Journal of Statistical Physics, Vol. 72, Nos. 5/6, 1993

On the Estimation of Topological Entropy

Sheldon Newhouse[1] **and Thea Pignataro**[2]

Received June 10, 1992; final April 27, 1993

We study a method for estimating the topological entropy of a smooth dynamical system. Our method is based on estimating the logarithmic growth rates of suitably chosen curves in the system. We present two algorithms for this purpose and we analyze each according to its strengths and pitfalls. We also contrast these with a method based on the definition of topological entropy, using (n, ε)-spanning sets.

KEY WORDS: Topological entropy; volume growth; entropy; length growth; dynamical system.

1. INTRODUCTION AND PRELIMINARIES

The topological entropy of a system is a quantitative measure of its orbit complexity. In a certain sense, it is the maximum amount of information lost per unit time by the system using measurements with finite precision. As such, the entropy is an important invariant to know. Since the definition of the entropy requires an exponentially growing number of objects, it is impractical to expect that the definition can effectively be used to estimate it. Fortunately, there are several recent theorems which aid in its estimation. Block *et al.*[2] use the kneading theory in one-dimensional dynamics to develop an algorithm which gives an accurate estimate of the entropy for unimodal maps of the interval when the entropy is not too small. Here we present two algorithms which are suggested by recent results (to be described later) relating the entropy in smooth systems to the logarithmic growth rates of the volumes of disks in the phase space. Our algorithms give good empirical results in a number of cases, but at the

[1] Mathematics Department, University of North Carolina, Chapel Hill, North Carolina 27599-3250.

[2] Mathematics Department, City College of New York, New York, New York 10031.

0022-4715/93/0900-1331$07.00/0

present time they are not well understood from the mathematical point of view. Vaienti[16] relates the entropy of hyperbolic attractors to the so-called Gibbs measures.

As a by-product of our first length growth algorithm, it is worth noting that we obtain an empirical method for detecting chaotic transients in smooth systems.

After our discussions of the use of length growth algorithms we briefly consider a method of estimation of entropy based on the Takens embedding technique. Our preliminary results show that even for low-dimensional systems, the latter method is not practical. Of course in the length growth methods we are given explicit forms for the systems, while the Takens method was designed for application to time series and hence involves the reconstruction of the system via time delay methods. Nevertheless, there are methods for the construction of Lyapunov exponents using the Takens method (e.g., refs. 4 and 17). Since our first length growth method is essentially a type of spatial average of Lyapunov exponents, it would be interesting to see if one could implement some combination of the length growth algorithm with the Takens embedding method to estimate entropy.

We now proceed to a brief summary of the basic definitions and facts concerning topological entropy.

Let (X, d) be a compact metric space with distance function d, and let $f\colon X \to X$ be a continuous self-map of X. Let $\varepsilon > 0$ be a positive real number, and let n be a positive integer. An *n-orbit* is a sequence x, $f(x),\ldots, f^{n-1}(x)$ of f-iterates of a point x in X. Two n-orbits $\{f^i x\}$, $\{f^i y\}$, $0 \leqslant i < n$, are *ε-distinguishable* if there is a $j \in [0, n)$ for which $d(f^j x, f^j y) > \varepsilon$. Let $r(n, \varepsilon, f)$ denote the maximal number of ε-distinguishable n-orbits. It is easy to see that there are numbers $C > 0$ and $\alpha > 0$ such that $r(n, \varepsilon, f) \leqslant Ce^{n\alpha}$ for $n \geqslant 0$.

Let

$$r(\varepsilon, f) = \limsup_{n \to \infty} \frac{1}{n} \log r(n, \varepsilon, f)$$

and let

$$h(f) = \lim_{\varepsilon \to 0} r(\varepsilon, f)$$

The number $h(f)$ is the *topological entropy* of f. For ε small, f has roughly $e^{nh(f)}$ ε-distinguishable n-orbits.

A set $Y \subset X$ is called (n, ε)-*spanning* if for each $x \in X$ there is a $y \in Y$ with $d(f^j x, f^j y) \leqslant \varepsilon$ for all $j \in [0, n)$. Let $s(n, \varepsilon, f)$ denote the minimum cardinality of such a set and define

$$s(\varepsilon, f) = \limsup_{n \to \infty} \frac{1}{n} \log s(n, \varepsilon, f)$$

It is easy to see that $s(\varepsilon, f) \leqslant r(\varepsilon, f) \leqslant s(\varepsilon/2, f)$ and, therefore,

$$h(f) = \lim_{\varepsilon \to 0} s(\varepsilon, f)$$

The number $h(f)$ satisfies the following properties.

