating the irrational ϕ. In this notation,

$$r_{min} = \min_{1 \leq k < N} k \cdot \epsilon_k(\phi), \tag{8.10}$$

and the problem of finding the minimum distance between pairs of points becomes a matter of rationally approximating ϕ. Fortunately, there is ample theory for this [2].

Let $[a_0,a_1,a_2,\ldots]$ denote the continued fraction expansion for ϕ. That is,

$$\phi = [a_0,a_1,a_2,\ldots] \equiv a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}} \tag{8.11}$$

It is truncations of the continued fraction expansion that provide the "best" rational approximations of ϕ, and that allow us to estimate the behavior of $\epsilon_k(\phi)$. We have from the theory of continued fractions the following inequality:

$$[a_0] < [a_0,a_1,a_2] < \ldots < [a_0,\ldots,a_{2n}] < \phi < [a_0,\ldots,a_{2n+1}] < \ldots < [a_0,a_1], \tag{8.12}$$

which enables us to put bounds on $\epsilon_k(\phi)$. We will in particular consider the case

where $\phi=[0,1,1,1,1,\ldots]=(\sqrt{5}-1)/2$ is the golden mean.

TABLE 8.1: Rational approximations to $\phi=\frac{\sqrt{5}-1}{2}$

ϕ		$[k\phi]/k$	ϵ_k	$k\cdot\epsilon_k$	$k^2\cdot\epsilon_k$
0.61803	1/2	0.5	0.11803	0.23607	0.47214
0.61803	2/3	0.66667	0.04863	0.14590	0.43769
0.61803	2/4	0.5	0.11803	0.47214	1.88854
0.61803	3/5	0.6	0.01803	0.09017	0.45085
0.61803	4/6	0.66667	0.04863	0.29180	1.75078
0.61803	4/7	0.57143	0.04661	0.32624	2.28367
0.61803	5/8	0.625	0.00697	0.05573	0.44582
0.61803	6/9	0.66667	0.04863	0.43769	3.93925
0.61803	6/10	0.6	0.01803	0.18034	1.80340
0.61803	7/11	0.63636	0.01833	0.20163	2.21789
0.61803	7/12	0.58333	0.03470	0.41641	4.99689
0.61803	8/13	0.61538	0.00265	0.03444	0.44774

We see that the best approximations are given by the truncated continued fractions. In particular, we find that $k\cdot\epsilon_k(\phi)$ is smallest when k is a truncated continued fraction denominator. In fact, the k for which the minimum holds in Equation (8.10) is just the largest truncated denominator less than N. This largest denominator is usually of order N; further, $k^2\cdot\epsilon_k(\phi) \rightarrow$ constant ($=1/\sqrt{5}\approx 0.4472$ when ϕ is the golden mean), so we expect

$$r_{min} = \min_{1\le k<N} k\epsilon_k(\phi) \tag{8.13}$$

$$= \tilde{k}\epsilon_{\tilde{k}}(\phi), \tag{8.14}$$

where $\tilde{k}$ is the value of k that minimizes $k\epsilon_k(\phi)$,

$$= \frac{\tilde{k}^2 \epsilon_{\tilde{k}}(\phi)}{\tilde{k}}. \tag{8.15}$$

In particular, for $\phi = \frac{\sqrt{5}-1}{2}$, $\tilde{k}$ is the largest Fibonacci number less than N. Thus, $\phi N \leq \tilde{k} < N$, so we can bound

$$\frac{\tilde{k}^2 \epsilon_{\tilde{k}}(\phi)}{N} \leq r_{min} \leq \frac{\tilde{k}^2 \epsilon_{\tilde{k}}(\phi)}{\phi N} \tag{8.16}$$

or

$$\frac{0.447}{N} \leq r_{min} \leq \frac{0.724}{N}. \tag{8.17}$$

This is to be contrasted with $r_{min} = O(1/N^2)$ that is expected for chaotic data.

Now, we can estimate the value of the correlation integral C(N,r) for $r = r_{min}$. If the minimum in Equation (8.13) is achieved for $k = \tilde{k}$, then the number of distances for which $r = r_{min}$ is just $N - \tilde{k}$. Thus,

$$C(N, r_{min}) = \frac{2}{N^2}(N - \tilde{k}). \tag{8.18}$$

Again, $\phi N \leq \tilde{k}_{min} < N$, so

$$\frac{2}{N^2} < C(N, r_{min}) < \frac{0.76}{N}. \tag{8.19}$$

For any value of N, $C(N, r_{min}) < 2r_{min}$. In particular, if we take a "typical" N, we expect $\tilde{k} \approx \sqrt{\phi} \cdot N$; and from this, we get

$$r_{min} \approx .56/N, \tag{8.20}$$

$$C(N,r_{min}) \approx 0.43/N. \tag{8.21}$$

In the model above we expect $C(N,r) = 2r$, at least for large r. In the example of perfectly uniform distribution of N points on the circle, we find $r_{min}=1/N$, but $C(N,r_{min})=(2/N^2)\cdot N=2/N=2r_{min}$. Thus, we have $C(N,r)=2r \propto r$ for the full range $r_{min} \leq r \leq \frac{1}{2}$. But for the case of quasiperiodic data with frequency ratio $\phi=(\sqrt{5}-1)/2$, we find that although $C(N,r)=2r$ for large r, $C(N,r_{min})<2r_{min}$. This leads to a correlation integral such as is shown in Figure 8.3.

For two reasons the correlation integral is not well suited for finding the dimension of quasiperiodic systems. The first is a matter of fundamental inefficiency: the range available is only O(N), even though there are $O(N^2)$ distances to compute. The second reason is that the curve is not even expected to show a constant slope over the range of distances that are available. The tail at the low end of the curve leads to systematically high estimates of dimension.

In principle, this analysis based on continued fractions can be applied to any irrational; the results we have quoted however, in Equations (8.17, 8.19, 8.20, 8.21) for instance, are specific to $\phi=(\sqrt{5}-1)/2$, and a new calculation would have to be done for each ϕ. Instead, we will consider the generic ϕ as a separate case, taking ϕ as a random variable uniformly distributed over [0,1).

8.3.3 Generic ϕ

We have shown that for the highly incommensurate $\phi=[0,1,1,1,\ldots]=$

$(\sqrt{5}-1)/2$, we have for quasiperiodic data a systematic effect that tends to overestimate the dimension. At the other extreme, we might consider a highly commensurate frequency ratio, one that is rational or very nearly rational. We would find in this case that r_{min} is very small (it is zero for rational ϕ if number of points N is greater than smallest denominator of ϕ), and that dimension could well be underestimated. A question to ask is: What about "typical" frequency ratios ϕ; is dimension on the average correctly estimated?

We can model the concept of a "typical" frequency ratio ϕ, by treating ϕ as a random variable that is uniformly distributed over the interval [0,1). For fixed N, we can compute exactly the expected value for r_{min}.

$$\langle r_{min} \rangle = \left\langle \min_{1 \le k < N} \left[k \cdot \epsilon_k(\phi) \right] \right\rangle = \int_0^1 \min_{1 \le k < N} \left[k \cdot \epsilon_k(\phi) \right] d\phi \tag{8.22}$$

The integrand is sketched in Figure 8.4. By numerical computation, we find for the large N behavior

$$\langle r_{min} \rangle \approx \frac{0.4215 \pm 0.0005}{N} \tag{8.23}$$

for $N \gtrsim 1000$, and where the "$\pm$" denotes the variation of $\langle r_{min} \rangle$ over N.

We see that the O(1/N) behavior of r_{min} is generic in ϕ. This compares with the $O(1/N^2)$ behavior of r_{min} that is exhibited by stochastic systems and by chaotic systems. The sample of points generated by a quasiperiodic system does not provide an efficient representation of the underlying attractor.

In Figure 8.5, we show pictures of the correlation integral for several

randomly chosen ϕ. We find that for "very irrational" ϕ, the shape of the correlation integral looks like that of the golden mean ϕ, fairly smooth but with a slope that is too steep for small r near r_{min}. For "nearly rational" ϕ, on the other hand, we see a "stair-steppy" shape from which a reasonable slope would be difficult to obtain.

We see that if ϕ is "very" irrational, dimension tends to be slightly overestimated, and if ϕ is "nearly" rational, dimension is likely to be badly estimated. In a numerical experiment, we considered a large sample of $\phi \varepsilon [0,1)$, and computed dimension of points on the unit interval [0,1] with the usual (not the wraparound) metric. The analysis was done automatically according to the Takens maximum likelihood formula (see §10.2), disregarding the shape of the correlation integral. With quasiperiodic data, we found an average dimension of $d = 1.075 \pm 0.02$, whereas for random data we obtained an average of $d = 0.971 \pm 0.006$ [3]. The five sigma effect suggests that on the average, quasiperiodic data can be expected to overestimate dimension.

There is some ambiguity in the physical meaning of the term "generic." For instance, the golden mean is very often the winding number in experiments at critical transitions [4]. It may be that the ϕ's that come up in typical *nonlinear* dynamical systems in a quasiperiodic regime do not spread themselves uniformly over the interval [0,1). This is a deep issue, beyond the scope of the work presented here.

As an example, though, the reader is referred to [5] in which the correlation integral is plotted for points from the logistic map $x_n = \lambda x_{n-1}(1 - x_{n-1})$ at $\lambda = 3.5699...$, the limit point of the period doubling bifurcation. Here, although

the system is by no means periodic, it has zero entropy (see §13.4.4), and we find it interesting that the correlation integral should look so similar to Figure 8.3 [6].

8.4 Notes and References

[1] For example: $x_{n+1}=x_n+y_n+\phi$; $y_{n+1}=0.5y_n$. Here $y_n \to 0$ as n increases, and the limit cycle is given by $x_{n+1}=x_n+\phi$; $y_n=0$.

[2] As well as the continued fractions that are discussed in the text, an interesting alternative scheme for rational approximation is given by Seung-hwan Kim and Stellan Ostlund. "Simultaneous rational approximations in the study of dynamical systems," Phys. Rev. A 34 (1986) 3426.

[3] The underestimate from the random data is caused by the "edge" effect (see §6.2.1); in the wraparound metric there is no edge effect (see §6.1), but for the sake of the numerical experiment, the correlation integral was based on the conventional distance.

[4] Mogens H. Jensen, Leo P. Kadanoff, Albert Libchaber, Itamar Procaccia, and Joel Stavans. "Global universality at the onset of chaos: results of a forced Rayleigh-Bénard experiment," Phys. Rev. Lett. 55 (1985) 2798; E. G. Gwinn and R. M. Westervelt. "Scaling structure of attractors at the transition from quasiperiodicity to chaos in electronic transport in Ge," Phys. Rev. Lett. 59 (1987) 157.

[5] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors," Physica 9D (1983) 189.

[6] See Figure 2 of Reference 5.

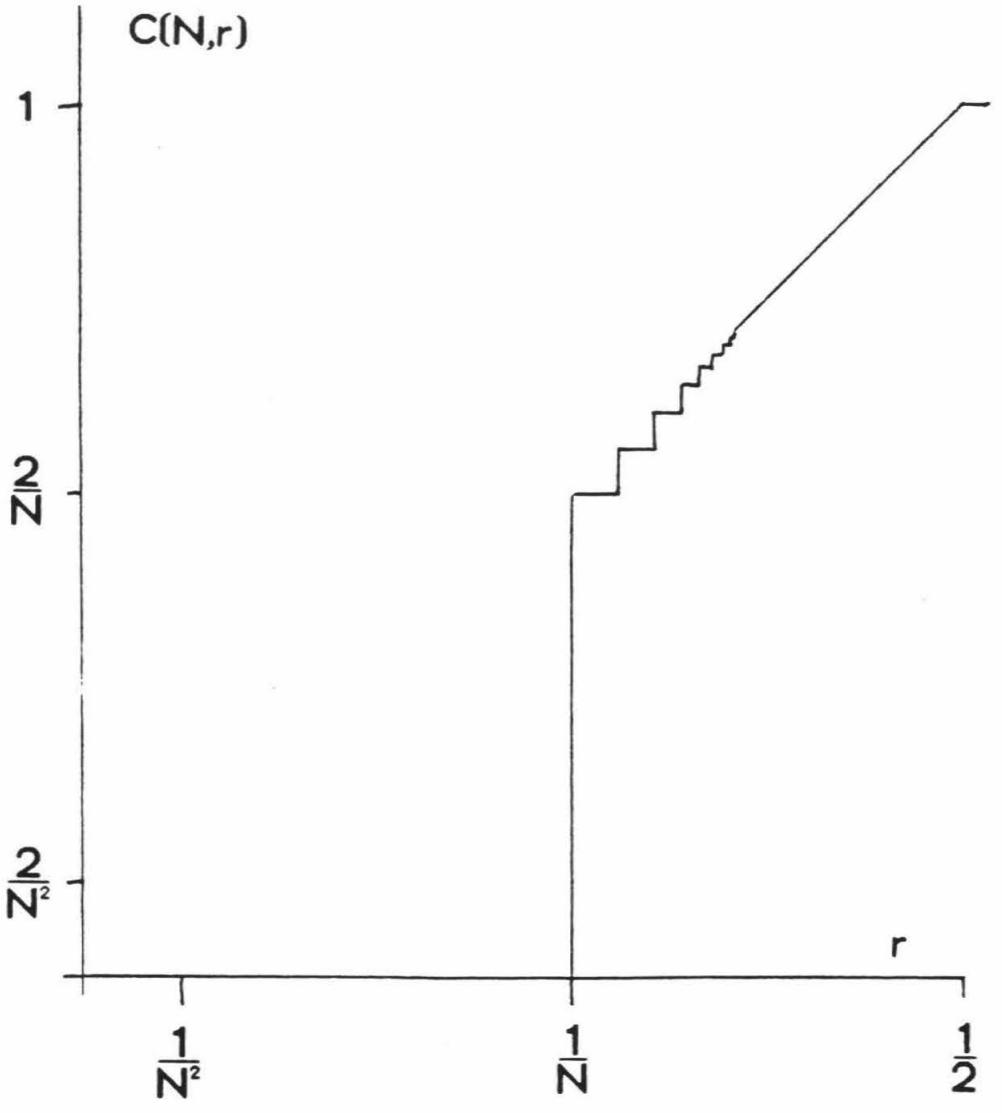


Figure 8.1 The correlation integral C(N,r) versus r on logarithmically scaled axes for N points equally spaced over [0,1) in the wraparound metric. Note that the useful range in C(N,r) is only O(N).

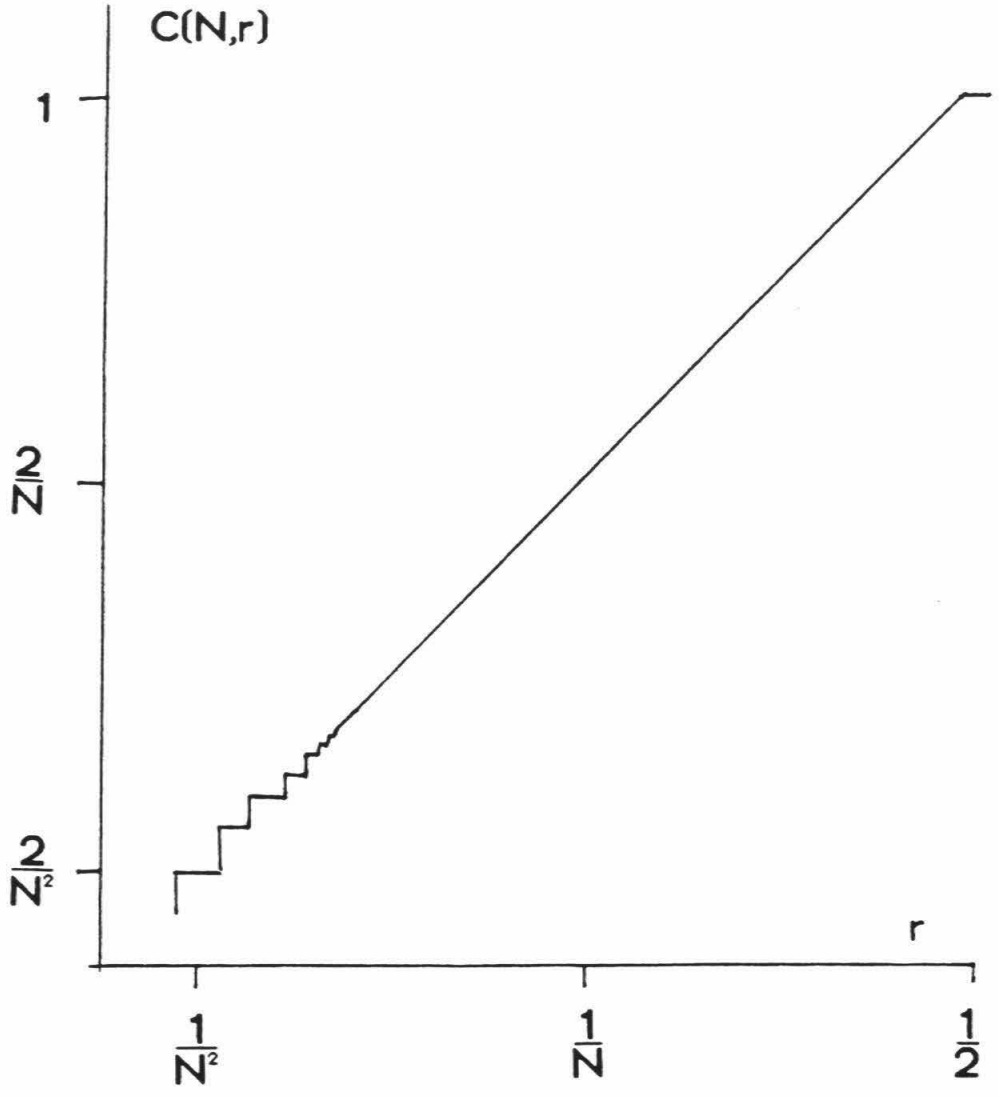


Figure 8.2 The correlation integral C(N,r) versus r on logarithmically scaled axes for N points randomly distributed over [0,1) in the wraparound metric. Here, the useful range in C(N,r) is $O(N^2)$.

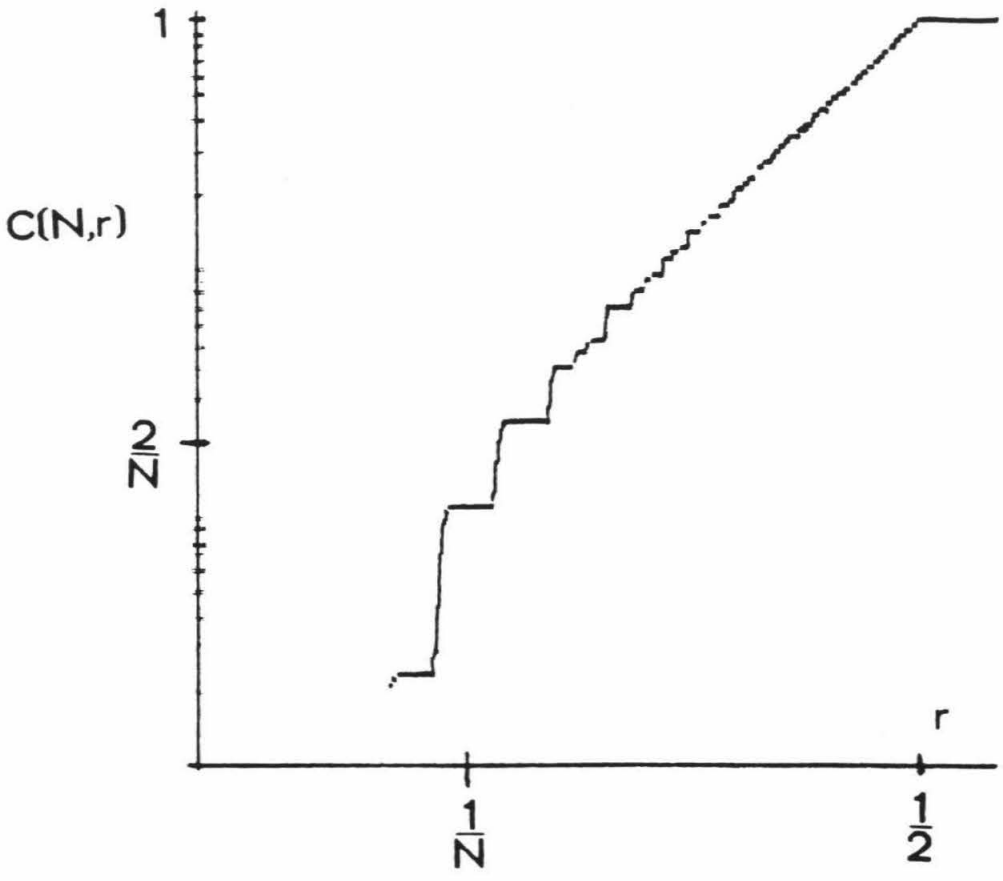


Figure 8.3 The correlation integral C(N,r) versus r on logarithmically scaled axes for N=100 points quasiperiodically generated according to winding number $\phi = 0.618034\cdots = (\sqrt{5}-1)/2 = [0,1,1,1,\ldots]$. Again, note that the range over which C(N,r) varies is only O(N).

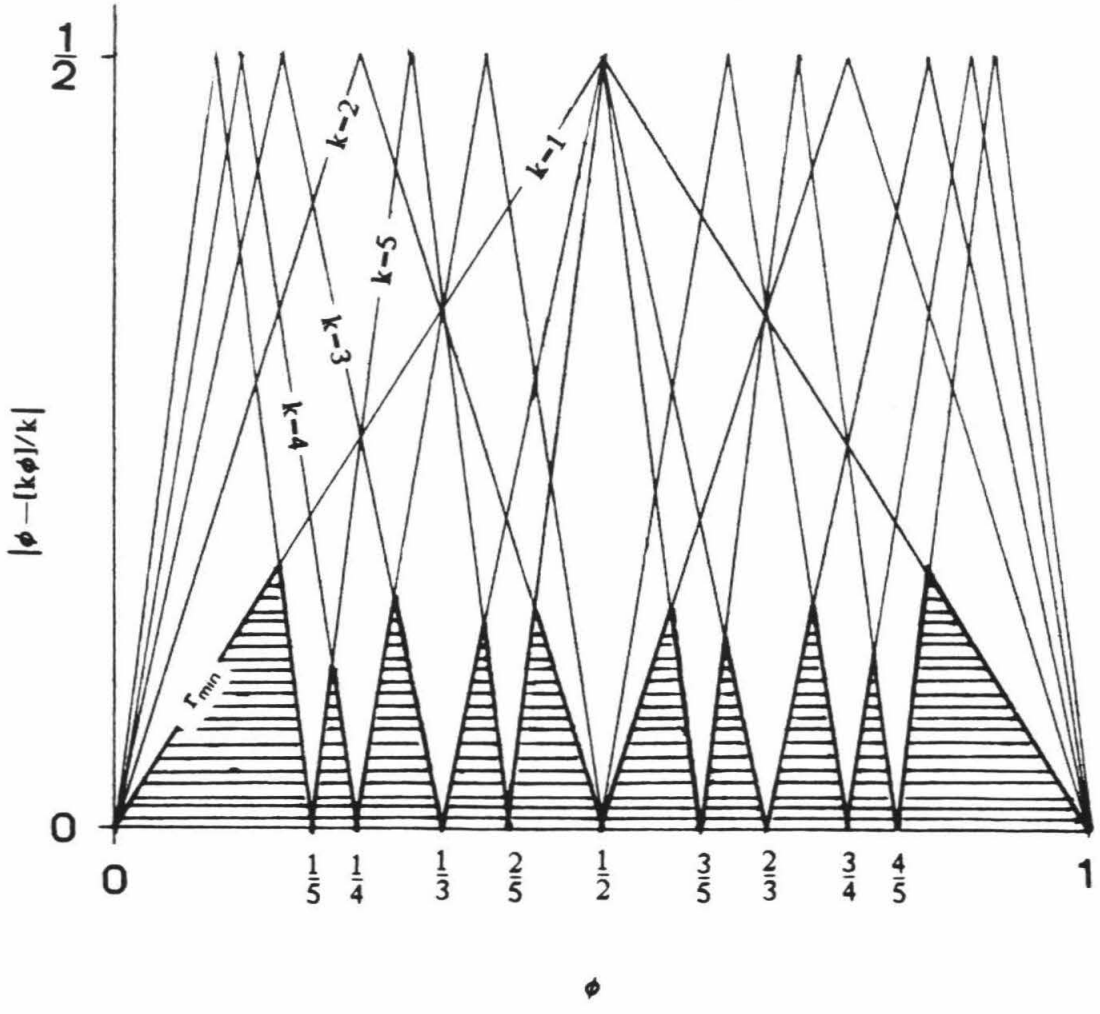


Figure 8.4 $|\phi-[k\phi]/k|$, where $[k\phi]$ is the integer nearest to $k\phi$, plotted as a function of ϕ for k=1,2,3,4,5. $r_{min}=\min_{1\leq k\leq N}|\phi-[k\phi]/k|$ is plotted as a function of ϕ for N=5. The average value $\langle r_{min}\rangle$ is the area of the shaded region.

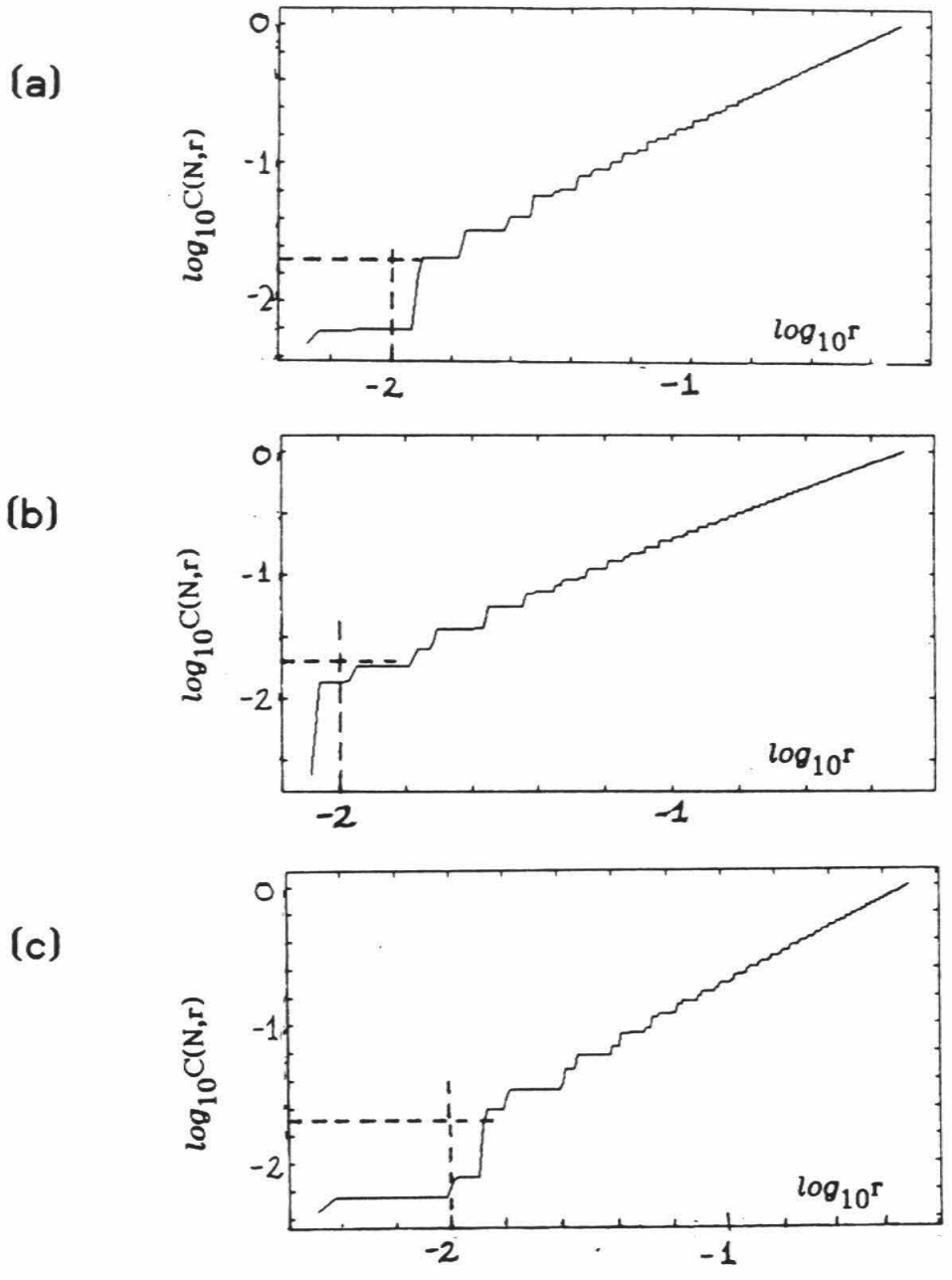


Figure 8.5 $log_{10}C(N,r)$ versus $log_{10}r$ for N=100 points quasiperiodically generated according to various winding numbers ϕ. In each case, the dotted lines correspond to r=1/N, C(N,r)=2/N. (a) ϕ=0.41421356 = [0,2,2,2,...]. (b) ϕ=0.30277564 = [0,3,3,3,...]. (c) ϕ=0.23606798 = [0,4,4,4,...].

CHAPTER NINE

9. AUTOCORRELATION

Autocorrelation is ubiquitous in time series data. For continuous signals x(t), there is always some time τ over which x(t) and x(t+τ) are strongly correlated. In this chapter, we point out an effect due to autocorrelation, an anomalous "shoulder" in the correlation integral, which can lead to inaccurate and possibly spurious estimates of dimension [1]. We have seen this shoulder in our own data (see §14.1.1) and in the literature [2,3,4].

The correlation integral probes the geometry of the attractor by measuring distances between pairs of points on the attractor. In Chapter Four, we showed that an ensemble (E_2) of distances between uncorrelated points on the attractor leads to accurate finite N estimates of the correlation integral [5]. As long as the dynamical system is chaotic [6], pairs of points well separated in time are essentially uncorrelated, and the distances between these pairs effectively satisfy the assumptions of the E_2 ensemble.

On the other hand, we *do* expect correlations for pairs of points that are not well separated in time. For time series with positive autocorrelation, pairs of points within a characteristic autocorrelation time τ of each other will be correlated, and this (dynamical, not geometrical) correlation can alter the shape of the finite N correlation integral. The distances between these correlated pairs of points do not really reflect the geometrical properties (such as the dimension) of the attractor.

In this chapter, we show how this dynamical correlation causes a kink to

form in the correlation integral at $C(N,r)\approx 2/N$, spoiling the $O(N^2)$ scaling range of $C(N,r)$. The kink vanishes in the $N\to\infty$ limit, but only for $N\gtrsim 2\tau^{\nu/2}$, where τ is the autocorrelation time (in units of sample time), and ν is the dimension of the attractor. We propose a generalized finite N correlation integral $C(W,N,r)$, which disregards pairs of points closer together in time than W. We find that $C(W\approx\tau,N,r)$ more accurately reflects the nature of the attractor, and in particular, converges toward $C(r)$ more efficiently than the standard $C(N,r)$.

We introduce a stochastic model from which analytic results are obtained; then we consider a deterministic dynamical system for which numerical results bear out the analytical conclusions.

9.1 The modified correlation algorithm

We begin with a time series $\{v_0,v_1,\ldots,v_{N-1}\}$ of m dimensional vectors $v_i\in\mathbb{R}^m$. For what follows, it does not matter whether the vectors were measured directly or constructed by embedding a one-dimensional time series into $\mathbb{R}^m$. Recall the definition of the correlation integral [7],

$$C(N,r) = \frac{2}{N(N-1)} \sum_{0\le i<j<N} H(r-\|v_i-v_j\|), \tag{9.1}$$

which we rewrite in equivalent form for the purposes of this exposition

$$C(N,r) = \frac{2}{N(N-1)} \sum_{n=1}^{N-1} \sum_{i=0}^{N-1-n} H(r-\|v_i-v_{i+n}\|). \tag{9.2}$$

Here, we propose a generalization that eliminates those pairs (v_i,v_{i+n}) with $n<W$.

$$C(W,N,r) = \frac{2}{(N+1-W)(N-W)} \sum_{n=W}^{N-1} \sum_{i=0}^{N-1-n} H(r - \|v_i - v_{i+n}\|) \tag{9.3}$$

Note that W=1 is just the standard algorithm. What we will show is that W>1 can improve the convergence properties of the correlation integral.

9.2 Autocorrelated Stochastic data

Consider a limited (finite N) time series of autocorrelated Gaussian noise [8]. Specify the mean $\mu=0$, the variance σ^2, and the autocorrelation $\alpha<1$. For convenience [9], we will further assume that the coordinates of v_i are independent of each other. Thus, we might write

$$v_i = (x_i, y_i, z_i, \ldots), \tag{9.4}$$

where $x_i \varepsilon R$ is one of the coordinates of v_i. And then

$$\langle x_i \rangle = \mu = 0 \tag{9.5}$$

$$\langle x_i^2 \rangle = \sigma^2 \tag{9.6}$$

$$\frac{\langle x_i x_{i+n} \rangle}{\sigma^2} = \alpha^n. \tag{9.7}$$

Note that as $\alpha \to 1$, the time series becomes more and more highly autocorrelated. The autocorrelation time (in units with $\Delta t=1$) is given by $\tau = 1/ln(1/\alpha)$ or $\tau \approx 1/(1-\alpha)$ for α near unity.

Because we have stochastic data, we can use statistical methods to obtain an analytical expression for the correlation integral. We will see that the behavior of the correlation integral can be classified into two qualitatively

distinct regimes. If N is large enough (or α small enough), then the effect of autocorrelation is negligible, the trajectory "fills out" the phase space, and the slope of the *log* C(W,N,r) versus *log* r curve approaches the embedding dimension m for $r \ll \sigma$. On the other hand, for N not sufficiently large (or for α too near unity), the effect of autocorrelation becomes noticeable and although the trajectory still fills out phase space, the correlation integral is not so well behaved; a structure is induced in the correlation curve, which inhibits good dimension estimates. And as we will see, the "sufficiently large" N that separates these two regimes can be extremely large.

Rewrite Equation (9.3) for the correlation integral, replacing the inner sum over Heaviside functions with its expectation value

$$C(W,N,r) = \frac{2}{(N+1-W)(N-W)} \sum_{n=W}^{N-1} (N-n)\, P(\|v_i - v_{i+n}\| \le r), \tag{9.8}$$

where P(**X**) denotes the probability that statement **X** is true. Now,

$$P(\|v_i - v_{i+n}\| \le r) = P(|x_i - x_{i+n}| \le r \ and \ |y_i - y_{i+n}| \le r \ and \ |z_i - z_{i+n}| \le r \ and \ \cdots) \tag{9.9}$$

$$= \left[P(|x_i - x_{i+n}| \le r)\right]^m . \tag{9.10}$$

For correlated Gaussian variables, the probability density is

$$P(x \le x_i \le x+dx,\ y \le x_{i+n} \le y+dy) = \frac{1}{2\pi\sigma^2\sqrt{1-\alpha^{2n}}} exp\left[-\frac{x^2 - 2\alpha^n xy + y^2}{2\sigma^2(1-\alpha^{2n})} \right] dx\, dy. \tag{9.11}$$

Then

$$P(|x_i - x_{i+n}| \le r) = \frac{1}{2\pi\sigma^2\sqrt{1-\alpha^{2n}}} \int_{-\infty}^{\infty} dx \int_{x-r}^{x+r} dy \; exp\left[- \frac{x^2 - 2\alpha^n xy + y^2}{2\sigma^2(1-\alpha^{2n})} \right] \tag{9.12}$$

$$= erf\left[\frac{r}{2\sigma\sqrt{1-\alpha^n}}\right], \tag{9.13}$$

where erf is the error function. Substitution back into Equation (9.8) yields our main analytical result

$$C(W,N,r) = \frac{2}{(N+1-W)(N-W)} \sum_{n=W}^{N-1} (N-n)\left[erf\left[\frac{r}{2\sigma\sqrt{1-\alpha^n}}\right] \right]^m. \tag{9.14}$$

9.2.1 Uncorrelated limit

In the case of zero autocorrelation, Equation (9.14) reduces to a simple form:

$$C(W,N,r) = [erf(r/2\sigma)]^m \tag{9.15}$$

for $\alpha = 0$. Recall that $erf(x) \propto x$ for $x \ll 1$ and $erf(x) \to 1$ for $x \gg 1$. The correlation integral looks like $C(W,N,r) \approx \frac{r^m}{(\sigma\sqrt{\pi})^m}$ for $r \ll 2\sigma$. And the exponent m is just the value that we want our log-log plot to pick out.

In fact, this is the same limit that is approached for $N \to \infty$, independent of α. Since $\alpha^n \to 0$ as $n \to \infty$, *most* of the terms in the sum will be error functions with arguments very near $r/2\sigma$; thus,

$$\lim_{N \to \infty} C(W,N,r) = [\mathrm{erf}(r/2\sigma)]^m \tag{9.16}$$

for any value of $\alpha<1$, and again the embedding dimension m will be approached.

9.2.2 Effect of Autocorrelation

On the other hand, if α is very nearly one, or if N is *not* sufficiently large, the sum in Equation (9.14) cannot be so simply expressed. For small n, $1-\alpha^n$ is noticeably less than unity and the argument of the error function will be noticeably larger than $r/2\sigma$. Also, since the *erf* is raised to the *m*th power, this effect is magnified with greater embedding dimension. Although $r/2\sigma$ is a good approximation to the argument of the *erf* for most of the terms (those with large n), those few for which this is not the case can actually dominate the sum for small r.

In particular, for $r \ll 2\sigma\sqrt{1-\alpha^W}$, the first term (n=W) of the right hand side of Equation (9.14) is

$$\frac{2}{N+1-W}\left[erf\left[\frac{r}{2\sigma\sqrt{1-\alpha^W}}\right]\right]^m \approx \frac{2(1-\alpha^W)^{-m/2}}{N}\left(\frac{r}{\sigma\sqrt{\pi}}\right)^m. \tag{9.17}$$

As $N\to\infty$ this vanishes. But if $N \ll 2(1-\alpha^W)^{-m/2}$, then the first term will be much larger than $[erf(r/2\sigma)]^m \approx \left(\frac{r}{\sigma\sqrt{\pi}}\right)^m$ and indeed will dominate the entire sum. In this case,

$$C(W,N,r) = \frac{2}{N+1-W}\left[erf\left[\frac{r}{2\sigma\sqrt{1-\alpha^W}}\right]\right]^m \tag{9.18}$$

for $r \ll 2\sigma\sqrt{1-\alpha^{W+1}}$. For $r \approx 2\sigma\sqrt{1-\alpha^W}$, the error function saturates at unity and a log-log plot of C(W,N,r) versus r displays a plateau at $C(W,N,r)\approx 2/N$. Finally, for $r > \sigma\sqrt{\pi}(2/N)^{1/m}$, the first term loses its significance and the

correlation integral C(W,N,r) begins to look like its corresponding $\alpha \to 0$ (or $N \to \infty$) limit.

Now, the least nonzero value that C(W,N,r) may have is $2/N^2$; this is because the number of distances must be integral. The *usable* range of C(W,N,r), *i.e.*, the range over which $C(W,N,r) \propto r^m$, will be between $2/N^2$ and of order 1 for the uncorrelated limit, and between $2/N^2$ and $2/N$ for the case where autocorrelation is important. These two cases are *qualitatively* different, and the first is better by a factor of N/2. On logarithmic axes, the first has almost double the range of the second.

The uncorrelated limit may be achieved by taking N sufficiently large, but "sufficiently large" can be tremendously large. With $W=1$, one needs $N \gg 2\tau^{m/2}$. For example, $\tau=10$ and $m=20$ demands $N \gg 2\times 10^{10}$. This is $N^2/2 \gg 10^{20}$ distances to compute! One possibility is to decrease the sampling rate in the original data, thereby decreasing τ. Another recommendation, which our notation has probably made obvious by now, is to take $W>1$.

9.2.3 Recommendations for W

As a minimum recommendation, we point out that if $W > \tau(2/N)^{2/m}$, then there will be "sufficiently many" data points N that the range of linearity in the log-log plot will not be compromised. We note that this W is typically much less than N, so the modification is actually quite minor.

Up to now, the problem has been discussed as one that "goes away" when $N \to \infty$. As has been seen, though, it is the first few terms that cause all the trouble: and they do not go away; they are merely overwhelmed. A better

algorithm, *even if* N *is sufficiently large*, is to toss out those overcontributing early terms right from the start. In the example of autocorrelated stochastic data, this is achieved with $W > \tau \ ln(m/2)$.

From a more intuitive point of view, the taking of $W \gtrsim \tau$ ensures that the small r behavior of the correlation integral counts only the "accidentally" close pairs of vectors; it is not biased by those pairs whose vectors are close in space only because they are close in time. In other words, it is the *geometrical* properties of the attractor that are measured, unbiased by the *dynamical* process by which the attractor points are sampled.

9.3 Numerical Results

In this section we verify our analytical results with two numerical examples. First, we consider autocorrelated stochastic data such as was discussed in the analytical model. Then, we show that the results are more general, and apply as well to a deterministic dynamical system with autocorrelation.

9.3.1 Stochastic data

With the initial goal of mimicking real data from a specific physical system (the Caltech research tokamak [10]), we created a one-dimensional time series with $\mu=0$, $\sigma=20$, $\alpha=0.9$ (so $\tau \approx 10$), and $N=10000$. We created m dimensional vectors, using the delayed coordinate embedding with a delay time of $T=5$ so that a typical vector is

$$v_i = (x_i, x_{i-5}, x_{i-10}, \cdots, x_{i-5(m-1)}). \tag{9.19}$$

Correlation integrals were computed with this data, see Figure 9.1(a). A quick and careless look at these curves might suggest a slope ν that saturates with increasing m. A closer look, however, reveals a more complicated structure. There is an extra shoulder, due almost entirely to the (anomalously large) $n=1$ term. When a $W=2$ curve is plotted (Figure 9.2), the shoulder disappears, though as the $W>2$ curves demonstrate, anomalous contributions come also from the $n=2$ term. Notice, however, that for $W\geq 3$ the correlation integrals are essentially unaffected by further increases in W. Figure 9.3(a) shows that the spurious saturation with m that was seen in the $W=1$ curves is not present for $W=\tau=10$; as m increases so does the slope of the *log* C(W=10,N,r) versus *log* r curve. These effects are more dramatically apparent in plots (Figures 9.1(b) and 9.2(b)) of the slopes of the C(W,N,r) curves as a function of r.

9.3.2 Deterministic dynamical data

Although the effect we describe is best modeled with stochastic data, it is, in fact, a general feature of autocorrelated input and can be seen in dynamical data as well. The Mackey-Glass [11] differential delay equation,

$$\frac{dx}{dt} = \frac{0.2x(t-s)}{1+[x(t-s)]^{10}} - 0.1x(t), \tag{9.20}$$

models a dynamical system of arbitrary complexity. Strictly there are an infinite number of degrees of freedom — note that the initial condition is the function x(t) specified over the range $t\varepsilon[t_o-s,t_o]$. Farmer [12], however, has

found that the *effective* number of degrees of freedom (*i.e.*, the dimension) is finite and increases with delay time s. Grassberger and Procaccia [ ] give a dimension of about 7.5 for the s=100 case. Using s=100 and the same numerical algorithm that is used in [ ], but increasing the sampling frequency so that the autocorrelation time is $\tau=10$, and renormalizing so that $\mu=0$, and $\sigma=20$ — all this so that comparisons can be made with the random data), we compute correlation integrals of the Mackey-Glass data for various W and m. As in the stochastic case, the standard (W=1) curves display the unwanted shoulders (Figure 9.4(a)) and the modified (W=10) curves do not (Figure 9.5(a)). Again, the plots of the slopes are especially compelling. Convergence of the slope to the attractor dimension ($\sim$7.5) is readily apparent for the modified correlation (Figure 9.5(b)), but no convergence is seen in the curves (Figure 9.4(b)) obtained by the standard algorithm.

9.4 Conclusion

We find that the introduction of a cutoff parameter W>1 improves the convergence of the standard correlation algorithm toward its $N\to\infty$ limit. Although we recommend $W\approx\tau$, where τ is the autocorrelation time of the input time series; we point out that as long as $W>\tau(2/N)^{2/m}$, where N is the number of points in the time series and m is the embedding dimension, the exact choice of W is not important.

As a final comment, we remark that these W>1 curves, once the standard W=1 curves had already been calculated, were very easy to obtain. We merely computed the n=1, 2, ..., W−1 terms separately (each of which required only 2/N, or 0.02% in our examples, of the work required to compute the whole curve)

and subtracted them from the W=1 curve.

9.5 Notes and References

[1] James Theiler. "Spurious dimension from correlation algorithms applied to limited time series data," Phys. Rev. **A** 34 (1986) 2427.

[2] S. J. Gee and J. B. Taylor. "Dimension measurement of fluctuations in HBTX1A," Proceedings of the 12th European Conference on Controlled Fusion and Plasma Physics (Budapest, September 1985). The authors cannot explain the characteristic shoulder that appears only when the time series signal is low-pass filtered. Since the effect of a low-pass filter is to increase autocorrelation time, we argue that the explanation given in this chapter is appropriate.

[3] A similar effect is described in P. Atten, J. G. Caputo, B. Malraison, and Y. Gagne. "Détermination de dimension d'attracteurs pour différents écoulements," Journal de Mécanique théorique et appliquée, Numéro spécial (1984) 133.

[4] M. L. Sawley, C. W. Simm, and A. Pochelon. "The correlation dimension of broadband fluctuations in a tokamak," Phys. Fluids 30 (1987) 129; also, C. W. Simm, M. L. Sawley, F. Skiff and A. Pochelon. "On the analysis of experimental signals for evidence of deterministic chaos," preprint (1986). The authors observe the shoulder we describe in this chapter (and in Reference [1]), and apply the remedy we prescribe.

[5] By "accurate" we mean that $\langle C(N,r)\rangle = C(r)$, where $\langle\rangle$ denotes an ensemble average. This technical definition is introduced in §4.1.

[6] Nonchaotic systems, such as the quasiperiodic systems of Chapter Eight, can have autocorrelation times as long as the time series itself. In this case, the

remedies of this chapter cannot be usefully applied.

[1] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors," Physica 9D (1983) 189.

[2] Formally, this is an Ornstein-Uhlenbeck process; see J. Doob. "The Brownian movement and stochastic equations," Ann. Math. 43 (1942) 351.

[3] This assumption is not always true, especially if the coordinates are time-delayed values from an original one-dimensional time series, but it does simplify the analysis. It is still possible to make analytical estimates, assuming some correlation between the coordinates, but the extra complication only obscures the main result. In any case, we find that our numerical experiments agree with the simplified analytical results.

[4] S. Zweben and R. Gould. "Scaling of edge-plasma turbulence in the Caltech tokamak," Nuclear Fusion 23 (1983) 1625.

[5] M. C. Mackey and L. Glass. "Oscillations and chaos in physiological control systems," Science 197 (1977) 287.

[6] J. D. Farmer. "Chaotic attractors of an infinite-dimensional dynamical system," Physica 4D (1982) 366.

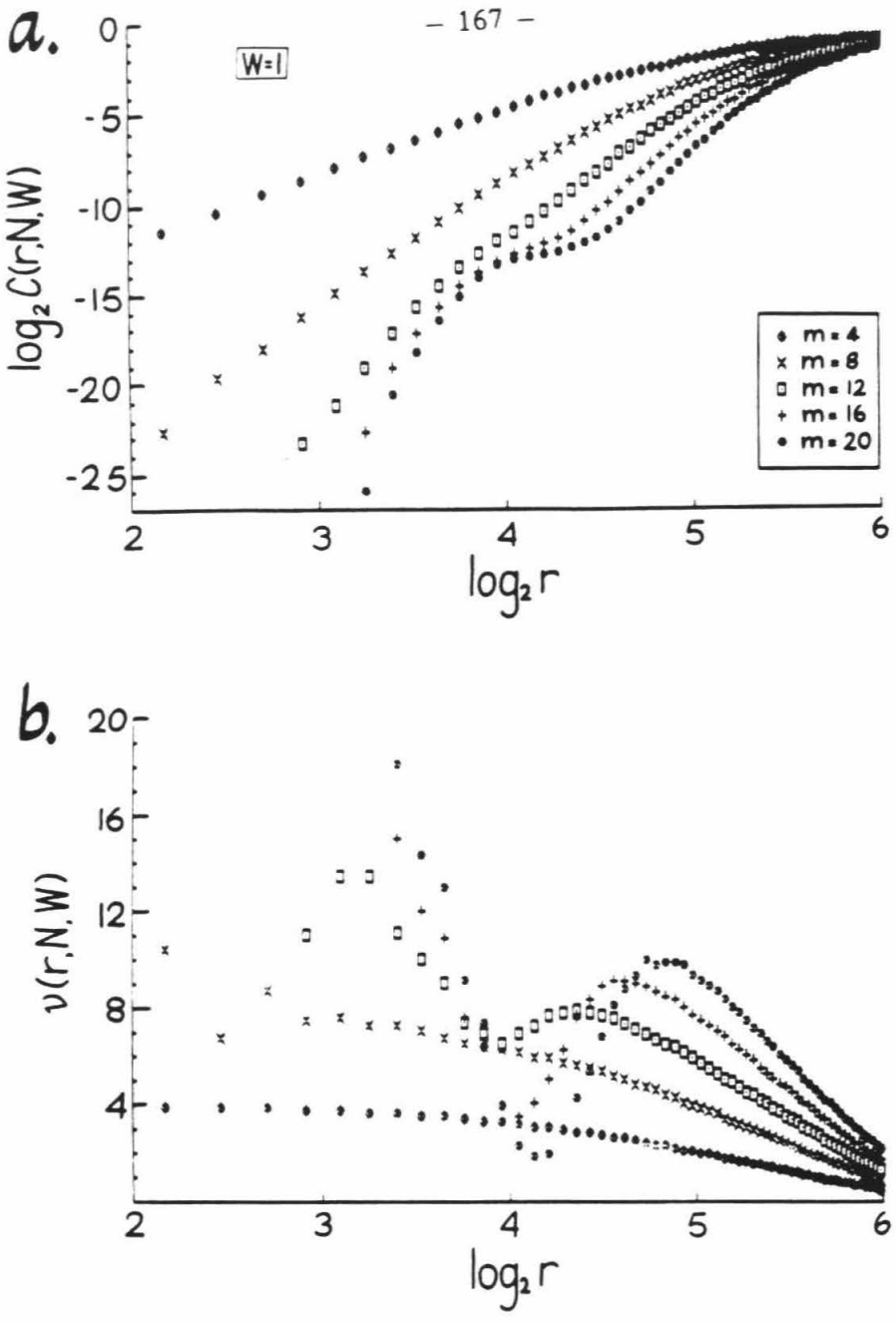


Figure 9.1 Standard correlation integrals for autocorrelated random data. (a) Log-log plot of the standard (W=1) correlation integral over a range of embedding dimension m for stochastic data with N=10000 points, standard deviation $\sigma=20$, and autocorrelation $\alpha=0.9$. Notice the horizontal plateau at $C(N,W,r)=2/N$; *n.b.*, $\log_2(2/N)\approx -12.29$. (b) Slope of the curves in (a). Here, the derivative $\nu(N,W,r) \equiv d[log\ C(N,W,r)]/d[log\ r]$ is approximated by $\nu(N,W,r) = \Delta[log\ C(N,W,r)]/\Delta[log\ r]$, where the operator Δ is defined by $\Delta f(r) \equiv f(r+1) - f(r)$.

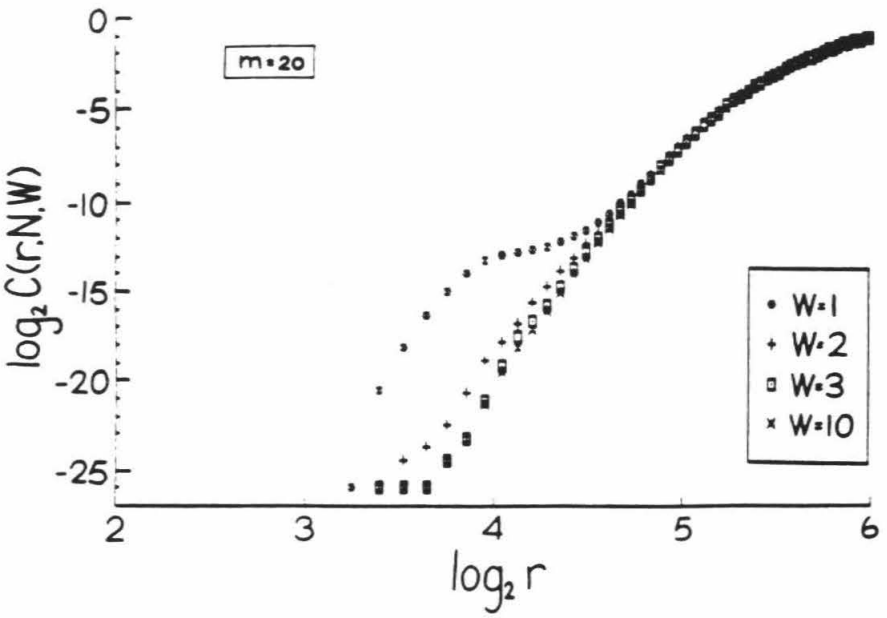


Figure 9.2 Modified correlation integrals for stochastic data. Embedding dimension is fixed at m=20 and the cutoff parameter W is varied.

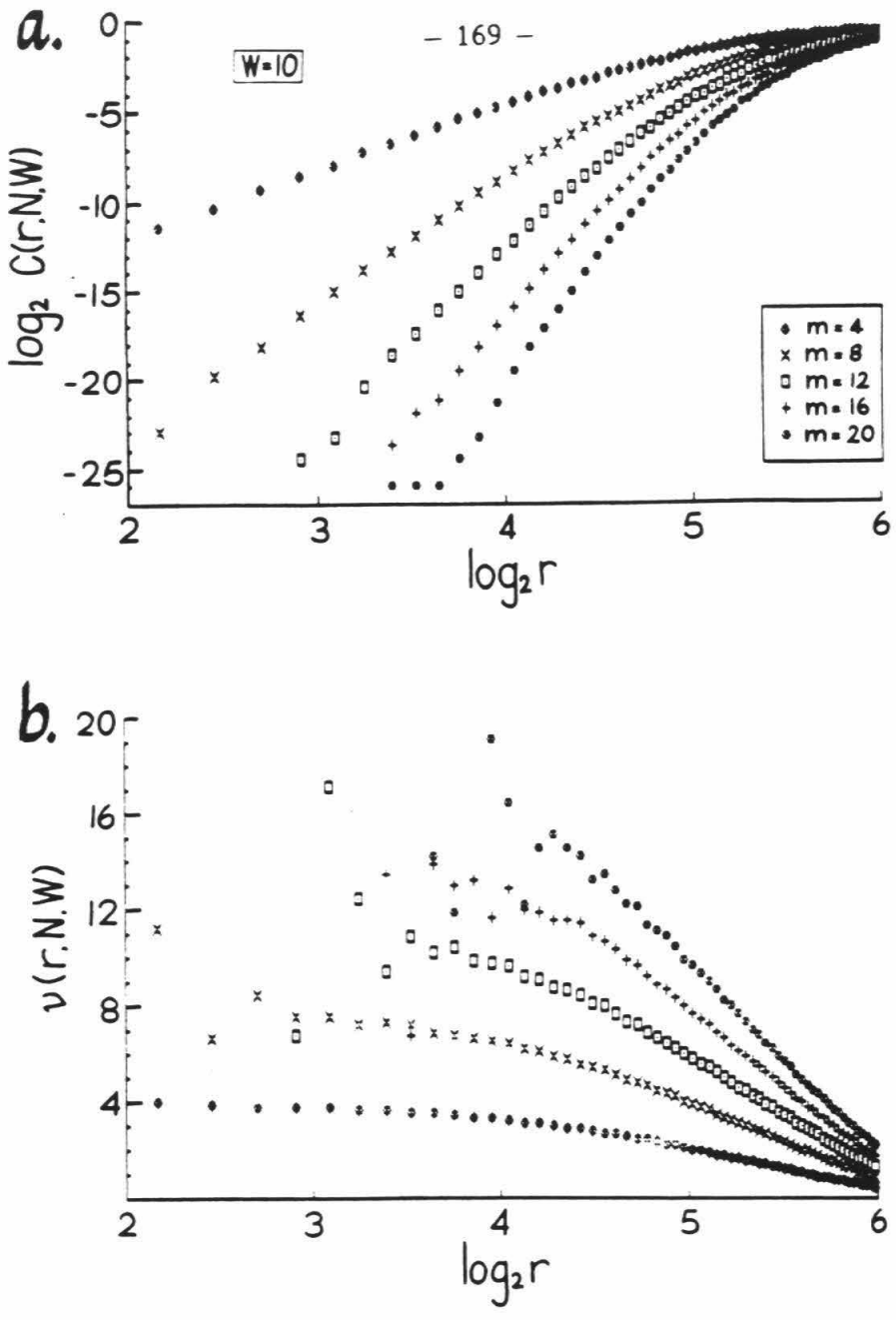


Figure 9.3 Modified correlation integrals for stochastic data. (a) Log-log plot of the modified (W=10) correlation integral over a range of embedding dimension m. (b) Slopes of the curves in (a).

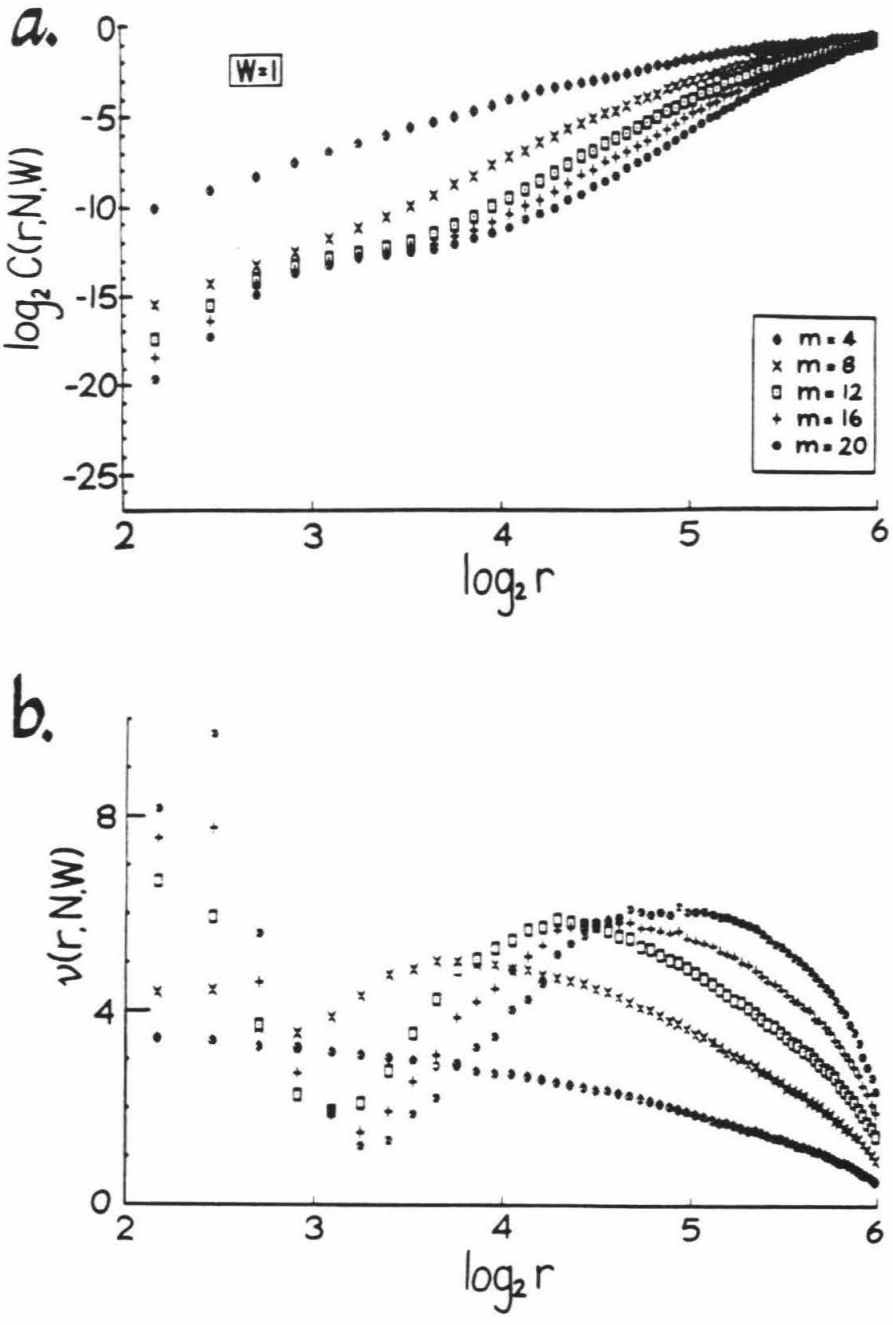


Figure 9.4 Standard correlation integral for Mackey-Glass dynamical data. (a) Log-log plot of standard correlation integral over a range of embedding dimension m for Mackey-Glass differential delay equation with N=10000 points, renormalized so that $\sigma=20$ and sampled at a rate $\Delta t=2$. (b) Slope of curves in (a).

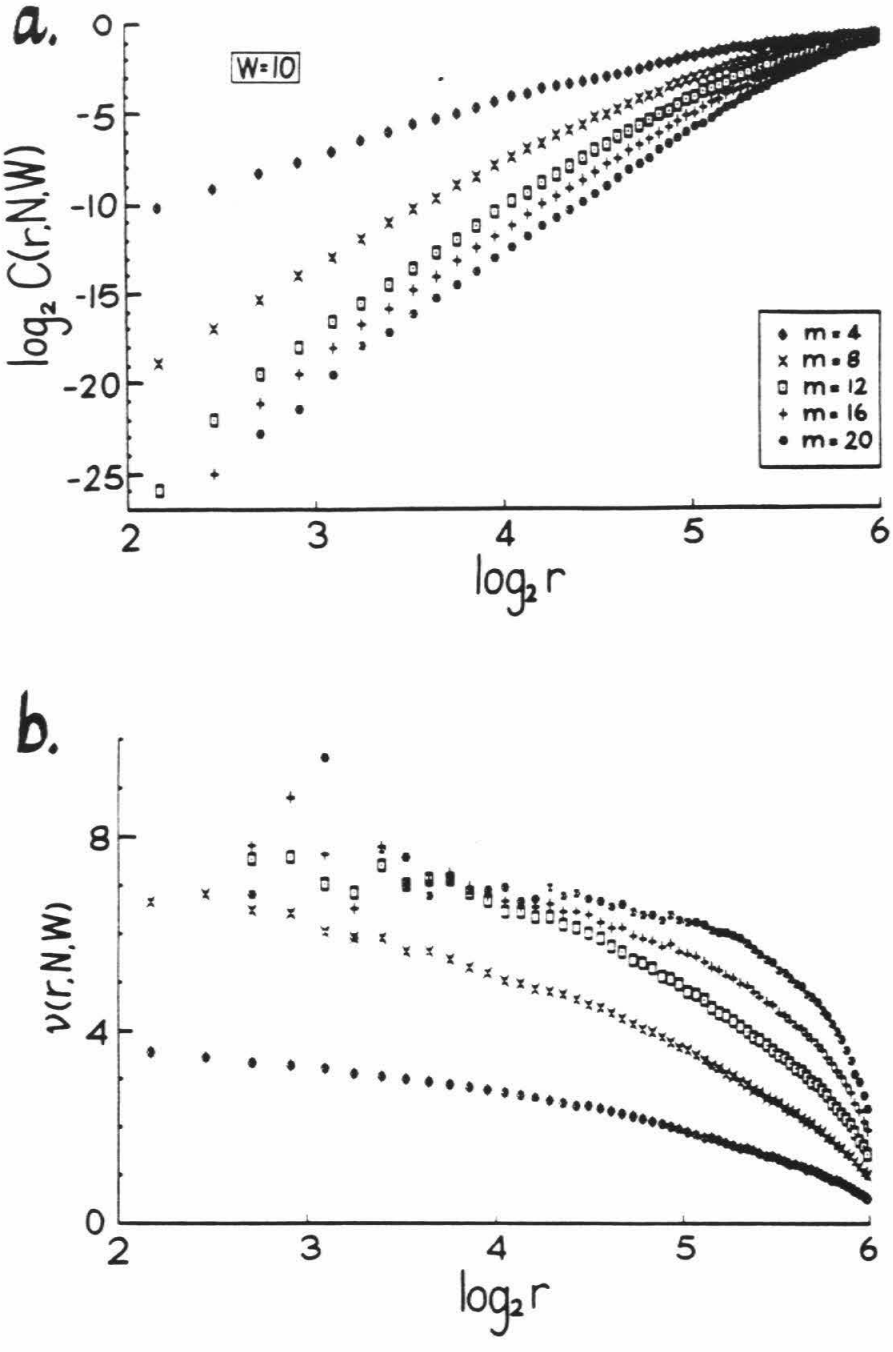


Figure 9.5 Modified correlation integral for Mackey-Glass dynamical data. (a) Log-log plot of modified (W=10) correlation integral over a range of embedding dimension m for Mackey-Glass differential delay equation. (b) Slope of curves in (a). Note convergence toward dimension $\nu \approx 7.5$.

CHAPTER TEN

10. STATISTICAL ERROR

To a dynamical system we associate an ensemble of finite N time series, each corresponding to a different initial condition on the attractor. For each time series we find a single dimension estimate $d(N;X_0)$, indexed by the initial condtion $X_0 \varepsilon \mathcal{A}$. We have, in effect, an ensemble of dimension estimates.

In practice, we are provided with only a single finite N time series, so our estimate of dimension is a single value $d(N;X_0)$ for some X_0. If there are no systematic effects to bias our estimate, then the ensemble average is equal to the actual attractor dimension,

$$d = \langle d(N;X_0) \rangle. \tag{10.1}$$

The root-mean-square deviation

$$\sigma(N) \equiv \sqrt{\langle d^2(N,X_0) \rangle - \langle d(N,X_0) \rangle^2}. \tag{10.2}$$

provides the statistical error bar that we associate with our dimension estimate. Of course, we expect $\sigma(N) \to 0$ as $N \to \infty$, so the more points N in the time series, the closer is a typical estimate $d(N;X_0)$ to the actual dimension d.

Despite the abundance of dynamical systems whose dimension is calculated, estimates of error bars in the literature have primarily been heuristic, "educated guesses."

10.1 Least-squares fit

The commonest method for estimating the error in the slope of a *log* C(N,r) versus *log* r curve has been to take the usual linear fit error bar corresponding to the spread of points from a straight line. This method, unfortunately, is flawed in a fundamental way, and nearly always underestimates the correct value of the error. Probably this, as much as anything else, has led to the folklore of distrust that surrounds the correlation algorithm.

Difficult as it may be to get an error bar for the final dimension estimate, it is fairly straightforward to put error bars on values of C(N,r) in the correlation integral. Recall that C(N,r) counts the fraction of distances (out of a total $N^2/2$) less than r. In particular, there are $D=N^2C(N,r)/2$ of these distances. Treating distances as independent [1] allows us to use the Poisson error of $\sqrt{D}$. The error bar on C(N,r) is therefore

$$\sigma_{C(N,r)} = \frac{2}{N^2}\sqrt{D} = \sqrt{\frac{2\ C(N,r)}{N^2}}. \tag{10.3}$$

Using these error bars on the individual C(N,r) points, a weighted fit can be made through the log-log plot of C(N,r) versus r.

The problem with this method is that the values of C(N,r) for various r are not independent. If C(N,r) counts the fraction of distances less than r, then $C(N,r+\epsilon)$ is the sum of C(N,r) and the fraction of distances between r and $r+\epsilon$. It is clear that $C(N,r+\epsilon)$ is quite dependent on C(N,r), especially for small ϵ. If C(N,r) is computed at equally spaced intervals in r, the effect is to overweigh

the points with large r. In principle this should not affect the dimension estimate itself, as long as r is in the scaling regime over which $C(N,r) \propto r^{\nu}$, but it does decidedly underestimate the error bar.

For ϵ large (more particularly, for $C(N,r+\epsilon) \gg C(N,r)$), this objection is not as significant. In this case, though, there are few points through which the straight line is to be fit, and with only a few points on the C(N,r) curve, it is easy to miss second order features (such as the shoulder that is discussed in Chapter Nine or the oscillations due to lacunarity in Chapter Seven), which are diagnostic of systematic defects in the correlation integral.

One possible compromise is to compute dimension by least-squares fit through all of the points in the scaling regime of the correlation integral, but then to estimate statistical error by considering the spread from this line of only a few well-separated points. This is something of a tortured compromise, and although we do not have any objections, we do not necessarily recommend it [2]. It provides, after all, a very *indirect* measurement of the error bar defined in Equation (10.2).

Experience has shown that *unweighted* least-squares fits often give better dimension estimates than the weighted fits described above. That is because the failure to take into account the extra relative accuracy of the large C(N,r) values tends to give more weight to the small r values. This tends to compensate for the overvaluation of large r values discussed above; also, the added weight given to the small r values tends to favor the $r \to 0$ limit. However, this coincidental cancellation of effects (the magnitudes of which we have no reason to think will be commensurate) does not by any means justify the use of

an unweighted fit.

10.2 Takens' maximum likelihood estimate

Rather than trying to find the best-fit slope on a log-log plot of the correlation integral C(N,r) versus r, Takens [3] provides an estimate of ν directly from the list of all the D distances $\{r_0, r_1, r_2, \ldots, r_{D-1}\}$ that are less than some fixed r_o. Using the method of maximum likelihood, and assuming that the distances are independently chosen from a probability density $P(r) \propto r^{\nu-1}$, Takens finds an estimator for ν,

$$\hat{\nu} = \frac{-1}{\frac{1}{D}\sum_{i=0}^{D-1} ln(r_i/r_o)}, \tag{10.4}$$

which has an expected error

$$\sqrt{<(\nu-\hat{\nu})^2>} = \frac{\hat{\nu}}{\sqrt{D}}. \tag{10.5}$$

Takens further shows that this estimator is optimal in that its expected error is the smallest of all estimators.

We note that the number of distances less than r_o is $D = N^2 C(N,r_o)/2$, which for fixed r_o, scales as N^2; so Takens' estimate has an error bar of

$$\sigma = \frac{\hat{\nu}}{N\sqrt{C(N,r_o)/2}}, \tag{10.6}$$

which scales as 1/N for fixed r_o. If the correlation integral scales as $C(N,r) \propto r^{\nu}$,

then we have

$$\frac{\sigma}{\nu} = O(N^{-1} r_o^{-\nu/2}). \tag{10.7}$$

Equations (10.4, 10.5) are based on assumptions that may not be strictly true, namely, that the distances $\{r_0, r_1, r_2, \ldots, r_{D-1}\}$ are truly independent, and that the correlation integral C(N,r) scales exactly as r^{ν}. For a deterministic dynamical system, the distances between pairs of points cannot be truly independent. On the other hand, if the dynamical system is chaotic, then as nearby trajectories diverge, they "forget" about their previous association, and distances well separated in time become effectively independent. The effect on Equation (10.5) is to replace D with D_{eff}, which we presume is proportional to D. Thus, the O(1/N) scaling still holds, but the coefficient is increased [4]. For the Hénon attractor, this increase amounts to a factor of about three. Note that for a quasiperiodic (or any nonchaotic) dynamical system, the assumption of independence fails completely and the O(1/N) scaling can no longer be trusted.

In other chapters (Chapters Six and Seven, most notably), we have seen examples of strange attractors for which C(N,r) does not scale strictly as r^{ν}. We might ask whether these effects will alter the *statistical* error bar predicted in Equation (10.6). A full investigation of this question remains to be undertaken.

Numerical evidence allows us to claim that an estimate of ν based on a linear fit through a log-log plot provides a value in most cases as good as $\hat{\nu}$. What the linear fit does not provide is an estimate of the error bar. We

suggest that the error bar in Equation (10.6) is appropriate, though we note that the coefficent may be off by a constant factor.

10.3 Average pointwise dimension

Because of the difficulty in obtaining reliable error bars on dimension estimates, Holzfuss and Mayer-Kress [5] have advocated individual pointwise dimension estimates at each point (or at a sampling of N_{ref} reference points) on the attractor. The immediate advantage of this approach is that from a single time series many estimates of (pointwise) dimension are taken. From these, a mean provides a natural estimate, and the standard deviation of the mean ($\sigma/\sqrt{N_{ref}}$, where σ is the ordinary standard deviation) provides an error bar.

Though more careful numerical experimentation is needed to justify the usefulness of this approach, a leading order analysis suggests that the approach is reasonable. Using Takens' estimate for pointwise dimension will give estimates of pointwise dimension that have a standard deviation of $O(1/\sqrt{D})$, where D is the number of distances in the estimate; we have $D \lesssim N$, so we can estimate the standard deviation as $O(1/\sqrt{N})$. The standard deviation of the mean is therefore $O(1/\sqrt{N_{ref}N})$. With $N_{ref}=N$, which means that all $O(N^2)$ distances are computed, we predict a statistical error of $O(1/N)$, the same as the full correlation integral.

10.3.1 Nonuniformity

A possible disadvantage to the use of the average pointwise dimension is that the pointwise dimension varies over the attractor. Although "almost all" of the points on the attractor have the same pointwise dimension, the finite

resolution at which the attractor is viewed leads to a nonuniformity in the *observed* pointwise dimension. This is an effect we shall attempt to quantify in this section.

Consider the model of the weighted Cantor set introduced in §7.1.2. In the language of that chapter, a typical point x could be expressed

$$x = ".s_1 s_2 \ldots" = \sum_j s_j n^{-j}, \tag{10.8}$$

but since the set would be viewed with a finite resolution of, say, $\epsilon = n^{-m}$, the point could be approximated by $x = ".s_1 s_2 \ldots s_m$," and the dimension at that point would be observed to be

$$d = \lim_{r \to 0} \frac{log\ \mu(B_x(r))}{log\ r} \approx \frac{log\ p_{s_1} p_{s_2} \cdots p_{s_m}}{log\ n^{-m}}, \tag{10.9}$$

where μ is the invariant measure, and $B_x(r)$ is the ball of radius r centered at the point x. Thus,

$$d = \frac{\sum_k N_k\ log\ p_k}{m\ log(1/n)}, \tag{10.10}$$

where N_k is the number of times "k" appears in the sequence "$s_1 s_2 \ldots s_m$." Note that the expected value $<N_k>$ is given by $m \cdot p_k$, and so the average value <d> of this measurement of pointwise dimension gives the information dimension,

$$<d> = \frac{\sum_k <N_k>\ log\ p_k}{m\ log(1/n)} = \frac{\sum_k p_k\ log\ p_k}{log(1/n)}. \tag{10.11}$$

To quantify how much a typical value of d varies from this average, compute the variance

$$\langle d^2 \rangle = \frac{\sum_j \sum_k \langle N_j N_k \rangle \log p_j \log p_k}{(m \log(1/n))^2}. \tag{10.12}$$

Now the joint probability distribution of N_j and N_k, for $j \neq k$, is given by

$$P(N_j=s, N_k=t) = \frac{m!}{s!t!(m-s-t)!} p_j^s p_k^t (1-p_j-p_k)^{m-s-t}. \tag{10.13}$$

For j=k, this reduces to

$$P(N_j=s, N_j=t) = P(N_j=s)\delta_{st} = \frac{m!}{s!(m-s)!} p_j^s (1-p_j)^{m-s} \delta_{st}, \tag{10.14}$$

from which it follows that

$$\langle N_j N_k \rangle = m(m-1) p_j p_k + m p_j \delta_{jk}. \tag{10.15}$$

Hence,

$$\langle d^2 \rangle = \frac{\sum_j \sum_k \left[m(m-1) p_j p_k + m p_j \delta_{jk} \right] \log p_j \log p_k}{(m \log(1/n))^2} \tag{10.16}$$

$$= \frac{m(m-1) \sum_j \sum_k p_j p_k \log p_j \log p_k + m \sum_j \sum_k p_j \delta_{jk} \log p_j \log p_k}{(m \log(1/n))^2} \tag{10.17}$$

$$= \frac{m(m-1)\left[\sum_k p_k \log p_k\right]^2 + m \sum_k p_k (\log p_k)^2}{(m \log(1/n))^2} \tag{10.18}$$

$$= \frac{m-1}{m} <d>^2 + \frac{1}{m} \frac{\sum_k p_k (\log p_k)^2}{(\log(1/n))^2}. \tag{10.19}$$

$$= <d>^2 + \frac{1}{m}\left[-<d^2> + \frac{\sum_k p_k (\log p_k)^2}{(\log(1/n))^2}\right]. \tag{10.20}$$

Write

$$<f(p)> \equiv \sum_k p_k f(p_k) \tag{10.21}$$

so that

$$<\log p> = \sum_k p_k \log p_k \tag{10.22}$$

and

$$<(\log p)^2> = \sum_k p_k (\log p_k)^2. \tag{10.23}$$

Note then that, from (10.10) and (10.21),

$$<d> = \frac{<\log p>}{\log(1/n)}. \tag{10.24}$$

Also note that

$$\sigma_d^2 \equiv <d^2> - <d>^2 = \frac{1}{m} \frac{<(\log p)^2> - <\log p>^2}{(\log(1/n))^2} \tag{10.25}$$

or

$$\sigma_d^2 = \frac{1}{\log \epsilon}\left\{\frac{<(\log p)^2> - <\log p>^2}{\log(1/n)}\right\}, \tag{10.26}$$

where $\epsilon = n^{-m}$ is the resolution to which the dimension in computed, and the term

in braces is a coefficient that corresponds to the "nonuniformity" of the set. We will show that this nonuniformity can be expressed in terms of the generalized dimensions of Hentschel and Procaccia [6]. In [6], this formula is given for the generalized dimension:

$$D_q = \frac{1}{q-1} \lim_{\epsilon \to 0} \frac{log \langle e^{(q-1)log\ p} \rangle}{log\ \epsilon}, \tag{10.27}$$

where the angle brackets <> denote the average of probabilities p in boxes of size ϵ. Since our model involves a strict self-similarity, we can write

$$D_q = \frac{1}{q-1} \frac{log \langle e^{(q-1)log\ p} \rangle}{log(1/n)}, \tag{10.28}$$

and now the angle brackets have the meaning assigned them in Equation (10.21). In particular, we find

$$\left.\frac{dD_q}{dq}\right|_{q=1} = \frac{\langle (log\ p)^2 \rangle - \langle log\ p \rangle^2}{2\ log(1/n)}, \tag{10.29}$$

so that with the definition [7],

$$\lambda \equiv -\frac{1}{D_q} \left.\frac{dD_q}{dq}\right|_{q=1}, \tag{10.30}$$

the variance of d, given by Equation (10.26), becomes

$$\sigma_d^2 = \frac{-2\lambda D_1}{log\ \epsilon}. \tag{10.31}$$

Suppose we have N points on the d dimensional attractor, and let us estimate the pointwise dimension at each of these N points. The resolution ϵ to which we can compute pointwise dimension will be limited by the finite N, and in particular, $\epsilon = N^{-1/d}$. We have

$$\sigma_d^2 = \frac{-2\lambda d}{log\ \epsilon} = \frac{\lambda d^2}{log\ N}. \tag{10.32}$$

Now an average of N estimates of d will have an error bar of $1/\sqrt{N}$ the size of a single estimate, so our error bar on d will be

$$\sigma_d = \frac{\lambda d^2}{\sqrt{N\ log\ N}}. \tag{10.33}$$

We see that the statistical error bar for the average pointwise dimension is much larger, $O(1/\sqrt{N\ log\ N})$ instead of O(1/N), for nonuniform attractors. Whether or not a similar problem affects the ordinary correlation dimension remains to be investigated.

10.4 Notes and References

[1] This invokes the assumption of the E_2 ensemble discussed in Chapter Four.

[2] On the other hand, we note that this method ***does*** predict an error bar of O(1/N), in agreement with §10.2.

[3] Floris Takens. "On the numerical determination of the dimension of an attractor," in *Dynamical Systems and Bifurcations, Groningen* 1984, Vol. 1125 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin 1985).

[4] In the language of Chapter Four, this paragraph is arguing that the E_1 ensemble can replace the E_0 ensemble, and the effect on the error bar is only that it is multiplied by a constant (the scaling with N is not altered). We note that the Takens estimate actually assumes the E_2 ensemble (*distances* are independent); that we are justified in replacing E_1 by E_2 is argued in Chapter Four.

[5] Joachim Holzfuss and Gottfried Mayer-Kress. "An approach to error-estimation in the application of dimension algorithms," in *Dimensions and Entropies in Chaotic Systems*, ed. G. Mayer-Kress (Springer-Verlag, Berlin, 1986).

[6] H. G. E. Hentschel and Itamar Procaccia. "The infinite number of generalized dimensions of fractals and strange attractors," Physica 8D (1983) 435.

[7] I have chosen λ to correspond to the "uniformity factor" described in Section IV of R. Badii and A. Politi. "Renyi dimensions from local expansion rates," Phys. Rev. A 35 (1987) 1288.

CHAPTER ELEVEN

11. LIMITATIONS ON DIMENSION IMPOSED BY FINITE N

In the previous chapter, we discussed the statistical error that arises from finite N time series. In this chapter we note that there are systematic finite N errors as well. Taking as a specific systematic error the edge effect that comes from a Gaussian distribution of points in the embedding space, we will find that the number of points N needed to maintain a given error tolerance increases exponentially with ν. Finally, by comparing the systematic and the statistical error, we will obtain a strategy for choosing the optimal scaling regime cutoff r_o. It is for values of $r \leq r_o$ that we seek scaling $C(N,r) \propto r^{\nu}$.

We recall the definition of the correlation dimension

$$\nu \equiv \lim_{r \to 0} \frac{log\ C(r)}{log\ r}, \tag{11.1}$$

where $C(r) \equiv \lim_{N \to \infty} C(N,r)$.

In practice, we cannot take $N \to \infty$, since we have only a finite amount of data, and so we usually approximate $C(r) \approx C(N,r)$ with large N. The approximation is a reasonable one only for large r. In particular, for $r < r_{min}$ where r_{min} is the smallest of the $N(N-1)/2$ distances, we have $C(N,r)=0$, and the limit

$$\lim_{r \to 0} \frac{log\ C(N,r)}{log\ r}, \tag{11.2}$$

cannot be taken. Instead, we might define $\nu(N,r) = \frac{log\ C(N,r)}{log\ r}$ and then say $\nu \approx \nu(N,r)$ "for small r." Or, somewhat more practically,

$$\nu(N,r) \equiv \frac{d[log\ C(N,r)]}{d[log\ r]} = \frac{r}{C(N,r)} \frac{dC(N,r)}{dr} \tag{11.3}$$

noting as usual the exceptions discussed in Chapter Seven.

In Chapter Six, we discuss

$$\nu(r) \equiv \frac{r}{C(r)} \frac{dC(r)}{dr} \tag{11.4}$$

as an approximation to ν at finite (nonzero) r in the context of the "edge effect." In this chapter, we will estimate r_{min} as a function of N and ν, and then, approximating $\nu(N,r)$ by $\nu(r)$ at $r=r_{min}$, we will obtain an expression for computed dimension as a function of N. (Note that we have discussed r_{min} as a function of N in the case $\nu=1$ in §4.2.2 and §4.3.2.)

We, of course, expect to find that $N\to\infty$ leads to the correct dimension, but what we are after is an indication of just how large N has to be in order to achieve accurate results.

11.1 Edge effect model

We will use the edge effect as a canonical example of a systematic error that vanishes as $N\to\infty$. We have seen (in Chapter Six) that the error vanishes as $r\to 0$ and that, typically, it vanishes quadratically with r. In this section we will relate the $N\to\infty$ and the $r\to 0$ limits and then will observe the systematic

error as a function of N for large N.

Although we are going to address a very specific source of systematic error, we do not mean for the results to be taken too literally. After all, if this were an accurate description of the error in the dimension estimate, then the thing to do would not be to quote it as an error bar, but to make the appropriate correction to the estimate. In the model below, the edge effect causes the algorithm to underestimate the actual dimension, yet we have seen (§6.3) examples where the the edge effect causes an overestimate.

Our model is a Gaussian distribution of points in $\mathbb{R}^m$. The "attractor" fills out the entire embedding space, so its dimension is $\nu = m$. We have for the correlation integral (see §9.2) that

$$C(r) = \left[erf(r/2\sigma)\right]^m = \left[erf(r/2\sigma)\right]^\nu \tag{11.5}$$

or $C(r) \approx \left(\frac{r}{\sigma\sqrt{\pi}}\right)^\nu$ at small r. We measure the dimension

$$\nu(r) = \nu \frac{r}{erf(r/2\sigma)} \frac{1}{\sigma\sqrt{\pi}} e^{-r^2/2\sigma^2} \tag{11.6}$$

so

$$\frac{\nu(r)}{\nu} = \frac{r}{erf(r/2\sigma)} \frac{1}{\sigma\sqrt{\pi}} e^{-r^2/2\sigma^2}. \tag{11.7}$$

Using the Taylor series expansions for the error function and the exponential,

$$erf(r/2\sigma) = \frac{1}{\sigma\sqrt{\pi}} \left(r - \frac{r^3}{12\sigma^2} + \cdots \right) \tag{11.8}$$

and

$$e^{-r^2/2\sigma^2} = 1 - \frac{r^2}{2\sigma^2} + \cdots, \tag{11.9}$$

we can get the small r behavior of Equation (11.7):

$$\frac{\nu(r)}{\nu} = \frac{r(1-r^2/2\sigma^2+\cdots)}{r(1-r^2/12\sigma^2+\cdots)} \approx 1 - \frac{5r^2}{12\sigma^2}. \tag{11.10}$$

We note that for fixed r, the relative error is independent of the actual dimension ν.

However, we will evalutate the expression at an r that is not fixed, but depends on N and ν. In this case, with N fixed, we have

$$C(N,r) \approx C(r) = \left[erf(r/2\sigma)\right]^{\nu}. \tag{11.11}$$

However, the smallest that C(N,r) is allowed to be is $2/N^2$. This gives an implicit expression for r_{min} [1].

$$C(N,r_{min}) = \left[erf(r_{min}/2\sigma)\right]^{\nu} = \frac{2}{N^2}, \tag{11.12}$$

which gives

$$r_{min} = 2\sigma\ erf^{-1}\left[\left[2/N^2\right]^{1/\nu}\right]. \tag{11.13}$$

The relative error depends on the square of r_{min}, according to Equation (11.10).

$$\frac{\nu-\nu(r_{min})}{\nu} \approx \frac{5r_{min}^2}{12\sigma^2} = \frac{5}{3}\left[erf^{-1}\left[\left[2/N^2\right]^{1/\nu}\right]\right]^2 \tag{11.14}$$

$$\frac{\nu-\nu(N)}{\nu} \approx \frac{5}{3}\left[erf^{-1}\left[exp\left[\frac{-2\ log(N/\sqrt{2})}{\nu}\right]\right]\right]^2 \tag{11.15}$$

Figure 11.1 shows $\nu(N)$ versus ν as given by this equation. For large N, so $log\ N \gg m$, we have

$$\frac{\nu-\nu(r)}{\nu} \approx \frac{5\pi}{12}\, exp\left[-4\,\frac{log(N/\sqrt{2})}{\nu}\right] = \frac{5\pi}{12}\left[\frac{N}{\sqrt{2}}\right]^{-4/\nu}. \tag{11.16}$$

11.2 Number of points N needed to achieve specified accuracy

That the error should depend on the term $\frac{log(N/N_o)}{\nu}$ implies that for a fixed relative error, say 10%, the number of points needed to get a dimension within that error will depend exponentially on ν. That is,

$$N = N_o e^{k\nu} \tag{11.17}$$

for some k. For Equation (11.16), an error of 10% requires at least $N=1.4e^{0.64\nu}$ points to see a dimension ν. This tells us, for instance, that we would need only $\sim$1000 points to see a dimension of $\nu=10$; experience tells us that this is a serious underestimate.

This expression for relative error assumes that the slope is taken at $r=r_{min}$, an extremely optimistic assumption. In practice, the slope is taken over some appreciable range $r_{min} \le r \le r_o$ of the correlation integral, so there are enough distances that statistical error is not a problem. If, for instance, we take the slope "halfway up the curve," where $C(N,r) \approx 1/N$, then our expression for the error in ν becomes

$$\frac{\nu-\nu(N)}{\nu} \approx \frac{5\pi}{12}\, exp\left[-2\,\frac{log(N)}{\nu}\right], \tag{11.18}$$

and the 10% error will take $N=e^{1.28\nu}$, so $\nu=10$ requires over 3×10^5 points (an order of magnitude more than is usually available).

It is clear that these estimates are very sensitive to the assumptions (they are also sensitve to the values of the coefficients $\frac{5\pi}{12}$ and $\sqrt{2}$, which themselves are by no means universal and depend on the nature of the system — our model of Gaussian distribution is meant merely as an example), and for this reason their value for quantitative predicition is limited. Nonetheless, they demonstrate the exponential scaling of N with ν, which, in turn, explains why it is so difficult to analyze high-dimensional time series data.

11.3 Optimal scaling regime

We have seen above that the systematic error decreases as r_o decreases.

$$\frac{\nu-\nu(r_o)}{\nu} \approx \frac{5r_o^2}{12\sigma^2} = O(r_o^2). \tag{11.19}$$

However, as r_o decreases, fewer distances will be actually calculated in the estimation of ν, and the statistical error will increase; in Chapter Ten, we have from Equation (10.5)

$$\frac{\sigma}{\nu} = O(N^{-1}r_o^{-\nu/2}). \tag{11.20}$$

To prevent one or the other error from dominating, we choose our scaling cutoff r_o so that both errors are roughly the same magnitude: equating $O(r_o^2)=O(N^{-1}r_o^{-\nu/2})$ gives

$$r_o = O(N^{-2/(4+\nu)}) \tag{11.21}$$

the scaling with N of the optimal cutoff r_o. This gives an estimate of how the relative error scales with N:

$$\frac{\sigma}{\nu} = O(N^{-4/(4+\nu)}). \tag{11.22}$$

Including both statistical and systematic error in this sense, we see that the number of points N needed to see a dimension ν to a specified accuracy varies as

$$N \approx N_o e^{k(\nu+4)}, \tag{11.23}$$

where $k \approx -\frac{1}{4} ln(\sigma/\nu)$ increases as the desired error σ decreases.

11.4 Notes and References

[1] This estimate of r_{min} is based on the assumption that the correlation integral is smooth all the way down to $C(N,r) \approx 2/N^2$; we have seen this in highly chaotic systems, but for quasiperiodic systems, as discussed in Chapter Eight, r_{min} typically satisfies an implicit equation of the form $C(N,r_{min}) = O(1/N)$. In this case, the error will be much greater; equivalently, many more points N will be needed to achieve the same accuracy.

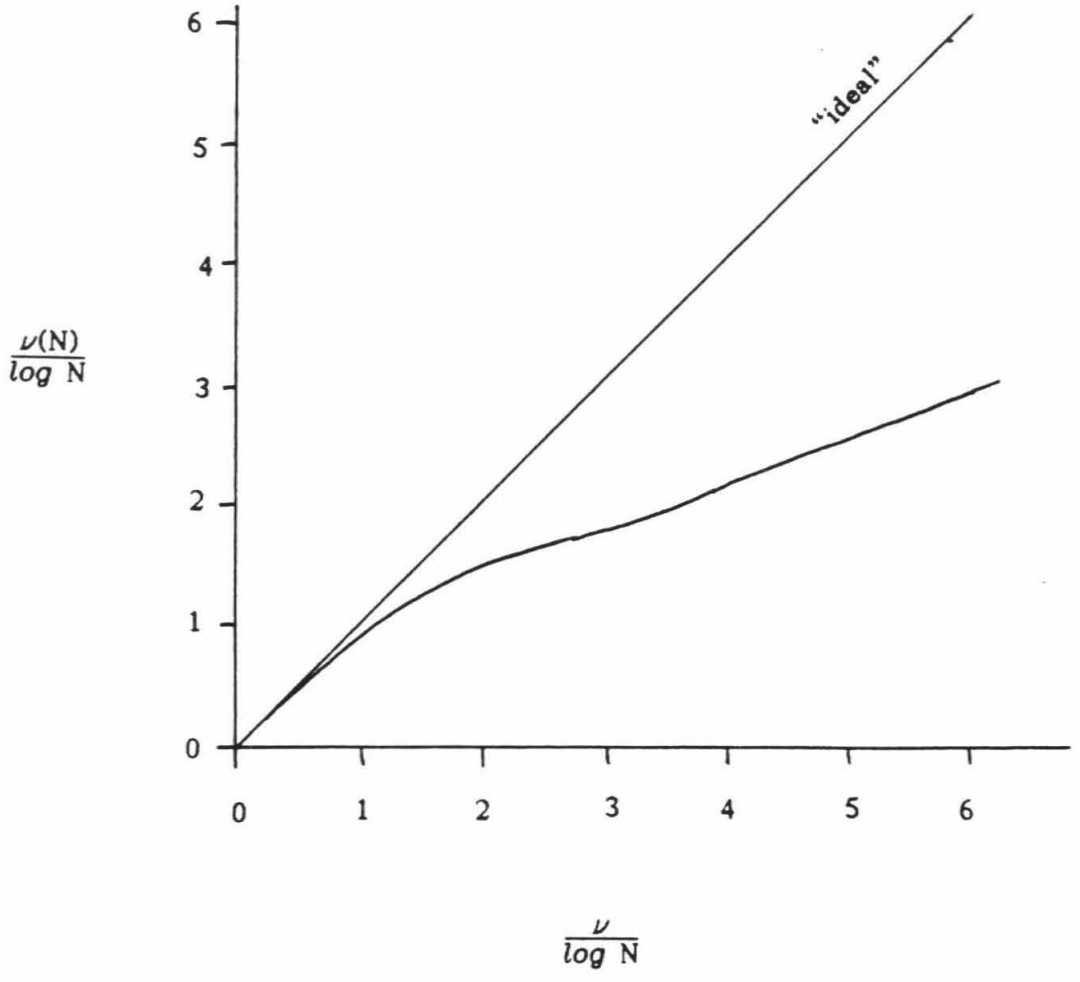


Figure 11.1 Finite N edge effect (see text for model): $\nu(N)$ versus ν, where ν is the correct dimension. $\nu = \lim_{N \to \infty} \nu(N)$. Note that both axes have been scaled by *log* N, so the curve is universal for all N.

CHAPTER TWELVE

12. COMPUTATION

Implementation of the correlation algorithm involves computing the $O(N^2)$ distances between every pair of points in a data set of N points. The operations involved in computing all these distances dominate the computation time and limit how large N can be. On the other hand, to compute dimensions accurately, N should be as large as possible. If the bottleneck is in the computation (and not, for instance, as it is for some systems, in the collection of the data), then one is motivated to improve either the hardware or the software that performs the computation.

12.1 Hardware: parallel processor

Our first approach in this direction was to enlist the aid [1] of the Caltech Hypercube Mark II, a parallel processor based on thirty-two nodes each containing an Intel 8086 microprocessor. Implementing any problem on a parallel computer involves decomposing the problem into separate tasks that can be performed simultaneously. Finding distances between pairs of points, it turns out, is a problem with a very straightforward decomposition. We will not discuss the details here but refer the reader to [2,3]. With the Hypercube, we were able to achieve a speedup very near the theoretical limit.

12.2 Software: box-assisted correlation

Recently, we devised a new algorithm [4] for computing C(N,r) more

efficiently. The idea is that although there are a total of $O(N^2)$ distances, it is only the small distances that contribute to the $r \to 0$ limit of the correlation integral. The standard algorithm computes all of the distances and then estimates a dimension from the slope of *log* C(r) versus *log* r, based only on distances less than some cutoff r_o.

The box-assisted algorithm computes *all* distances $r \leq r_o$ and *some (but not all)* of the distances $r > r_o$. It is by *not* computing all those extra distances that this algorithm is able to achieve its advantage.

Indeed, we find that for appropriately chosen r_o, we can find the smallest O(N) distances in O(N *log* N) time [5]. This can be dramatically faster than the $O(N^2)$ time that is usually required. We present an example below with N=64000 points that can be implemented on a Personal Computer; the box-assisted algorithm cuts the computation time by a factor of over a thousand.

In this procedure, points are distributed into m-dimensional "boxes" of size r_o. Then, rather than compute distances between *every* pair of points, we compute distances only between points that are either in the same box or else are in neighboring boxes. This way, we get all of the distances in the range $0 \leq r \leq r_o$. In the process, we also compute a few extra distances in the range $r_o < r \leq 2r_o$ [6], which are discarded. See Figure 12.1.

There is, to be sure, an inefficiency in these discarded distances, but it is no more inefficient than the standard algorithm that computes and discards *all* distances $r > r_o$. On the other hand, there is a certain amount of "overhead" with the box-assisted correlation algorithm: one must keep track of which points are in which box and which boxes are neighbors of each other.

The primary extra work, we find, comes from searching for boxes' neighbors. When we set up our grid of boxes we do not actually provide a separate memory location for each box. If we did, then we would find, for a typical attractor of dimension $\nu < m$, that *most* of the boxes would be empty. Instead, we have what amounts to a list of box positions. From the point of view of computer storage, empty boxes do not exist. Given the position of a box's neighbor, the searching routine must determine whether or not a box exists at that neighboring position, and if it does, which points are inside it.

In this algorithm, we associate a box position with each point. Then the points are sorted lexicographically [7] according to the position of each box. This effectively sorts our list of boxes and at the same time provides a convenient way to determine which points are in a given box. Furthermore, since the boxes are sorted in linear order, a binary search can efficiently find which if any box is at a given position.

The cost of all this extra sorting and searching is only O(N *log* N). For small r_o, therefore, the total execution time for computing a correlation dimension can be dramatically reduced.

12.2.1 Evaluation of effectiveness

In this section, we analyze the efficiency of the box-assisted correlation method and discuss in more detail the various operations that comprise the algorithm's overhead.

12.2.1.1 Run time

We divide time required to execute the box-assisted correlation computation into four components: a reading time, a sorting time, a distance computation time, and a neighbor searching time. We write

$$T_{box\text{-}assisted} = T_{read} + T_{sort} + T_{dist} + T_{search}. \quad (12.1)$$

These correspond more or less to the chronological operations of the algorithm. First, it reads all the points, next it sorts the points, and finally, for the rest of the time it switches back and forth between searching for neighboring boxes and computing distances between points in those boxes.

The reading time we can write directly,

$$T_{read} = \tau_{read} N, \quad (12.2)$$

where τ_{read} is the time to read a single point and associate a box location with it. Sorting N points can be done in O(N *log* N) time, so we write

$$T_{sort} = \tau_{sort} N \log_2 N. \quad (12.3)$$

The time spent computing distances is proportional to the number of distances D that are ultimately computed; that is,

$$T_{dist} = \tau_{dist} D, \quad (12.4)$$

which still leaves us to estimate D for this algorithm. Just as the searching routine is the most complicated part of the algorithm, estimation of T_{search} is the most difficult. For now we will write it

$$T_{search} = \tau_{search} S \cdot B, \tag{12.5}$$

where B is the number of nonempty boxes and the variable S — into which all of the complication is incorporated — is the average number of "search steps" per box.

In practice, it is these last two terms, the distance computations and the neighbor searches, which take all the time. Except for very low-dimensional systems ($m \leq 2$) and/or very small boxes $r_0 \to 0$, the reading and sorting times are comparatively negligible.

In the standard algorithm, by contrast, there is no searching and sorting, but *all* of the distances are computed, so

$$T_{standard} = \tau_{read} N + \tau_{dist} \cdot \tfrac{1}{2} N^2. \tag{12.6}$$

As long as $D \ll \frac{1}{2}N^2$ and the searching term does not dominate* [the sorting time *cannot* dominate, since it is manifestly O(N *log* N)], the running time for the box-assisted correlation algorithm will be much less than for the standard algorithm.

The coefficients τ_{read}, τ_{sort}, τ_{dist}, and τ_{search} are machine-dependent [8], though they are all of essentially the same magnitude. τ_{dist} increases

directly with embedding dimension m; τ_{sort} and τ_{search} increase more or less linearly with m but level off for large m. For our program, points are read in directly as a time series and then embedded [9] into $\mathbb{R}^m$, so τ_{read} is independent of m.

12.2.1.2 Choice of box size

The box size r_0 is the only parameter over which the user has full control in the box-assisted correlation algorithm. How D, B, and especially S depend upon r_0 is nontrivial, though we will make some estimates below.

Two attitudes can be taken toward optimum choice of box size r_0. We might for instance specify beforehand that we want all distances less than that value of r_0 that determines the scaling region. Arguing that the more distances the better the statistics [10], we say that we want to compute as many distances as possible and exclude only those beyond which the r^ν scaling fails.

The second approach chooses r_0 so that O(N) of the $\frac{1}{2}N^2$ distances are less than r_0. The standard algorithm provides the full $O(N^2)$ range of C(N,r), but it takes a time that is also $O(N^2)$. Since it is a log-log plot of C(N,r) versus r that will be ultimately constructed, we may want to optimize the logarithmic range obtained per unit of computing time; that is $(log\ D_*)/T$, where D_* is the number of distances less than r_0, and T is the time to do the computation. We will later see that

$$T = O(D_*) + O(N\ log\ N). \tag{12.7}$$

Thus, choosing $D_* = O(N)$ optimizes $(log\ D_*)/T$.

Now we can estimate $C(N,r)$ with $C(\sqrt{N},r)$; this will cost $O(N)$ and will give a range in the correlation integral of $O(N)$ — in particular, this estimates $C(N,r)$ for the larger distances. If we can compute the smallest $O(N)$ distances cheaply, in $O(N\ log\ N)$ time, say, then we will have obtained an $O(N)$ range for the $C(N,r)$ curve, which is distinct from the large-distance $O(N)$ range. These two ranges may then be "pasted together" on logarithmic axes, providing $O(N^2)$ range with significantly less than $O(N^2)$ work.

To find r_o so that N distances are less than r_o, we invoke the definition of the correlation integral to get the implicit equation

$$N = \tfrac{1}{2}N^2C(N,r_o), \tag{12.8}$$

which, when we apply the approximation $C(N,r_o) \approx (r_o/R)^{\nu}$, we can invert for r_o.

$$r_o = R(2/N)^{1/\nu} \tag{12.9}$$

This may seem a bit circular, defining r_o in terms of ν, the quantity we are ultimately after, but rough estimates of ν and R are usually available, and at any rate can be estimated from $C(\sqrt{N},r)$.

12.2.1.3 Number of distances

Of the D distances we compute, the "desirable" distances are those less than r_o. We can write this number down exactly, in terms of the correlation

integral

$$D_* = \tfrac{1}{2}N^2C(N,r_o) \approx \tfrac{1}{2}N^2(r_o/R)^\nu. \tag{12.10}$$

Now the actual D includes some distances that are greater than r_o, and so will be larger than this. We estmate the ratio of "desirable" distances to the total D by considering distances from a single (typical) point. See Figure 12.1. The number of "desirable" distances measured *from* this particular point will be proportional to $(2r_o)^\nu$, since a hypercube with diameter $2r_o$ centered on the particular point will contain all those points *to* which the distance is less than r_o. By the same token, a hypercube of diameter $3r_o$ centered not at the particular point but at the center of the box in which the particular point resides will contain all the points *to* which distances are measured. Hence, the ratio of desirable distances to total distances computed will be $(2/3)^\nu$, and

$$D = \tfrac{1}{2}N^2(3r_o/2R)^\nu. \tag{12.11}$$

For $r_o \gtrsim 2R/3$, this equation overestimates D; the number of distances calculated is never larger than $\frac{1}{2}N^2$. It follows that the time spent computing distances is

$$T_{dist} = \tau_{dist}D = (3/2)^\nu D_*, \tag{12.12}$$

which for fixed ν varies linearly with the number of desirable distances. Should we desire O(N) distances, T_{dist} will be O(N) as long as the coefficient $(3/2)^\nu \ll N$. That is,

$$\nu < log_{3/2}N = 1.7\ log_2 N. \tag{12.13}$$

In fact, as we will see later, it is not the extra distances but the increased search time that limits how large a dimension we are able to compute efficiently with the box-assisted correlation algorithm.

12.2.1.4 Number of boxes

If R is the "radius" of the attractor, then $(2R/r_o)$ estimates the number of boxes along any one dimension, and $(2R/r_o)^\nu$ [11] is the number of boxes expected to cover the attractor. Now if r_o is sufficiently large that there are many points *per* box, then $(2R/r_o)^\nu$ is a reasonable estimate of B. On the other hand, since the number of points N is finite, we expect that as r_o decreases, more and more of these covering boxes will be empty. In particular, as the boxes become ever tinier, the points will eventually come to be individually wrapped — a separate box for each point. That is,

$$\lim_{r_o \to 0} B = N. \tag{12.14}$$

If we model the distribution of points among the available boxes with the Possion formula [12], then the number of nonempty boxes is

$$B = (2R/r_o)^\nu \left[1 - e^{-N(r_o/2R)^\nu}\right]. \tag{12.15}$$

12.2.1.5 Number of neighbor searches

Unlike D and B, which depend on the geometry of the attractor, S — the average number of searching steps from each box — depends on the cleverness of our search strategy. Although this makes it difficult to give a good general estimate for S in terms of the other parameters, we can provide some upper bounds. We will provide two upper bounds in particular, each associated with a different strategy for neighbor searching. Our algorithm uses a hybrid of these two strategies, so both bounds apply. In many cases, it turns out that the actual S is much less than both bounds.

From every box, we can search each of the neighboring positions to see if there is a nonempty box at that position. Since the boxes are sorted, each search can be done in $log_2 B$ steps. If $(b_1,...,b_m)$ is the position of the "from" box, then the positions of the "to" boxes will be of the form $(b_1+\Delta b_1,..., b_m+\Delta b_m)$ where $\Delta b_i \varepsilon \{-1,0,1\}$ for i=1,..,m. Thus, there are 3^m "to" positions. We can write

$$S \leq \frac{3^m-1}{2} log_2 B \tag{12.16}$$

where the -1 is to exclude the case $\Delta b_1=\Delta b_2= \cdots =\Delta b_m=0$ (the "to" box is the same as the "from" box) and and the factor of two stems from the symmetry of distance: $d(A,B)\equiv d(B,A)$. Having found all the distances from points in Box #1 to those in Box #2, we needn't compute distances from points in Box #2 to those in Box #1.

For intermediate numbers of boxes, $3^m<B\ll N$, and with $\nu\sim m$, Equation

(12.16) not only bounds but reasonably estimates S. For low-dimensional attractors, $\nu \ll m$, a box typically will have many empty neighbors, and the effective S will be much lower than this bound.

The alternative strategy is to go through the list of boxes one at a time and to ask each of them: Are you my neighbor? There are B/2 such candidates (with the factor of two arising as above), and the binary search is avoided, so

$$S \leq \tfrac{1}{2}B. \tag{12.17}$$

This second bound provides a reasonable estimate of S only when B is very small or when ν is very large (see §12.3).

We can use these bounds on S to bound the search time. Using $B \leq N$ and Equation (12.16), we have

$$T_{search} \leq \tau_{search} \frac{3^m - 1}{2} N \log_2 N, \tag{12.18}$$

which formally is O(N *log* N). We note, however, that this "order" is sensible only if the coefficient is not too large; this is for $3^m \ll N$, or

$$m \ll \log_3 N = 1.6 \log_2 N. \tag{12.19}$$

In practice, we find the search time begins to overrun the total execution time of the standard algorithm at $m \approx 0.75 \log_2 N$. From the second bound, in Equation (12.17), we have

$$T_{search} \leq \tau_{search} \cdot \tfrac{1}{2}N^2, \qquad (12.20)$$

which shows that in the worst case the search time is $O(N^2)$. In this case, the total execution time for the box-assisted correlation method may exceed that of the standard method — this certainly is a case to be avoided! But at least the search time is never any worse than $O(N^2)$. No matter how large m or how poorly chosen r_o, the search time will never be atrociously longer than the total execution time for the standard algorithm.

12.2.1.6 Large and small box size limits

Our bound in Equation (12.17) tells us that the search time is negligible if r_o is so large that all the points fall into a single box. In that case all of the distances are computed, and all of the computation time goes into computing distances. In other words, the $r_o \gg R$ limit of box-assisted correlation does exactly what the standard algorithm does and does it for essentially the same computational cost.

We consider also the limit $r_o \to 0$. In this case, $B \to N$, and although formally both bounds on the search time are at their maximum here, the boxes are becoming sparser and more isolated from each other. Eventually, none of the boxes have any neighbors and the actual search time plummets to as low as $\tau_{search} N$. In this limit we also have $D \to 0$, so the total execution time is very small. Of course, with no distances computed, not much is learned about the attractor (this much is learned: that the smallest distance is greater than r_o), so there is not much practical benefit in this limit.

12.2.2 Implementation

A program [14] to implement this algorithm has been written in the C language and tested on an IBM PC running at 4.77 MHz with 640K RAM. We find that memory limitations [15] (not time constraints!) prevent us from processing a time series of more than 64000 points.

As an example, we compute the dimension of the Hénon attractor [16] from a sample of N=64000 points. Using r_o=0.0005 and m=2, the whole computation takes about 36 minutes. Only four of those minutes are actually spent computing distances (and of the 1.88×10^5 distances that are computed, only 1.15×10^5 are actually used); the rest of the time is spent "setting up:" it takes 7 minutes to read in and box the points, 15 minutes to sort the points, and 10 minutes to search for neighboring boxes. However inefficient this seems at first, it is still dramatically faster than the standard approach of computing all $\frac{1}{2}N^2 = 2.05\times10^9$ distances, which on our PC would take over thirty days!

This choice of r_o=0.0005 is much smaller than what might conventionally be considered the scaling regime, but it enables us to get the shortest O(N) distances computed and tabulated. As a separate computation, we can take a smaller sample of $O(\sqrt{N})$ points and get an estimate for what the "rest" of the correlation curve looks like, and again this will take only O(N) time. What we end up with, in this case, is the full $O(N^2)$ dynamic range in C(N,r) computed with O(N *log* N) work. See Figure 12.2.

12.3 Prism-assisted correlation

Because there are so many potential neighbors ($\sim 3^m$) in systems of large embedding dimension, we find that the search time increases very rapidly (though not quite exponentially — it is limited ultimately by the bound in Equation (12.20)) with increasing m. Empirically, we find with random data [17] that the box-assisted algorithm becomes worse than the standard algorithm for embedding dimensions larger than $m \approx 0.75\ log_2 N$.

However, we have devised a variant of this algorithm, which gets around the large m limitation. In the m-dimensional space we impose a b-dimensional grid where b is less than m. The "boxes" in this space are m-dimensional rectangular prisms with b short sides of length r_o and $m-b$ long sides, which extend the entire length of the attractor. As before, we compute distances between pairs of points only if those points are in the same or in adjoining prisms. Note that b=m is just the regular box-assisted algorithm and b=0 corresponds to the standard algorithm.

With $b<m$, there are fewer neighbors, and the coefficient of the search time looks like 3^b instead of 3^m. We can take m as large as we like, and the search time will not increase.

On the other hand, for fixed m, a smaller value of b means that more distances are computed. This is because of the distances we have to compute between points at opposite "ends" of these long prisms. Following Equation (12.11), we have

$$D_{prism} = \tfrac{1}{2}N^2(3r_o/2R)^b. \qquad (12.21)$$

distances to compute. Let us take r_o, as usual, so that O(N) desirable distances are computed. We substitute Equation (12.9) into the above to get

$$D_{prism} = \tfrac{1}{2}N^2\left[\tfrac{2}{3}(N/2)^{1/\nu}\right]^{-b} \tag{12.22}$$

In choosing the best value for b, we have these two competing effects. The number of distances computed *decreases* with increasing b, and the number of neighbors the program has to search for *increases*. We can estimate the sum of the two times,

$$T_{prism} = \tau_{dist}\cdot\tfrac{1}{2}N^2\left[\tfrac{2}{3}(N/2)^{1/\nu}\right]^{-b} + \tau_{search}\frac{3^b-1}{2}\,B\,log_2 B, \tag{12.23}$$

noting that the term for search time is valid only for $b<\nu$. Formally, we can optimize by setting $\partial T/\partial b=0$. The resulting expression is quite unwieldy, but in the limit of large N and large ν ($\gg log$ N), we have

$$b = \frac{log\ N}{log\ (9/2)} \approx 0.5\ log_2 N. \tag{12.24}$$

Numerical experiments with random data confirm this estimate. See Figure 12.3.

12.4 Notes and References

[1] We are indebted to Professor Geoffrey C. Fox and the Caltech Concurrent Computing Project (C^3P) for assistance and use of their machine.

[2] G. C. Fox, M. Johnson, G. Lyzenga, S. Otto, J. Salmon, and D. W. Walker. *Solving Problems on Concurrent Processors* (Prentice-Hall, Englewood Cliffs, New Jersey, 1987).

[3] James Theiler. "Strange attractors and turbulence," Caltech Concurrent Computation Program (C^3P) Technical Bulletin 56.43, 56.44 (1985) 1:11.

[4] James Theiler. "Efficient algorithm for estimating the correlation dimension from a set of discrete points," Phys. Rev. A *(in press)*.

[5] A general discussion of related problems from a computer science point of view can be found in Jon Louis Bentley. "Multidimensional Divide-and-Conquer," Comm. ACM 23 (1980) 214. We are indebted to J. D. Farmer for this reference.

[6] This assumes the L_∞ or "maximum" metric. For the L_1 or "taxicab" metric, the appropriate inequality is $r \leq 2mr_o$; for the L_2 or Euclidean metric, it is $r \leq 2\sqrt{m} \cdot r_o$.

[7] In the "lexicographic" ordering, we say that $\mathbf{a} < \mathbf{b}$, where in coordinate notation $\mathbf{a} = (a_1, a_2, \ldots, a_m)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_m)$, if and only if $a_k < b_k$ for some $k \leq m$, and $a_i = b_i \ \forall \ i < k$.

[8] On an IBM PC, we find $\tau_{dist} = (0.5 + 0.2m)$ msec. In the Takens method, we also have to take the logarithm of all distances less than r_o; these cost 4.3 msec each (with an 8087 floating point coprocessor). Also $\tau_{search} \approx (1.5 + 0.6m)$ msec and $\tau_{sort} = (0.44 + 0.09m)$ msec, with some levelling off at large m. Finally, τ_{read} = 6.7 msec. There is a small memory compiler option that cuts these times in half, but it can be used only for $N < 5000$.

[9] Floris Takens. "Detecting Strange Attractors in Turbulence," in *Dynamical Systems and Turbulence, Warwick*, 1980, Vol. 898 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1981).

[10] The error bar on ν will scale as $1/\sqrt{D_*}$, where D_* is the number of distances less than r_o. Floris Takens. "On the numerical determination of the dimension of an attractor," in *Dynamical Systems and Bifurcations, Groningen*, 1984, Vol. 1125 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1985).

[11] Properly, we should write $B \approx (2R/r_o)^d$, where d is the "capacity" of the set. Though capacity and correlation dimension are not precisely the same thing, our approximations do not distinguish between them.

[12] What we really assume is that points are distributed uniformly among those boxes that cover the attractor; this presumes not only that the distribution of points is uniform over the attractor, but that the intersection of the attractor with boxes is uniform — in fact, there are often a lot of "clipped edges." For details of this effect in another context, see W. E. Caswell and J. A. Yorke. "Invisible Errors in Dimension Calculations: Geometric and Systematic Effects," in *Dimensions and Entropies in Chaotic Systems*, ed. G. Mayer-Kress (Springer, Berlin, 1986).

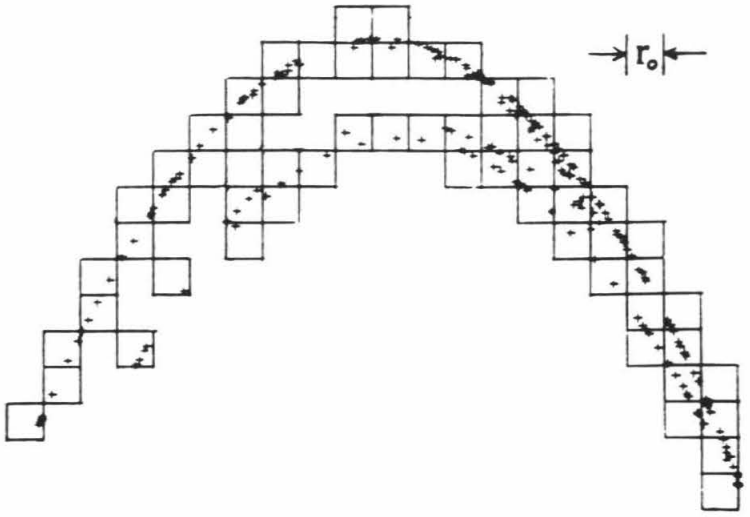
[13] The source code, further documentation, and executable files, which run on an IBM PC (or compatible), are available from the author.

[14] In our Grassberger-Procaccia routine, we use $8N + 4xr_o$ bytes, where x is the "expansion factor," which is multiplied by the floating point input before discretizing into integers. In our Takens maximum likelihood routine, we use 14N bytes, since the floating point input is stored in double precision.

[15] M. Hénon. "A two-dimensional mapping with a strange attractor," Comm. Math. Phys. <u>50</u> (1976) 69. The mapping is given by $x_{i+1}=y_i+1-ax_i^2$; $y_{i+1}=bx_i$. Following Hénon, we use a=1.4, b=0.3.

[16] For low-dimensional attractor data, we find that the search time increases so slowly with m that O(N) distances can *always* be computed faster with the box-assisted algorithm than with the standard algorithm. But even in these cases, we find that the "prism-assisted" variant provides further improvement.

(a)



(b)

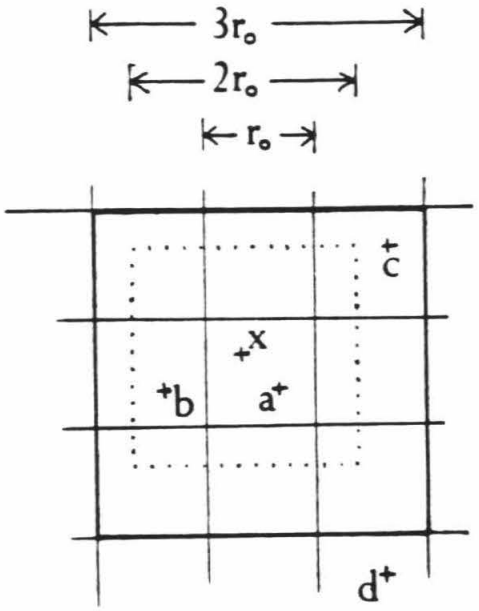


Figure 12.1 (a) A square grid of boxes of width r_o is placed over the attractor points. (b) Here, distances are computed from point x to other points on the attractor. Distances are computed to points that are in the same (*e.g.*, point a) or in adjacent boxes (*e.g.*, points b,c). If any of those distances are greater than r_o (*e.g.*, to point c), then those distances are discarded. Points (*e.g.*, point d) not in adjacent boxes are not considered. In other words, distances to points (a,b,c, not d) inside the box are computed, but only those (a,b, not c) inside the dotted box are used.

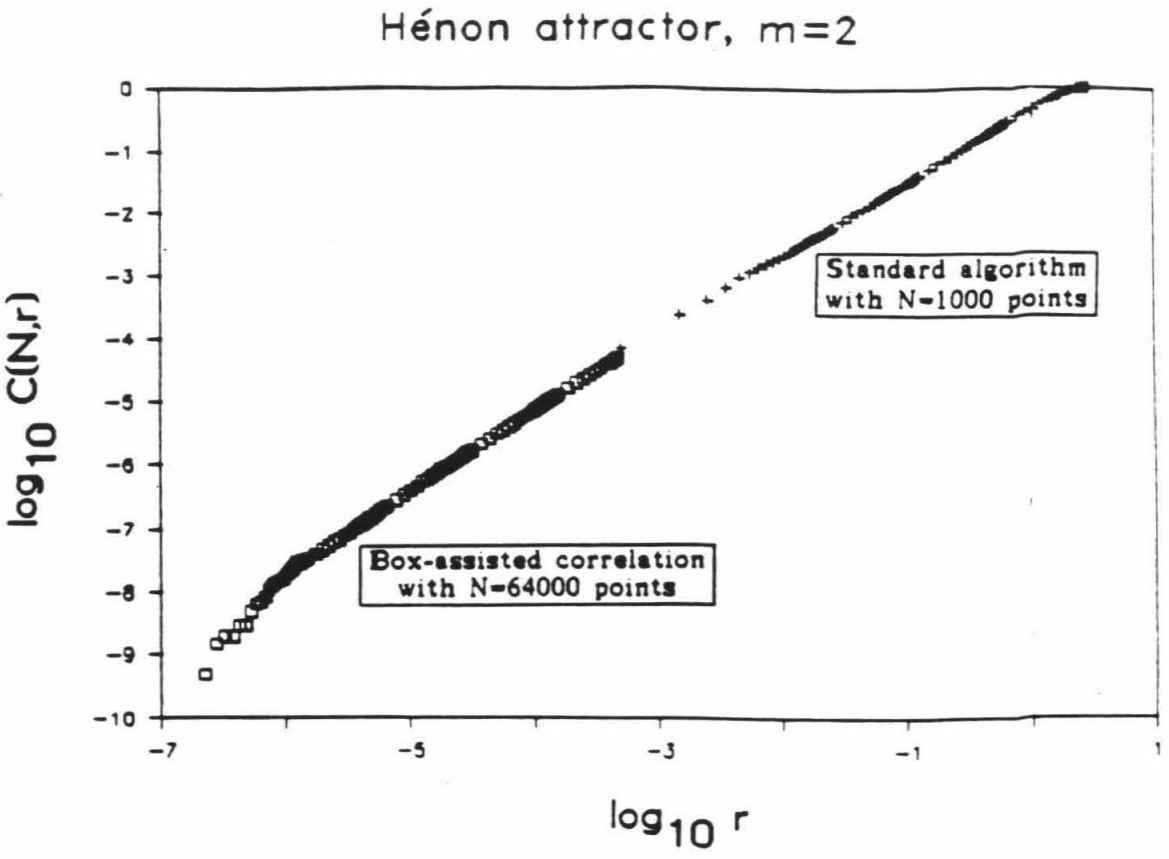


Figure 12.2 Log-log plot of the correlation integral for the Hénon attractor. The small distances in the lower half (□) of the curve were computed with $N=64000$ points, using a box size of $r_0=0.0005$. The upper half (+) was computed with a much smaller sample of $N=1000$ points and a box size so large ($r_0=3.0$) that all the distances were computed. Both computations together took less than an hour on a personal computer. To compute the entire curve in one piece with the standard algorithm would have taken over a month.

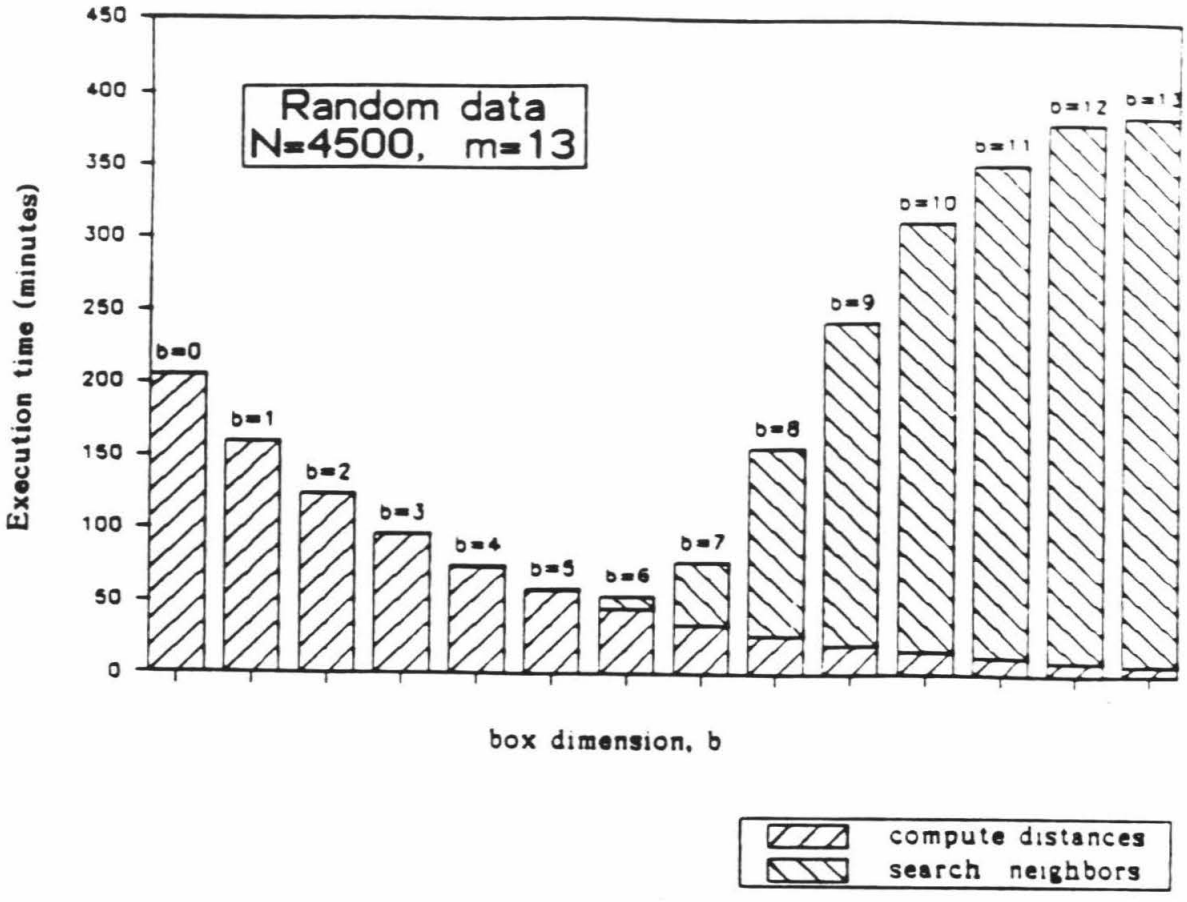


Figure 12.3 Prism-assisted correlation for N=4500 points of random data (uniformly distributed over $[-1,1]^m \subset \mathbf{R}^m$) embedded in m=13 space with r_o=0.66, chosen so that ~4500 "desirable" distances are computed. Taking b=m corresponds to the usual box-assisted algorithm, and b=0 is the standard unassisted algorithm. Although this is a case where the standard (b=0) algorithm is better than the box-assisted (b=m) method, we find that we do get improvement from the "prism-assisted" correlation algorithm with $1 \leq b \leq 8$ and that the best performance is achieved at $b=6 \approx 0.5 \log_2 4500$. The contribution to total execution time due to initial reading and sorting is about one minute, a negligible amount.

CHAPTER THIRTEEN

13. LINEAR ANALYSIS OF TIME SERIES

In this chapter, we will introduce some of the more conventional methods of time series analysis and will discuss their potential applicability to analyzing the time series of nonlinear dynamical systems. These methods are fundamentally linear, and in most cases have been known for some time. The results reported here are from Robinson's book [1], though much of the work was originally done in the early 1940's by Kolmogorov and Wiener (independently of each other).

Linear analysis offers two hopes: one, as a direct means of inferring properties of the system from its time series; and two, as an enhancement to nonlinear methods of analysis, such as the correlation integral.

Some of the first excitement in the field of nonlinear dynamics began when researchers noticed broadband Fourier spectra from deterministic chaos, something formerly associated with nondeterministic noise. In general, it is not possible [2] to distinguish deterministic from nondeterministic systems directly from the power spectrum — it is for this purpose that *non*linear methods (such as correlation dimension) were developed.

Directly from the power spectrum, however, we *will* be able to distinguish systems which, on the one hand, are nonchaotic and deterministic from systems which, on the other hand, are either chaotic or stochastic; in the vocabulary of fluid dynamics: we can distinguish laminar from turbulent flow.

Linear methods can be useful enhancements to the nonlinear methods of

analysis such as the dimension algorithm. Among these enhancements is a modification of the usual embedding procedure so that the axes are uncorrelated with each other. We propose further enhancements based on the decomposition of time series into linear and "purely nonlinear" components. The decomposition allows us, for instance, to filter out the linear component, and apply the nonlinear analysis just to the nonlinear component. Though one might imagine that this provides a more efficient analysis, experience suggests the opposite. A more successful application of the decompositon is the creation of a "linearly equivalent" time series that has the same linear component as the original time series, but its nonlinear component is generated stochastically. We can use the new time series as a benchmark against which the original (possibly deterministic — we do not always know *a priori*) time series may be compared.

13.1 Autocorrelation and Fourier spectrum

Among the immediate characteristics of a time series, we can write the mean μ, and standard deviation σ, defined by

$$\mu = \langle x_n \rangle, \tag{13.1}$$

$$\sigma^2 = \langle (x_n - \mu)^2 \rangle, \tag{13.2}$$

where the angle brackets, $\langle\rangle$, denote the average over n. The autocorrelation is defined

$$A(\tau) \equiv \langle (x_n - \mu)(x_{n+\tau} - \mu) \rangle = \langle x_n x_{n+\tau} \rangle - \mu^2. \tag{13.3}$$

Note that this implies that the autocorrelation is even; that is, $A(-\tau)=A(\tau)$. It is a common practice, *not* adopted here, to normalize the autocorrelation, dividing $A(\tau)$ by σ^2, so that $A(0)=1$. On the other hand, much of the analysis will be with "standardized" time series, for which $\langle x_n \rangle = 0$ and $\langle x_n^2 \rangle = 1$. In this case, $A(0)=1$.

We define the Fourier transform, a continuous periodic function with a period of 2π,

$$X(\omega) = \sum_{n=-\infty}^{+\infty} x_n e^{-i\omega n}. \tag{13.4}$$

It is invertible, of course;

$$x_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega)\, e^{i\omega n}\, d\omega. \tag{13.5}$$

The Wiener-Khintchine relations for a discrete time series allow us to write the Fourier spectrum in terms of the autocorrelation

$$|X(\omega)|^2 = A(0) + 2 \sum_{\tau=1}^{\infty} A(\tau) \cos(\tau\omega), \tag{13.6}$$

and vice-versa,

$$A(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\omega)|^2 \cos(\tau\omega)\, d\omega. \tag{13.7}$$

13.2 White noise

White noise is an important example of a purely stochastic system. In a white noise time series $\{\epsilon_n\}$, each ϵ is chosen *independently* from some random distribution. It follows from this independence that distinct elements of the time series will be *uncorrelated*; that is, $\langle\epsilon_r\epsilon_s\rangle=\langle\epsilon_r\rangle\langle\epsilon_s\rangle$ for $r\neq s$ [3].

Standard white noise has mean zero and standard deviation unity. Since it is uncorrelated, $\langle\epsilon_r\epsilon_s\rangle = \delta_{rs}$. Thus, $A(0)=1$ and $A(\tau)=0$ for $\tau\geq 1$. The Fourier spectrum in this case $|X(\omega)|^2=1$ is said to be flat (or "white") since it has no dependence on ω. We note that predicting the future of a time series of white noise is essentially hopeless; successive values of ϵ depend in no way upon previous values.

13.3 Prediction

There is an obvious motivation for being able to predict x_{n+1} from time series history $x_n, x_{n-1}, x_{n-2}, \ldots$. Methods of nonlinear prediction are still being developed [4,5,6], but linear prediction is a very straightforward technique. Furthermore, we will find the language and results of prediciton theory useful in other applications of linear methods.

13.3.1 Linear prediction

In linear prediction theory, we estimate x_{n+1} with $\hat{x}^{(m)}_{n+1}$, the linear combination of the previous m values of the time series (we assume $\mu=0$).

$$\hat{x}^{(m)}_{n+1} = k_0 x_n + k_1 x_{n-1} + k_2 x_{n-2} + \cdots + k_{m-1} x_{n-m+1}, \tag{13.8}$$

where the coefficients k_0, k_1, k_2, ..., k_{m-1} are chosen to minimize the mean-square deviation of x_{n+1} from its predictor. Here

$$\chi_m^2 \equiv \left\langle (x_{n+1} - \hat{x}_{n+1}^{(m)})^2 \right\rangle \tag{13.9}$$

measures that deviation. The more values of k that we are free to introduce, the better our predictor can be. Thus, χ_m^2 decreases monotonically with m. On the other hand, χ_m^2 is bounded from below by zero (it is by definition positive). Thus, we can define the limit

$$\chi^2 \equiv \lim_{m \to \infty} \chi_m^2 \tag{13.10}$$

as a measure of the "linear unpredictability" of the time series. If $\chi^2 = 0$, then the system is completely predictable. Such time series we classify as linearly deterministic, and included in this class are periodic and quasiperiodic systems, though the class is not limited to these types. However, there is no reason to expect generic nonlinear deterministic dynamical systems to be linearly deterministic.

Nonlinear deterministic systems are those for which a function f exists, which satisfies

$$x_{n+1} = f(x_n, x_{n-1}, x_{n-2}, \ldots), \tag{13.11}$$

but the function is not linear in its arguments. Attempting to predict the future from a linear function such as Equation (13.8) leads to $\chi^2 > 0$. Linear analysis

cannot predict the future exactly, no matter how many terms are taken. The nonlinear deterministic system, *from the point of view of the linear analysis,* has a component that is linearly predictable and a component that is like white noise; this second component is the "purely nonlinear component" that we discuss further in §13.6.

13.3.2 Least Squares

We begin by computing the linear coefficients $k_0,\ldots,k_{m-1}$ in terms of the autocorrelation. We want the miminum value of χ^2_m, so take partial derivatives with respect to each of the parameters k_j for $j=0,\ldots,m-1$ and set these to zero. Thus,

$$0 = \frac{\partial}{\partial k_j} \left\langle (x_{n+1} - \hat{x}^{(m)}_{n+1})^2 \right\rangle \tag{13.12}$$

$$= \frac{\partial}{\partial k_j} \left\langle \left[x_{n+1} - \sum_{i=0}^{m-1} k_i x_{n-i} \right]^2 \right\rangle. \tag{13.13}$$

Expand the square.

$$= \frac{\partial}{\partial k_j} \left\langle x^2_{n+1} - 2x_{n+1} \sum_{i=0}^{m-1} k_i x_{n-i} + \left[\sum_{i=0}^{m-1} k_i x_{n-i} \right]^2 \right\rangle \tag{13.14}$$

Bring the averages "inside" the summation

$$= \frac{\partial}{\partial k_j} \left[< x^2_{n+1} > -2 \sum_{i=0}^{m-1} k_i <x_{n-i} x_{n+1}> + \sum_{r=0}^{m-1} \sum_{s=0}^{m-1} k_r k_s <x_{n-r} x_{n-s}> \right]. \tag{13.15}$$

Use the definition of autocorrelation $A(\tau)$.

$$= \frac{\partial}{\partial k_j}\left[A(0) - 2\sum_{i=0}^{m-1} k_i A(i+1) + \sum_{r=0}^{m-1}\sum_{s=0}^{m-1} k_r k_s\, A(r-s)\right] \tag{13.16}$$

Take the partial derivative. Use $\partial k_i/\partial k_j = \delta_{ij}$.

$$= \left[-2\sum_{i=0}^{m-1} \delta_{ij} A(i+1) + \sum_{r=0}^{m-1}\sum_{s=0}^{m-1} (\delta_{rj} k_s + k_r \delta_{sj})\, A(r-s)\right] \tag{13.17}$$

$$= -2A(j+1) + \sum_{s=0}^{m-1} k_s A(j-s) + \sum_{r=0}^{m-1} k_r A(r-j) \tag{13.18}$$

Combine the sums, using $A(-\tau)=A(\tau)$, to get an equation that holds for $j=0,\ldots,m-1$.

$$0 = -2A(j+1) + 2\sum_{i=0}^{m-1} k_i A(j-i) \tag{13.19}$$

We introduce the **covariance matrix** $\Xi^{(m)}$, defined in terms of the autocorrelation. We have $\Xi^{(m)}_{ij} = \Xi^{(m)}_{ji} = A(i\text{-}j) = A(|i-j|)$; that is,

$$\Xi^{(m)} = \begin{bmatrix} A(0) & A(1) & \cdots & A(m-1) \\ A(1) & A(0) & \cdots & A(m-2) \\ \vdots & \vdots & \ddots & \vdots \\ A(m-1) & A(m-2) & \cdots & A(0) \end{bmatrix} \tag{13.20}$$

Elements along any subdiagonal are identical, so this is a Toeplitz matrix [7]. Though there are m^2 entries in the matrix, only m of them are distinct. The equations obtained above for the k_j's can be expressed in terms of the covariance matrix.

$$\begin{bmatrix} A(0) & A(1) & \cdots & A(m-1) \\ A(1) & A(0) & \cdots & A(m-2) \\ \vdots & \vdots & \ddots & \vdots \\ A(m-1) & A(m-2) & \cdots & A(0) \end{bmatrix} \begin{bmatrix} k_0 \\ k_1 \\ \vdots \\ k_{m-1} \end{bmatrix} = \begin{bmatrix} A(1) \\ A(2) \\ \vdots \\ A(m) \end{bmatrix} \tag{13.21}$$

Having solved the matrix equation, we can compute our "goodness" parameter,

$$\chi_m^2 = \left\langle \left[x_{n+1} - \sum_{i=0}^{m-1} k_i x_{n-i} \right]^2 \right\rangle. \tag{13.22}$$

$$= A(0) - 2\sum_{i=0}^{m-1} k_i A(i+1) + \sum_{r=0}^{m-1}\sum_{s=0}^{m-1} k_r k_s A(r-s). \tag{13.23}$$

$$= A(0) - \sum_{i=0}^{m-1} k_i A(i+1) . \tag{13.24}$$

Kác [8] gives the more compact formula in terms of the determinants of the covariance matrices,

$$\chi_m^2 = \det(\Xi^{(m)})/\det(\Xi^{(m-1)}). \tag{13.25}$$

The information about whether or not our system is linearly predictable is all contained in the autocorrelation values. Because they are Fourier transforms of each other, that information is also in the power spectrum. In fact, there is an explicit form, due to Grenander and Szegö [9], for the limit as $m \to \infty$ of χ_m^2:

$$\chi^2 = exp\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} ln\left(|X(\omega)|^2 \right) d\omega \right]. \tag{13.26}$$

This formula provides a measure of linear unpredictability directly from the Fourier power spectrum. If $\chi^2>0$, then the system is linearly unpredictable, and we call it "turbulent." If $\chi^2=0$, then the system is linearly deterministic, and we call it "laminar."

The condition for linear unpredictability is that the Fourier spectrum be almost everywhere nonzero. For example, systems with quasiperiodic time series, the Fourier spectra of which are sums of delta functions, are linearly predictable. In the next section, we discuss the linear predictability of a number of examples.

13.4 Linear predictability: some examples

Here are some deterministic and nondeterministic systems. We will measure and discuss their linear predictability χ^2. We'll find that $\chi^2=0$ for nonchaotic deterministic (laminar) systems and that $\chi^2>0$ for chaotic or stochastic (turbulent) systems.

13.4.1 Periodic (sinusoidal)

Consider the sinusoidal time series with unit amplitude and zero initial phase.

$$x_n = \sin(n\phi) \tag{13.27}$$

We note that this prescription is not manifestly time-independent. Although we

can recast the equation in deterministic nonlinear form

$$x_{n+1} = \sin\left[\sin^{-1}(x_n) + \phi\right], \tag{13.28}$$

we note that there is an ambiguity in the inverse sine. It turns out that we can also write this as a deterministic linear equation in terms of the previous two time series values:

$$\begin{aligned}
x_{n+1} &= \sin((n+1)\phi) && (13.29)\\
&= \sin(n\phi)\cos(\phi) + \sin(\phi)\cos(n\phi) && (13.30)\\
&= \sin(n\phi)\cos(\phi) + \sin(\phi)\left[\cos((n-1)\phi + \phi)\right] && (13.31)\\
&= \sin(n\phi)\cos(\phi) + \sin(\phi)\left[\cos((n-1)\phi)\cos(\phi) - \sin((n-1)\phi)\sin(\phi)\right] && (13.32)\\
&= \sin(n\phi)\cos(\phi) + \sin(\phi)\left[\cos((n-1)\phi)\cos(\phi) - \sin((n-1)\phi)\sin(\phi)\right]. && (13.33)
\end{aligned}$$

But we have from Equation (13.30) that

$$\cos(n\phi) = \frac{x_{n+1} - \sin(n\phi)\cos(\phi)}{\sin(\phi)} \tag{13.34}$$

thus,

$$\cos((n-1)\phi) = \frac{x_n - \sin((n-1)\phi)\cos(\phi)}{\sin(\phi)}, \tag{13.35}$$

so we can finally write

$$\begin{aligned}
x_{n+1} &= \sin(n\phi)\cos(\phi) + \\
&\quad \sin(\phi)\left[\frac{x_n - \sin((n-1)\phi)\cos(\phi)}{\sin(\phi)}\cos(\phi) - \sin((n-1)\phi)\sin(\phi)\right] && (13.36)\\
&= x_n\cos(\phi) + \sin(\phi)\left[\frac{x_n - x_{n-1}\cos(\phi)}{\sin(\phi)}\cos(\phi) - x_{n-1}\sin(\phi)\right] && (13.37)
\end{aligned}$$

$$= [2\cos(\phi)]\, x_n - x_{n-1}. \tag{13.38}$$

Periodic sinusoidal time series are linearly determinisitic, and depend only on the previous two values of the time series. Thus $\chi^2_m=0$ for all $m\geq 2$, and in particular, in the limit $m\to\infty$, we have $\chi^2=0$. Since the dependence is on *two* previous values, we can, given starting values x_0 and x_1, produce the entire time series. These two free parameters, it turns out, specify the amplitude and phase of the sine wave. Thus Equation (13.38) is in fact more general than (13.27).

13.4.2 Quasiperiodic (sinusoidal)

Here, consider the sum of two sinusoidal time series. There are four free parameters, two amplitudes and two initial phases (we do not think of the frequencies as free parameters since they are fixed by the system).

$$x_n = A_1\sin(n\phi_1+\Phi_1) + A_2\sin(n\phi_2+\Phi_2) \tag{13.39}$$

Again, this formulation is not manifestly time-invariant. However, we can write

$$\begin{aligned} x_{n+1} &= [2\cos(\phi_1) + 2\cos(\phi_2)]\, x_n \\ &\quad - [2 + 4\cos(\phi_1)\cos(\phi_2)]\, x_{n-1} \\ &\quad + [2\cos(\phi_1) + 2\cos(\phi_2)]\, x_{n-2} \\ &\quad - x_{n-3}. \end{aligned} \tag{13.40}$$

In general, a sum of N sine waves may be combined into a single linear difference

equation with 2N terms. And again, $\chi^2=0$, so the system is linearly deterministic.

13.4.3 Periodic (not sinusoidal)

Consider the following map:

$$x_{n+1} = x_n + \phi \quad \text{(modulo 1)}. \tag{13.41}$$

(This is the twist map that was introduced in §8.3.) But for the modulo 1, this is already linear. However, the modulo operator is decidedly nonlinear (it is not even continuous). To compare it to the sinusoidal case, we note that we can write this time series also as

$$x_n = x_0 + n\phi \quad \text{(modulo 1)} \tag{13.42}$$

$$= f(x_0 + n\phi), \tag{13.43}$$

where the function f is the fractional part. This is not the same as the sine function, but it is periodic and we know that we can express it as a sum of sine waves. (There are a few details: the function f has a nonzero offset, and a different period than the sine waves in the previous examples, but these are only details.) We might expect, therefore, that with enough terms (to account for enough of the sine wave components), we can write an arbitrarily precise predictor; that is, that $\chi^2=0$. And this is true, but we will use another argument to verify that $\chi^2=0$.

Given $\epsilon>0$, choose some N for which the fractional part of $N\phi$ is less

than ϵ. It is not hard to see that this can always be done (for instance, take N to be the denominator of a continued fraction expansion of ϕ — see §8.3.2). Then, take as a predictor

$$\hat{x}^{(N)}_{n+1} = x_{n-N+1}. \tag{13.44}$$

This is probably not the *best* predictor, but its deviation from the actual x_{n+1} will be less than ϵ, and in particular, $\chi^2_N<\epsilon^2$. Thus, we can see that $\chi^2=0$.

13.4.4 Feigenbaum attractor [10]

Take the logistic map

$$x_{n+1} = \lambda x_n(1-x_n) \tag{13.45}$$

at $\lambda=\lambda_{onset}=3.5699...$, which is the limit point of the bifurcation cascade. The time series is aperiodic but it is not chaotic; that is, nearby trajectories do not diverge.

In fact, every orbit on the attractor is shadowed by (unstable) orbits of period $N=2^K$. That is, for any orbit $\{x_0,...\}$ and any $\epsilon>0$, there is an orbit $\{x'_0,...\}$ of period 2^K for some K and with $|x'_n-x_n|<\epsilon$ for all n. Now we can construct predictors of the form

$$\hat{x}^{(N)}_{n+1} = x_{n-N+1} \tag{13.46}$$

with $N=2^K$, which will be at most ϵ away from the actual value; thus, $\chi^2_m<\epsilon^2$

and

$$\lim_{m\to\infty} \chi^2_m = 0. \tag{13.47}$$

The system is linearly determinsitic.

13.4.5 Damped random walk

This time, we consider a system that is manifestly nondeterminsitic. White noise has been added to what would otherwise be a linear map.

$$x_{n+1} = \lambda x_n + \gamma\epsilon_{n+1} \tag{13.48}$$

Here, $\lambda<1$ is the damping factor. Without the noise term, the time series would damp to zero. We have

$$\langle x\rangle = \frac{\gamma\langle\epsilon\rangle}{1-\lambda} = 0 \tag{13.49}$$

and

$$\langle x^2\rangle = \frac{\gamma^2\langle\epsilon^2\rangle}{1-\lambda^2} = \frac{\gamma^2}{1-\lambda^2}\,. \tag{13.50}$$

Further,

$$\langle x_n x_{n+1}\rangle = \lambda\langle x_n^2\rangle + \gamma\langle x_n\epsilon_{n+1}\rangle \tag{13.51}$$

$$= \lambda\langle x_n^2\rangle \tag{13.52}$$

implies

$$A(1) = \lambda A(0). \tag{13.53}$$

More generally,

$$x_{n+\tau} = \lambda x_{n+\tau-1} + \epsilon_{n+\tau} \tag{13.54}$$

$$x_n x_{n+\tau} = \lambda x_n x_{n+\tau-1} + x_n \epsilon_{n+\tau} \tag{13.55}$$

$$\langle x_n x_{n+\tau} \rangle = \lambda \langle x_n x_{n+\tau-1} \rangle + \langle x_n \epsilon_{n+\tau} \rangle, \tag{13.56}$$

and since the noise at time $n+\tau$ is uncorrelated with the value of x at time n, $\langle x_n \epsilon_{n+\tau} \rangle = 0$ and

$$\langle x_n x_{n+\tau} \rangle = \lambda \langle x_n x_{n+\tau-1} \rangle, \tag{13.57}$$

so that

$$A(\tau) = \lambda A(\tau - 1) = \lambda^{\tau} A(0). \tag{13.58}$$

The autocorrelation decreases exponentially with τ. Our best linear estimate in this case can be shown to be $k_0 = \lambda$, $k_1 = k_2 = \cdots = 0$. And this leads to

$$\chi^2 = A(0)(1 - \lambda^2) = \gamma^2 \neq 0. \tag{13.59}$$

Our conclusion that this is not linearly predictable comes as no surprise, for we have directly and deliberately added noise to the system. The future is not completely predictable *by any means*.

13.4.6 Positive expansion rate

Every one of the deterministic examples so far have been nonchaotic; nearby trajectories did not diverge. And each turned out as well to be linearly deterministic. In the example in this section, we take

$$x_{n+1} = \eta x_n \qquad \text{(modulo 1).} \tag{13.60}$$

If $\eta > 1$, we have a positive expansion rate, nearby trajectories diverge, and chaos ensues. (If $\eta < 1$, then the time series converges, $x_n \to 0$ as $n \to \infty$.)

Another way to write this rule is

$$x_n = f(\eta^n x_0), \tag{13.61}$$

where the function f is the fractional part. In this example, we will take η to be an integer.

We will compute the autocorrelation function,

$$A(\tau) \equiv \langle (x_{n+\tau} - \mu)(x_n - \mu) \rangle = \langle x_{n+\tau} x_n \rangle - \mu^2, \tag{13.62}$$

noting the need to subtract the mean since $\mu \neq 0$. Further, instead of averaging over n, we will average over $x \in [0,1)$. This is valid if the orbit fills the interval uniformly. It can be shown that this is the case for generic orbits [11] as long as η is an integer. Thus,

$$\langle x_{n+\tau} x_n \rangle = \int_0^1 f(\eta^\tau x)\, x\, dx \tag{13.63}$$

$$= \int_0^{1/\eta^\tau} f(\eta^\tau x)\, x\, dx + \int_{1/\eta^\tau}^{2/\eta^\tau} f(\eta^\tau x)\, x\, dx + \cdots + \int_{(\eta^\tau - 1)/\eta^\tau}^{1} f(\eta^\tau x)\, x\, dx \tag{13.64}$$

$$= \sum_{k=0}^{\eta^\tau - 1} \left[\int_{k/\eta^\tau}^{(k+1)/\eta^\tau} f(\eta^\tau x)\ x\ dx \right] \tag{13.65}$$

$$= \sum_{k=0}^{\eta^\tau - 1} \left[\int_{k/\eta^\tau}^{(k+1)/\eta^\tau} (\eta^\tau x - k)\ x\ dx \right] \tag{13.66}$$

$$= \int_0^1 (\eta^\tau x)\ x\ dx + \sum_{k=0}^{\eta^\tau - 1} \left[\int_{k/\eta^\tau}^{(k+1)/\eta^\tau} -k\ x\ dx \right] \tag{13.67}$$

$$= \frac{\eta^\tau}{3} + \sum_{k=0}^{\eta^\tau - 1} -k \left[\frac{(k+1)^2}{2\eta^{2\tau}} - \frac{k^2}{2\eta^{2\tau}} \right] \tag{13.68}$$

$$= \frac{\eta^\tau}{3} - \frac{1}{2\eta^{2\tau}} \left[\sum_{k=0}^{\eta^\tau - 1} 2k^2 + k \right] \tag{13.69}$$

$$= \frac{\eta^\tau}{3} - \frac{1}{2\eta^{2\tau}} \left[\frac{\eta^\tau(\eta^\tau - 1)(2\eta^\tau - 1)}{3} + \frac{\eta^\tau(\eta^\tau - 1)}{2} \right] \tag{13.70}$$

$$= \frac{\eta^\tau}{3} - \frac{1}{2\eta^\tau} \left[\frac{4\eta^{2\tau} - 3\eta^\tau - 1}{6} \right] \tag{13.71}$$

$$= \frac{3\eta^{2\tau} + \eta^\tau}{12\eta^{2\tau}} = \frac{1}{4} + \frac{1}{12\eta^\tau}. \tag{13.72}$$

Now, $\mu = \frac{1}{2}$, so we have that

$$A(\tau) = \langle x_{n+\tau} x_n \rangle - \mu^2 = \frac{1}{12\eta^\tau}. \tag{13.73}$$

The autocorrelation decreases exponentially with τ. In fact, this is the same autocorrelation (with $\eta=1/\lambda$) that was observed in Equation (13.58) of the damped random walk. So here, following Equation (13.59), we have

$$\chi^2 = A(0)(1-\frac{1}{\eta^2}) = \frac{\eta^2-1}{12\eta^2} \neq 0. \tag{13.74}$$

We have a deterministic system that is not *linearly* deterministic.

For noninteger η, the result is basically the same, though the derivation is not as clean. We do not, for instance, have uniformity over the interval [0,1). However, we do have the result for noninteger $\eta \gg 1$, that $A(\tau)=c(\tau)/\eta^\tau$ with $c(\tau)$ a slowly varying pre-exponential [12].

13.4.7 Generic attractor

$$x_{n+1} = 1 - ax_n^2 + bx_{n-1} \tag{13.75}$$

This is the well-known Hénon map. A plot of x_{n+1} vs. x_n shows the appearance of a chaotic attractor (see Figure 13.1(a)). There are few conclusions that can be derived analytically from the map, so we resort to numerical experiment. Figure 13.1(b) shows how χ_m^2 varies with m, and that $\chi^2>0$. Like most nonlinear deterministic systems, this one is not linearly deterministic.

13.4.8 In general

We find in general for chaotic or expanding systems that $\chi^2>0$. For systems in which nearby trajectories do not diverge, we can always find some near periodic motion, and given this, we can find nearly perfect predictors. In these cases, taking large enough m reduces χ^2_m to be arbitrarily small. That is, $\chi^2=0$ for nonchaotic systems. We have then, *from the linear analysis*, a way to distinguish chaotic from nonchaotic motion. As an example, in Figure 13.2(a), we plot χ^2 as a function of the logistic parameter λ in the map $x_{n+1} = \lambda x_n(1-x_n)$. There is a general increase in linear unpredictability, though windows of nonchaotic $\chi^2=0$ motion can be seen; the same general increase, and the same windows of nonchaotic motion, are seen in Figure 13.2(b), which is a plot of Lyapunov exponent as a function of λ.

13.5 Orthogonal embedding

The method of time-delayed coordinates for embedding a one-dimensional time series into $\mathbb{R}^m$ defines the m dimensional vector

$$v_i = \begin{bmatrix} x_i \\ x_{i+T} \\ \vdots \\ x_{i+(m-1)T} \end{bmatrix}. \tag{13.76}$$

The idea of Broomhead and King [13,14] is to generalize the embedding, so that each component is a linear combination of the available time-delayed coordinates.

$$v_i = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ s_{21} & s_{22} & \cdots & s_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & s_{m3} & s_{mm} \end{bmatrix} \begin{bmatrix} x_i \\ x_{i+T} \\ \vdots \\ x_{i+(m-1)T} \end{bmatrix}. \tag{13.77}$$

That is, the kth component of v_i is

$$v_i^{(k)} = \sum_{j=1}^{m} s_{kj}\, x_{i+(j-1)T} \cdot \tag{13.78}$$

What Broomhead and King show is that if $\left[s_{kj} \right]$ is the matrix of eigenvectors of the covariance matrix Ξ, then the components of the vector v_i will be linearly uncorrelated. That is,

$$\left\langle v_i^{(k)} v_i^{(k')} \right\rangle = \delta_{kk'} , \tag{13.79}$$

where the angle brackets $\langle\rangle$ denote an average over i. The advantage of this embedding is that points are efficiently spread over $\mathbb{R}^m$.

The issue of optimal coordinates for the embedding process is discussed in a nonlinear context by Fraser and Swinney [15]. Here the criterion that axes be uncorrelated is replaced by the criterion that axes be as independent as possible.

13.6 Nonlinear component

In this section, we decompose nonlinear deterministic time series into a linear component and a "purely nonlinear" component. The nonlinear component is distinguished by its lack of autocorrelation; it is by linear standards equivalent to white noise.

Given this decomposition, we can take two approaches. One approach is to

just "subtract off" (or to "filter out") the linear component and be left with the purely nonlinear component, and then to restrict analysis to this standardized time series. Though very straightforward, there are practical difficulties with this approach. A second approach is to create a new time series which are "linearly equivalent" to the original time series; these time series are created by replacing the nonlinear component with the equivalent white noise. The new time series will be nondeterministic and have many degrees of freedom, but its properties are well understood. It provides a benchmark against which to test the original time series.

13.6.1 Filtering out the linear component of a time series

We have seen that any time series can be separated into a linearly predictable component and a "noise" term. We can write

$$x_{n+1} = k_0 x_n + k_1 x_{n-1} + \cdots + k_{m-1} x_{n-m+1} + \gamma \epsilon_{n+1} , \tag{13.80}$$

where k_0, k_1,...,k_{m-1} are the best fit in the sense of Equation (13.12). We can solve for ϵ_{n+1},

$$\epsilon_{n+1} = \frac{x_{n+1} - (k_0 x_n + k_1 x_{n-1} + \cdots + k_{m-1} x_{n-m+1})}{\gamma}, \tag{13.81}$$

and treat $\{\epsilon_n\}$ as the time series of interest. This time series should in some sense be equivalent to the original time series [16]; it is completely deterministic for instance, but there will be no autocorrelation (it will have been "subtracted out") and the Fourier spectrum will be flat, or white.

In practice, however, we find that when we do this the new time series looks "fuzzy." Though the system is strictly deterministic, the linear combination of values well separated in time and therefore nearly uncorrelated from each other leads to behavior that looks very random. For the example of the Hénon map, see Figure 13.3. We have not investigated the effect of this subtracting out on numerical computations of Kolmogorov entropy (and there is still some hope that they may be improved), but we *have* seen that dimension estimates are adversely affected.

13.6.2 Creating linearly equivalent time series

Our original time series can be rewritten in terms of a linearly predictable component and a "noise" term,

$$x_{n+1} = k_0 x_n + k_1 x_{n-1} + \cdots + k_{m-1} x_{n-m+1} + \gamma \epsilon_{n+1}, \tag{13.82}$$

where the noise level γ is given by $\gamma = \sqrt{\chi^2_m}$. The so-called noise term may be completely determinsitic (if the original time series was deterministic) but its correlation properties will be very much like white noise. As $m \to \infty$, in fact, it will satisfy $\langle \epsilon_r \epsilon_s \rangle = \delta_{rs}$. If we replace the deterministic ϵ_n with real white noise, we will obtain a new time series that is linearly indistinguishable from the original, but that is fundamentally stochastic.

As an aside, we point out that there is another way, conceptually more direct, though computationally awkward, of creating linearly equivalent time series. The trick is that the original and the equivalent time series will have the same power spectrum. So, from the Fourier spectrum $P(\omega) = |X(\omega)|^2$, we can

get a Fourier *transform* just by taking the "square root" of the power. An ordinary square root gives the magnitude of the Fourier transform; there is no phase information in the power spectrum. Take a random function $\Phi(\omega)$, and so the transform of the new series is

$$\tilde{X}(\omega) = \sqrt{P(\omega)}\, e^{i\Phi(\omega)}. \tag{13.83}$$

The series itself is just the Fourier transform of this, and is given by Equation (13.5).

Such a time series, we argue, could be very useful as a benchmark against which the original may be measured. The new time series is essentially stochastic, but its linear properties are indistinguishible from the original time series. We can test a dimension algorithm, for instance, by applying it to both time series. If the correlation curves are nearly the same, we have evidence that the original series came from a high dimensional system. On the other hand, if we see a low dimension for our original system, and a high or unsaturating dimension for our new time series, then we have all the more reason to trust the low dimensional result. We can tell if the experiment is a determinsitic or a stochastic system.

13.7 Notes and References

[1] Enders A. Robinson. *Time Series Analysis and Applications* (Goose Pond Press, Houston, 1981).

[2] Actually, a method for distinguishing stochastic and (infinitely

differentiable) deterministic time series directly from the power spectrum has been proposed by D. Sigeti and W. Horsthemke. "High frequency power spectra for systems subject to noise," Phys. Rev. **A** 35 (1987) 2276. However, their method involves the asymptotic high frequency behavior of the power spectrum and this may not be available in a discrete time series. Also, it does not measure the number of degrees of freedom in a deterministic system.

[3] The converse is not true. To say that two random variables are uncorrelated does *not* imply that they are independent. This is a crucial weakness of the linear theory; it can measure correlation but it is insensitive to true dependence or independence.

[4] J. Doyne Farmer and John J. Sidorovich. "Predicting chaotic time series," Phys. Rev. Lett. 59 (1987) 845.

[5] Alan Lapedes, (CNLS, Los Alamos) has been working on approximating the prediction function with a neural network (*private communication*).

[6] Peter Grassberger. "Information content and predictability of lumped and distributed dynamical systems," preprint (1987).

[7] Michael E. Fisher and Robert E. Hartwig. "Toeplitz determinants: some applications, theorems, and conjectures," in *Stochastic Processes in Chemical Physics* (volume XV of *Advances in Chemical Physics*), ed. K. E. Shuler (Interscience publishers, John-Wiley and Sons, New York, 1969), p. 333.

[8] M. Kac. *Summer Institute on Spectral Theory and Statistical Mechanics*, J. D. Pincus, ed, Brookhaven National Laboratories Report BNL 993 (1966) T-442; quoted in Reference 7.

[9] U. Grenander and G. Szegö. *Toeplitz Forms and Their Applications* (University of California Press, Berkeley, California, 1958); quoted in

Reference 7.

[10] Mitchell J. Feigenbaum. "Quantitative universality for a class of nonlinear transformations," J. Stat. Phys. 19 (1978) 25.

[11] There is a measure zero set of exceptions. For instance, if x_0 is a fraction whose denominator is of the form η^K, then $x_k=0$ for all $k \geq K$.

[12] G. M. Zaslavsky. *Chaos in Dynamic Systems* (Harwood Academic Publishers, Chur, Switzerland, 1985) 55.

[13] D. S. Broomhead and Gregory P. King. "Extracting qualitative dynamics from experimental data," Physica 20D (1986) 217.

[14] Broomhead and King, in Reference 13, make further claims regarding the analytical power of linear methods; some of these are refuted in A. I. Mees, P. E. Rapp, and L. S. Jennings. "Singular-value decomposition and embedding dimension," Phys. Rev. A 36 (1987) 340.

[15] Andrew M. Fraser and Harry L. Swinney. "Independent coordinates for strange attractors from mutual information," Phys. Rev. A 33 (1986) 1134.

[16] This equivalence is discussed in W. A. Brock. "Distinguishing random and deterministic systems: abridged version," J. Econ. Theory 40 (1986) 168.

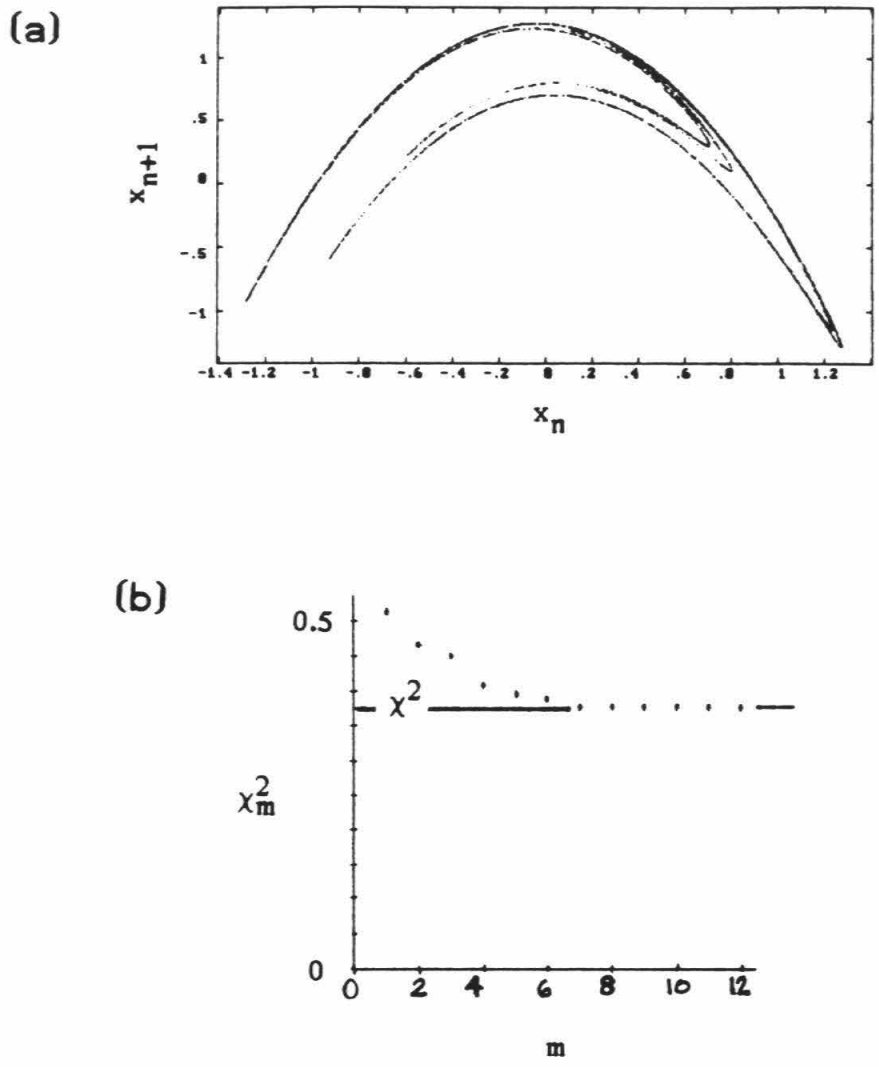


Figure 13.1 (a) x_{n+1} versus x_n for the Hénon map. (b) Measure of linear unpredictability χ^2_m versus m for Hénon attractor: note $\chi^2>0$.

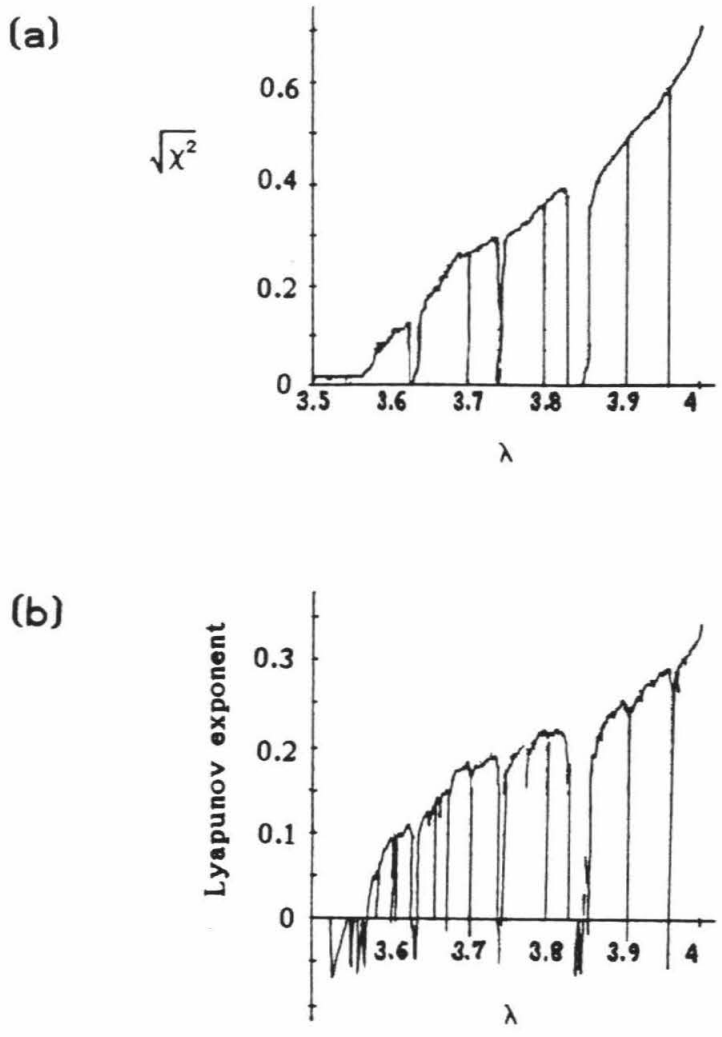


Figure 13.2 Measures of nonlinearity versus logistic parameter λ in the map: $x_{n+1}=\lambda x_n(1-x_n)$. (a) Linear unpredictability $\sqrt{\chi^2}$. (b) Lyapunov exponent.

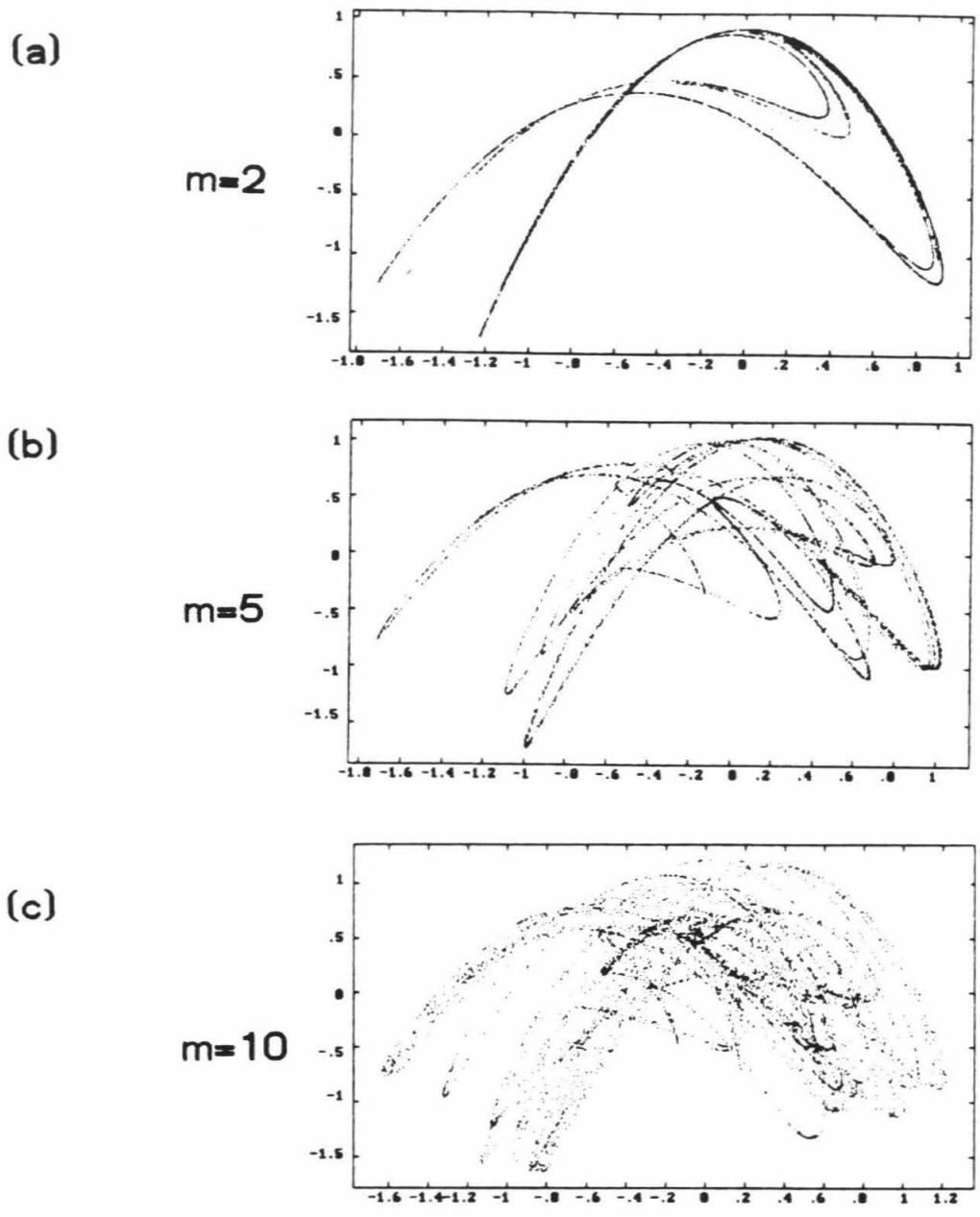


Figure 13.3 Decorrelated time series: ϵ_{n+1} versus ϵ_n for Hénon attractor, where

$$\epsilon_{n+1} \equiv x_{n+1} - \left(k_0 x_n + k_1 x_{n-1} + \cdots + k_{m-1} x_{n-m+1}\right),$$

and the values of $k_0,\ldots,k_{m-1}$ are chosen to minimize $\langle \epsilon^2 \rangle$. As m is increased, the plot of ϵ_{n+1} versus ϵ_n becomes ever fuzzier, and the dimension tends to be overestimated.

CHAPTER FOURTEEN

14. ANALYSIS OF PHYSICAL DATA

Though the bulk of this thesis is devoted to test cases and numerical examples, the dimension algorithm was developed primarily for analyzing time series data from a physical experiment.

There is a vast literature [1] describing physical systems that exhibit deterministic chaos as evidenced by their finite dimensional time series. Most of these cases are either simple systems (*e.g.*, electric circuits) or systems just beyond their onset to turbulence (*e.g.*, fluid flows). It is in principle possible for more complicated systems or for more fully developed turbulence to exhibit low-dimensional chaos. We have little evidence to *expect* that result, but in this chapter, we explore the possibility.

We will discuss some experiments in plasma physics [2]. The two cases we describe involve complicated systems in a regime of well-developed turbulence. Our results are negative, in that we did not observe a dimension; at best we can estimate a lower limit. The turbulence observed in the Caltech tokamak and in the Texas tokamak (TEXT), we *can* assert, is motion involving many ($\gtrsim$10) effective degrees of freedom.

Whether the systems are fundamentally stochastic (with an "infinite" number of degrees of freedom) or "merely" chaotic (with, say, $\lesssim$100 degrees of freedom) we cannot say. But knowing that the dimension is larger than $\sim$10 is still useful; it tells us, for instance, that we could not expect to realistically model the system with less than ten variables.

14.1 The Caltech tokamak

It was in an attempt to understand certain time series output from the Caltech research tokamak [3,4] that the project of this thesis to isolate and identify features of the correlation integral began.

The time series itself was taken from a biased Langmuir probe inserted into the interior of the tokamak machine. Since the probe is biased with respect to the potential of the plasma, it draws a current that is more or less proportional to the local density of the plasma at the probe. Precise proportionality is not important for the time series analysis; we view the time series as a one-dimensional projection from the actual state space, and what Takens [5] proved is that virtually all projections permit a faithful re-embedding into $\mathbb{R}^m$ for large enough m.

A time series from a typical "shot" is shown in Figure 14.1. The tokamak is not quite an autonomous system; each shot must be started up individually, and a single shot lasts 14 milliseconds, and only for about a third of that time can the system be considered stationary. In fact, there is a general drift caused by the deterioration of the plasma (it is compensated somewhat by "puffing," a process of slowly pumping in extra gas to be ionized in the plasma), and our approach has been to subtract a linear fit to the drift from the data, leaving a time series with zero mean and a standard deviation, after discretization, of $\sigma = 20$.

14.1.1 Case history: how not to compute a correlation dimension

Since our goal was to classify tokamak turbulence either as stochastic noise or as deterministic chaos, we created stochastic time series (this in keeping with the philosophy of §13.6.2) with the same variance σ^2 and the same autocorrelation α (see §9.2 for the model) as the tokamak time series.

With the usual time-delayed coordinate embedding procedure (see §1.4.1), we embedded both time series in $\mathbb{R}^m$ for m = 4, 8, 12, and 16. We computed the correlation integral by the standard recipe

$$C(N,r) = \frac{\text{\# of distances less than r}}{\text{\# of distances altogether}}, \tag{14.1}$$

evaluating at logarithmic increments in r. These are plotted in Figure 14.2. From these plots, we were led to the tentative conclusion that there was a seven-dimensional attractor in the tokamak system.

Our first clue that something was amiss came with the realization that we could evaluate C(N,r) at *all* r more efficiently than at just the logarithmic increments. The efficiency was just an issue of better bookkeeping, but having this extra resolution enabled us to see a "shoulder" in the C(N,r) curves for both the tokamak and the stochastic data. When we understood this effect (it is discussed in Chapter Nine), we were able to eliminate it with a slightly modified correlation integral:

$$C(W,N,r) = \frac{\text{\# of distances less than r, except for those from pairs of points closer together in time than W}}{\text{\# of distances altogether}}. \tag{14.2}$$

A less wordy definition is given in Equation (9.3). Plots of the modified (W=10) correlation integrals showed nearly identical curves for the tokamak and the

stochastic data.

14.1.2 Conclusion for the Caltech tokamak

Our dimension algorithms are unable to distinguish the tokamak time series from a stochastic time series of the same autocorrelation. We can provide no evidence that plasma turbulence is deterministic chaos.

14.2 Texas tokamak (TEXT)

Similar analysis was performed on data from the Texas tokamak [6]. We found the Texas data to be a little bit "cleaner" than the Caltech data; the time series was stationary over a longer time, and it had been measured with greater precision (Texas: 12 bits, Caltech: 8 bits).

Nonetheless, as Figure 14.3 shows, our conclusion for the Texas tokamak is about the same as for the Caltech tokamak. There is no evidence for saturation of slope with increasing m, no evidence of a low dimensional strange attractor, and no evidence that plasma turbulence is not a fundamentally stochastic process.

14.3 Notes and References

[1] See References in Chapter One, such as the review by N. B. Abraham, J. P. Gollub, and H. L. Swinney. "Testing nonlinear dynamics," Physica 11D (1984) 252.

[2] Strange attractors *have* been observed for less-than-fully-developed turbulence in plasmas. W. Arter and D. N. Edwards. "Non-linear studies of

Mirnov oscillations in the DITE tokamak: evidence for a strange attractor," Phys. Lett. **114A** (1986) 84; P. Y. Cheung and A. Y. Wong. "Chaotic behavior and period doubling in plasmas," Phys. Rev. Lett. **59** (1987) 551. In the case of more fully developed turbulence, the strange attractor has not been observed: M. L. Sawley, C. W. Simm, and A. Pochelon, "The correlation dimension of broadband fluctuations in a tokamak," Phys. Fluids **30** (1987) 129.

[3] S. J. Zweben and R. W. Gould. "Scaling of edge-plasma turbulence in the Caltech tokamak," Nucl. Fusion **23** (1983) 1625.

[4] Dr. Roy Gould provided raw data from the Caltech tokamak.

[5] Floris Takens. "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick*, 1980, Vol. 898 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1981).

[6] We are indebted to Dr. Christoph W. Ritz (Fusion Research Center, University of Texas at Austin) for providing TEXT data.

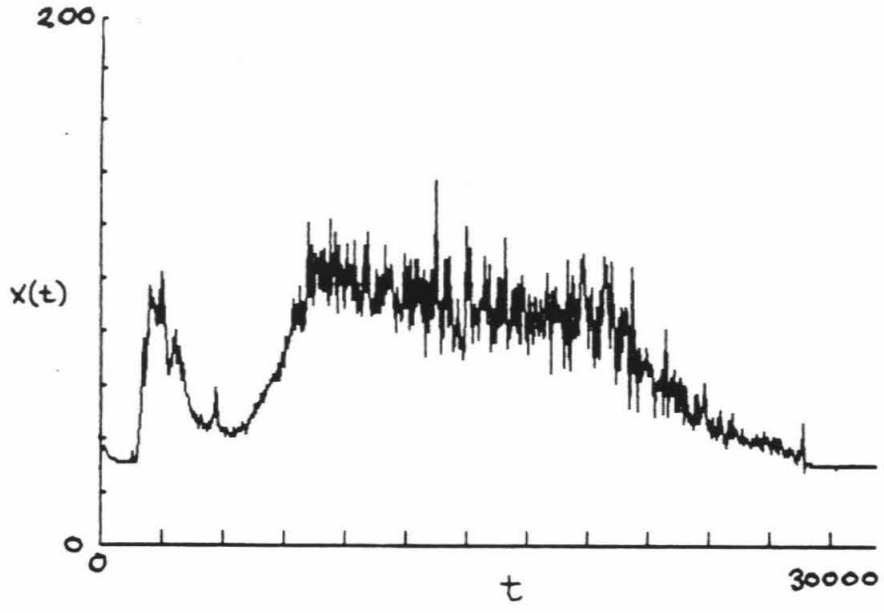


Figure 14.1 Typical "shot" time series from Caltech tokamak. There are N=32768 values in a shot, corresponding to a total time of about 15 msec.

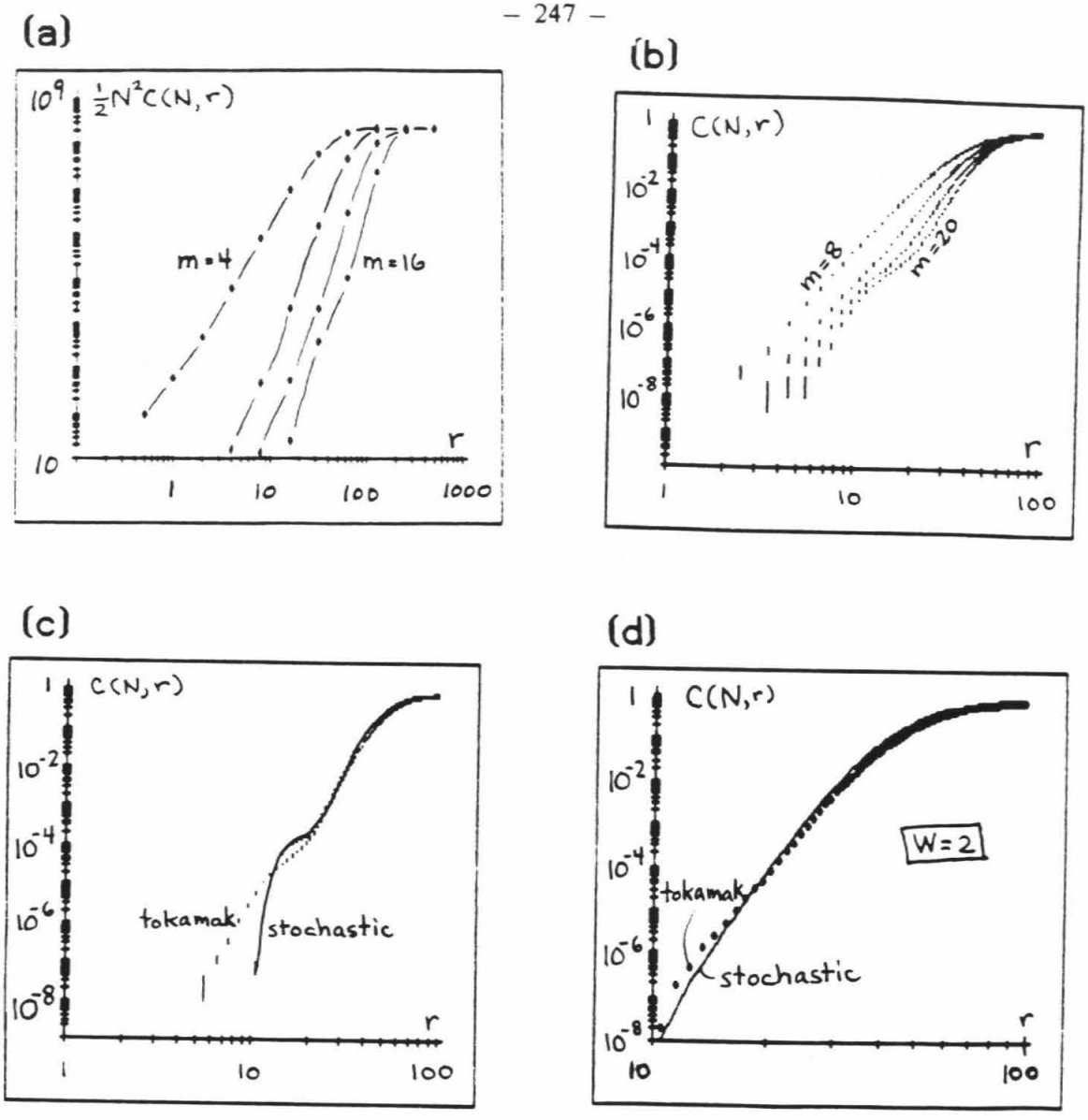


Figure 14.2 Correlation integral for Caltech tokamak data. (a) Standard correlation integral for embedding dimensions m=4,8,12,16. There is an apparent convergence to a slope of about seven, suggesting a seven-dimensional attractor. (b) Standard correlation integral for the same Caltech tokamak data, here at more closely spaced values of r, and for m=8,12,16,20. Here a shoulder is observed for the large m curves. (c) Comparison of autocorrelated stochastic data with tokamak data (same autocorrelation time $\tau=10$) shows that at small r, the value of C(N,r) is much larger for tokamak data than for stochastic data. One inference is that the tokamak *data are not* "as random" as the stochastic data; however, this notion is suspicious since the difference appears to be in the shape of the shoulders. (d) Modified correlation integral (W=2, see Chapter Nine) eliminates most of the shoulder and shows much less distinction between the tokamak and the stochastic correlation integrals. For W=10 (not shown), the tokamak and stochastic correlation integrals are nearly identical.

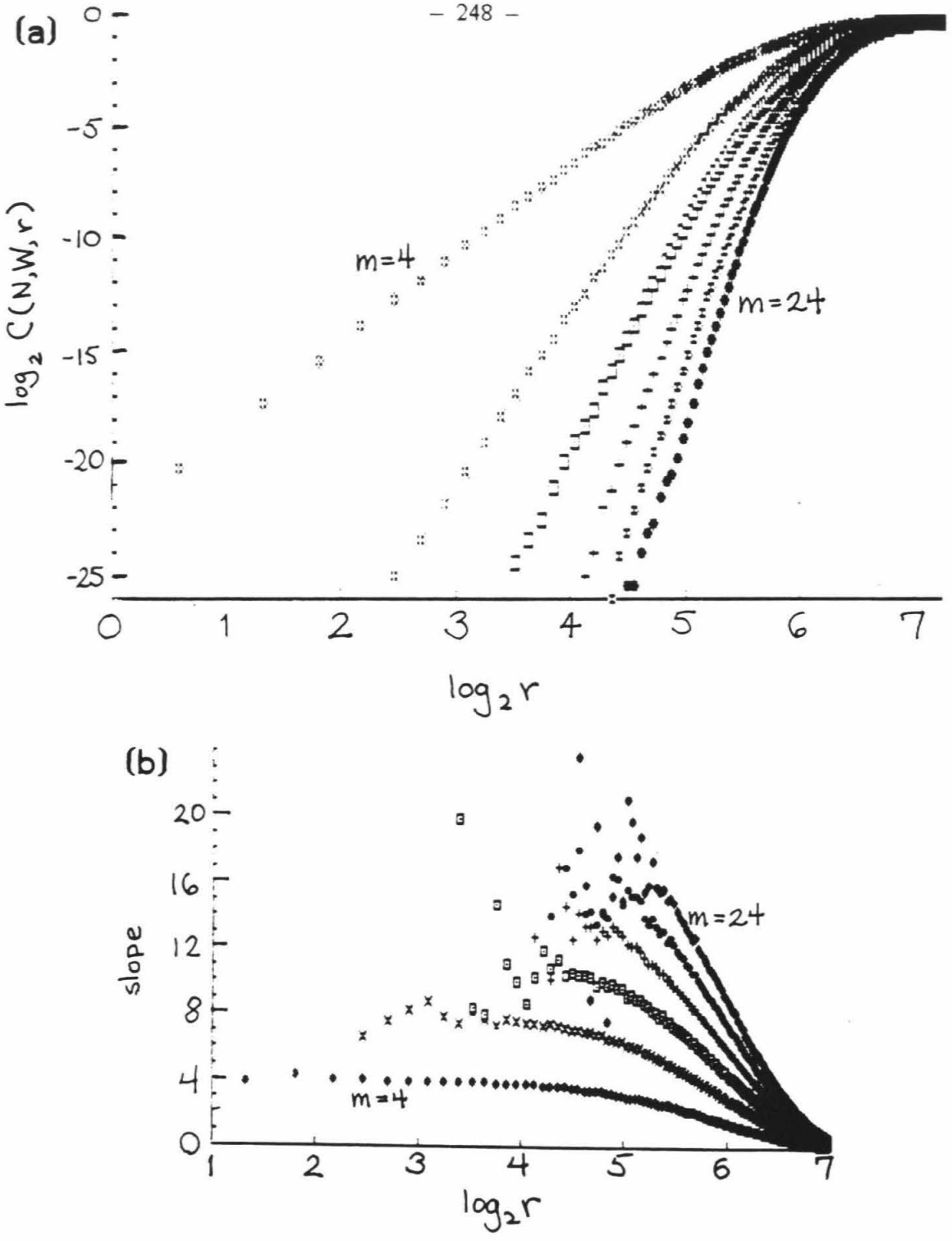


Figure 14.3 Correlation integral for Texas tokamak (TEXT) data. (a) Modified correlation integral (W=10, see Chapter Nine) for embedding dimensions m=4,8,12,16,20,24. (b) Slope of the correlation integrals in (a). We note that no convergence toward dimension is observed.

BIBLIOGRAPHY

N. B. Abraham, J. P. Gollub, and Harry L. Swinney. "Testing nonlinear dynamics," Physica 11D (1984) 252.

K. Aihara, G. Matsumoto, and M. Ichikaura. "An alternating periodic-chaotic sequence observed in neural oscillators," Phys. Lett. 111A (1985) 251.

W. Arter and D. N. Edwards. "Non-linear studies of Mirnov oscillations in the DITE tokamak: evidence for a strange attractor," Phys. Lett. 114A (1986) 84.

P. Atten, J. G. Caputo, B. Malraison, and Y. Gagne. "Détermination de dimension d'attracteurs pour différents écoulements," Journal de Mécanique théorique et appliquée, Numéro spécial (1984) 133.

Ditza Auerbach, Predrag Cvitanović, Jean-Pierre Eckmann, Gemumu Gunaratne, and Itamar Procaccia. "Exploring chaotic motion through periodic orbits," Phys. Rev. Lett. 58 (1987) 2387.

A. Babloyantz, J. M. Salazar, and C. Nicolis. "Evidence of chaotic dynamics of brain activity during the sleep cycles," Phys. Lett. 111A (1985) 152.

R. Badii and A. Politi. "Intrinsic oscillations in measuring the fractal dimension," Phys. Lett. 104A (1984) 303.

Remo Badii and Antonio Politi. "Statistical description of chaotic attractors: the dimension function," J. Stat. Phys. 40 (1985) 725.

R. Badii and A. Politi. "Renyi dimensions from local expansion rates," Phys. Rev. A 35 (1987) 1288.

Hao Bai-Lin. *Chaos* (World Scientific, Singapore, 1984).

M. F. Barnsley and S. Demko. "Iterated function systems and the global construction of fractals." Proc. Roy. Soc. London A 399 (1985) 243.

R. P. Behringer. "Rayleigh-Bénard convection and turbulence in liquid helium," Rev. Mod. Phys. 57 (1985) 657.

Jon Louis Bentley. "Multidimensional Divide-and-Conquer," Comm. ACM 23 (1980) 214.

D. Bessis, J.-D. Fournier, G. Servizi, G. Turchetti, and S. Vaienti. "Mellin transforms of correlation integrals and generalized dimension of strange sets," Phys. Rev. A 36 (1987) 920.

A. Brandstater and Harry L. Swinney. "Strange attractors in weakly turbulent Couette-Taylor flow," Phys. Rev. A 35 (1987) 2207.

W. A. Brock. "Distinguishing random and deterministic systems: abridged version," J. Econ. Theory 40 (1986) 168.

D. S. Broomhead and Gregory P. King. "Extracting qualitative dynamics from experimental data," Physica 20D (1986) 217.

W. E. Caswell and J. A. Yorke. "Invisible Errors in Dimension Calculations: Geometric and Systematic Effects," in *Dimensions and Entropies in Chaotic Systems*, ed. G. Mayer-Kress (Springer, Berlin, 1986).

P. Y. Cheung and A. Y. Wong, "Chaotic behavior and period doubling in plasmas," Phys. Rev. Lett. 59 (1987) 551.

Aviad Cohen and Itamar Procaccia. "Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems," Phys. Rev. A 31 (1985) 1872.

Stephen Demko, Laurie Hodges, and Bruce Naylor. "Construction of fractal objects with Iterated Function Systems," Computer Graphics 19 (1985) 271.

J. Doob. "The Brownian movement and stochastic equations," Ann. Math. 43 (1942) 351.

Ivan Dvorak and Jaromir Siska. "On some problems encountered in the estimation of the correlation dimension of the EEG," Phys. Lett. 118A (1986) 63.

J.-P. Eckmann and D. Ruelle. "Ergodic theory of chaos and strange attractors," Rev. Mod. Phys. 57 (1985) 617.

J.-P. Eckmann, S. Oliffson Kamphorst, D. Ruelle, and S. Ciliberto. "Liapunov exponents from a time series," Phys. Rev. A 34 (1986) 4971.

J. D. Farmer. "Chaotic attractors of an infinite-dimensional dynamical system," Physica 4D (1982) 366.

J. Doyne Farmer, Edward Ott, and James A. Yorke. "The dimension of chaotic attractors," Physica 7D (1983) 153.

J. Doyne Farmer and John J. Sidorowich. "Predicting Chaotic Time Series," Phys. Rev. Lett. 59 (1987) 845.

Mitchell J. Feigenbaum. "Quantitative universality for a class of nonlinear transformations," J. Stat. Phys. 19 (1978) 25.

Michael E. Fisher and Robert E. Hartwig. "Toeplitz determinants: some applications, theorems, and conjectures," in *Stochastic Processes in Chemical Physics*, Vol. XV of *Advances in Chemical Physics*, ed. K. E. Shuler, (Interscience publishers, John-Wiley and Sons, New York, 1969) 333.

G. C. Fox, M. Johnson, G. Lyzenga, S. Otto, J. Salmon, and D. W. Walker. *Solving Problems on Concurrent Processors* (Prentice-Hall, Englewood Cliffs, New Jersey, 1987).

Andrew M. Fraser and Harry L. Swinney. "Independent coordinates for strange attractors from mutual information," Phys. Rev. A 33 (1986) 1134.

H. Froehling, J. P. Crutchfield, D. Farmer, and R. Shaw. "On determining the dimension of chaotic flows," Physica 3D (1981) 605.

S. J. Gee and J. B. Taylor. "Dimension measurement of fluctuations in HBTX1A," Proceedings of the 12th European Conference on Controlled Fusion and Plasma Physics (Budapest, September 1985).

G. Gibson and C. Jeffries. "Observation of period doubling and chaos in spin wave instabilites in yttrium iron garnet," Phys. Rev. A 29 (1984) 811.

R. S. Gioggia and N. B. Abraham. "Routes to chaotic output from a single-mode, dc-excited laser," Phys. Rev. Lett. 51 (1983) 650.

R. S. Gioggia and N. B. Abraham. "Anonalous mode-pulling, instabilities, and chaos in a single-mode, standing-wave 3.39 micron HeNe laser," Phys. Rev. A 29 (1984).

L. Glass, M. R. Guevara, A. Shrier, and R. Perez. "Bifurcation and chaos in a periodically stimulated cardiac oscillator," Physica 7D (1983) 89.

Jean-Michel Grandmont and Pierre Malgrange. "Nonlinear economic dynamics: introduction," J. Econ. Theory 40 (1986) 3.

Peter Grassberger and Itamar Procaccia. "Characterization of strange attractors," Phys. Rev. Lett. 50 (1983) 346.

Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors," Physica 9D (1983) 189.

Peter Grassberger and Itamar Procaccia. "Estimation of the Kolmogorov entropy from a chaotic signal," Phys. Rev. A 28 (1983) 2591.

P. Grassberger. "Generalizations of the Hausdorff dimension of fractal measures," Phys. Lett. 107A (1985) 101.

Peter Grassberger. "Do climatic attractors exist?" Nature 323 (1986) 609.

Peter Grassberger. "Information content and predictability of lumped and distributed dynamical systems," preprint (1987).

H. S. Greenside, A. Wolf, J. Swift, and T. Pignataro. "Impracticality of a box-counting algorithm for calculating the dimensionality of strange attractors," Phys. Rev. A 25 (1982) 3453.

U. Grenander and G. Szegö. *Toeplitz Forms and Their Applications* (University of California Press, Berkeley, California, 1958).

John Guckenheimer and Philip Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Vol. 42 of *Applied Mathematical Sciences* (Springer-Verlag, New York, 1983), 236.

E. G. Gwinn and R. M. Westervelt. "Scaling structure of attractors at the transition from quasiperiodicity to chaos in electronic transport in Ge," Phys. Rev. Lett. 59 (1987) 157.

Thomas C. Halsey, Mogens H. Jensen, Leo P. Kadanoff, Itamar Procaccia, and Boris I. Shraiman. "Fractal measures and their singularities: the characterization of strange sets," Phys. Rev. **A** 33 (1986) 1141.

Felix Hausdorff. "Dimension und äußeres Maß," Mathematische Annalen 79 (1919) 157.

James. W. Havstad and Cindy L. Ehlers. "Attractor dimension from small data sets," preprint (1987).

M. Hénon. "A two-dimensional mapping with a strange attractor," Comm. Math. Phys. 50 (1976) 69.

H. G. E. Hentschel and Itamar Procaccia. "The infinite number of generalized dimensions of fractals and strange attractors," Physica 8D (1983) 435.

Joachim Holzfuss and Gottfried Mayer-Kress. "An approach to error-estimation in the application of dimension algorithms." in *Dimensions and Entropies in Chaotic Systems*, ed. G. Mayer-Kress (Springer-Verlag, Berlin, 1986).

F. A. Hopf, D. L. Kaplan, M. H. Rose, L. D. Sanders, and M. W. Derstine. "Characterization of chaos in a hybrid optically bistable device," Phys. Rev. Lett. 57 (1986) 1394.

Carson D. Jeffries. "Chaotic dynamics of instabilities in solids," Phys. Scr. T9 (1985) 11.

Mogens H. Jensen, Leo P. Kadanoff, Albert Libchaber, Itamar Procaccia, and Joel Stavans. "Global universality at the onset of chaos: results of a forced Rayleigh-Bénard experiment," Phys. Rev. Lett. 55 (1985) 2798.

M. Kac. *Summer Institute on Spectral Theory and Statistical Mechanics*, ed. J. D. Pincus, Brookhaven National Laboratories Report BNL 993 (1966) T-442.

Seung-hwan Kim and Stellan Ostlund. "Simultaneous rational approximations in the study of dynamical systems," Phys. Rev. **A** 34 (1986) 3426.

P. E. Kloeden and A. I. Mees. "Chaotic phenomena," Bull. Math. Biol. 47 (1985) 697.

Edward N. Lorenz. "Deterministic nonperiodic flow," J. Atmos. Sci. 20 (1963) 130.

M. C. Mackey and L. Glass. "Oscillations and chaos in physiological control

systems," Science **197** (1977) 287.

Benoit B. Mandelbrot. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).

Ricardo Mañé. "On the dimension of the compact invariant sets of certain non-linear maps," in *Dynamical Systems and Turbulence, Warwick*, 1980, Vol. 898 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1981).

Mark J. McGuinness. "A computation of the limit capacity of the Lorenz attractor," Physica **16D** (1985) 265.

J. L. Martiel and A. Goldbeter. "Autonomous chaotic behavior of the slime mould *Dictyostelium discoideum* predicted by a model for cyclic AMP signalling," Nature **313** (1985) 590.

Gottfried Mayer-Kress and Scott P. Layne. "Dimensionality of the human electro-encephalogram," preprint (1986).

A. I. Mees, P. E. Rapp, and L. S. Jennings. "Singular-value decomposition and embedding dimension," Phys. Rev. **A 36** (1987) 340.

N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. "Geometry from a time series," Phys. Rev. Lett. **45** (1980) 712.

A. Renyi. *Probability Theory* (North-Holland, Amsterdam, 1970).

Enders A. Robinson. *Time Series Analysis and Applications* (Goose Pond Press, Houston, 1981).

Filipe J. Romeiras and Edward Ott. "Strange nonchaotic attractors of the damped pendulum with quasiperiodic forcing," Phys. Rev. **A 35** (1987) 4404.

J. C. Roux, R. H. Simoyi and H. L. Swinney. "Observation of a strange attractor," Physica **8D** (1983) 257.

M. L. Sawley, C. W. Simm, and A. Pochelon. "The correlation dimension of broadband fluctuations in a tokamak," Phys. Fluids **30** (1987) 129.

Willam M. Schaffer and Mark Kot. "Nearly one-dimensional dynamics in an epidemic," J. Theor. Bio. **112** (1985) 403.

R. S. Shaw. *The Dripping Faucet as a Model Chaotic System* (Aerial Press, Santa Cruz, 1985).

D. Sigeti and W. Horsthemke. "High frequency power spectra for systems subject to noise," Phys. Rev. **A 35** (1987) 2276.

C. W. Simm, M. L. Sawley, F. Skiff and A. Pochelon. "On the analysis of experimental signals for evidence of deterministic chaos," preprint (1986). Submitted for publication to Helvetica Physica Acta.

L. A. Smith, J.-D. Fournier, and E. A. Spiegel. "Lacunarity and Intermittency in Fluid Turbulence." Physics Letters **114A** (1986) 465.

Floris Takens. "Detecting Strange Attractors in Turbulence," in *Dynamical Systems and Turbulence, Warwick*, 1980, Vol. 898 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1981).

Floris Takens. "Invariants related to dimension and entropy," in *Atas do* 13^{O} (colóqkio brasiliero de matemática, Rio de Janeiro, 1983).

Floris Takens. "On the numerical determination of the dimension of an attractor," in *Dynamical Systems and Bifurcations, Groningen*, 1984, Vol. 1125 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1985).

S. W. Teitsworth, R. M. Westervelt, and E. E. Haller. "Non-linear oscillations and chaos in electrical breakdown in Ge," Phys. Rev. Lett. **51** (1983) 825.

Yves Termonia and Zeev Alexandrowicz. "Fractal dimension of strange attractors from radius versus size of arbitrary clusters," Phys. Rev. Lett. **51** (1983) 1265.

J. Testa, J. Perez, and C. Jeffries. "Evidence for universal chaotic behavior of a driven nonlinear oscillator," Phys. Rev. Lett. **48** (1982) 714.

James Testa. "Fractal dimension at chaos of a quasiperiodic driven tunnel diode," Phys. Lett. **111A** (1985) 243.

James Theiler. "Strange attractors and turbulence," Caltech Concurrent Computation Program (C^3P) Technical Bulletin 56.43, 56.44 (30 October 1985 and 6 November 1985) 1:11.

James Theiler. "Spurious dimension from correlation algorithms applied to limited time series data," Phys. Rev. **A 34** (1986) 2427.

James Theiler. "Efficient algorithm for estimating the correlation dimension from a set of discrete points," Phys. Rev. **A** (*in press*).

Alan Wolf, Jack B. Swift, Harry L. Swinney, and John A. Vastano. "Determining Lyapunov exponents from a time series," Physica **16D** (1985) 285.

L.-S. Young. "Capacity of attractors," Ergod. Theory Dynam. Syst. **1** (1981) 381.

G. M. Zaslavskii, Phys. Lett. **69A** (1978) 145.

G. M. Zaslavsky. *Chaos in Dynamic Systems* (Harwood Academic Publishers, Chur, Switzerland, 1985) 55.

S. J. Zweben and R. W. Gould. "Scaling of edge-plasma turbulence in the Caltech tokamak," Nucl. Fusion **23** (1983) 1625.

Arc to Arcturus