1. $h(f^n) = nh(f)$, $n \geqslant 0$.
2. $h(f^{-1}) = h(f)$ for f a homeomorphism.
3. If f: $X \to X$, g: $Y \to Y$, and π: $X \to Y$ are continuous onto mappings such that $g\pi = \pi f$, then $h(f) \geqslant h(g)$. Thus, if π is a homeomorphism (i.e., topological conjugacy), then $h(f) = h(g)$.
4. $h(f)$ is independent of the metric d on X.
5. $h(f^t) = |t| h(f^1)$ if $\{f^t\}$ is a continuous flow.
6. $h(f) = \sup_{\mu \in \mathcal{M}(f)} h_\mu(f)$, where $\mathcal{M}(f)$ denotes the set of f-invariant probability measures on X and $h_\mu(f)$ denotes the measure-theoretic entropy.

Note that the last property implies that if $h(f)$ is positive, then f has invariant probability measures with positive entropy, indicating that f has some chaotic dynamics.

The topological entropy $h(f)$ is usually difficult to compute proceeding from its definition. In smooth systems, the following theorems will suggest algorithms which will enable us to estimate $h(f)$ in many cases.

Let M be a smooth manifold and f: $M \to M$ be a smooth map. Let I denote an interval in the set of real numbers and let n be a positive integer. A mooth map γ: $I \to M$ is called a smooth *curve* in M. For such a curve we can let

$$|\gamma| = \int |\gamma'(t)|\, dt$$

denote its arc length. Then $|f^n(\gamma)|$ is the arc length of the nth iterate of γ. The quantity

$$\bar{G}(\gamma, f) = \limsup_{n \to \infty} \frac{1}{n} \log^+ |f^n(\gamma)|$$

where $\log^+(u) = \max(\log u, 0)$ is called the *upper growth rate* of the length of γ. When $\bar{G}(\gamma, f)$ is actually a limit instead of a lim sup, i.e., when the limit exists, we write $G(\gamma, f)$ instead of $\bar{G}(\gamma, f)$. If our space M happens to be a real interval, then the map f is itself a curve, and we write $\bar{G}(f, f)$ as $\bar{G}(f)$.

In this connection, Misiurewicz and Szlenk proved the following theorem.

Theorem 1 (M-S). Let f be a piecewise monotonic map of the interval I. Then

$$h(f) = \max(0, G(f))$$

This result was generalized as follows.[10,11,19,18]

If f is a smooth self-map of the compact manifold M, and D^k is the unit k-dimensional disk in $\mathbf{R}^k$, a smooth k-disk in M is a smooth map $\gamma: D^k \to M$. We define its *k-dimensional volume* (with multiplicities) to be

$$|\gamma| = \int_{D^k} |\Lambda^k T\gamma(t)|\, d\lambda(t)$$

where $\Lambda^k T\gamma(t)$ denotes the k^{th} exterior power of the derivative $T\gamma$ of γ at the point t in D^k and $d\lambda$ denotes the standard volume element in D^k. Using k-dimensional volume in the same way that arc length is used above, we define $\bar{G}(\gamma, f)$ to be

$$\limsup_{n\to\infty} \frac{1}{n} \log^+ |f^n(\gamma)|$$

and

$$G(\gamma, f) = \lim_{n\to\infty} \frac{1}{n} \log^+ |f^n(\gamma)|$$

when the latter limit exists.

We define $\bar{G}(f) = \sup_\gamma \bar{G}(\gamma, f)$, and

$$G(f) = \sup_{\gamma \text{ for which } G(\gamma,f) \text{ exists}} G(\gamma, f)$$

Theorem 2 (N, Y). 1. Let f be a C^∞ self-map of the compact smooth manifold M. Then

$$h(f) = G(f)$$

2. If f is a C^∞ diffeomorphism of the two-dimensional manifold M^2, or the time-t map of a three-dimensional differential equation, then

$$h(f) = \sup_{\text{smooth curves } \gamma} G(\gamma, f)$$

3. If f is a C^∞ diffeomorphism of a compact two-dimensional manifold M with piecewise smooth boundary ∂M into itself, which decreases area, then

$$h(f) = \max_{\mathscr{C} \text{ a component of } \partial M} G(\mathscr{C}, f)$$

The last part of this theorem states that an area dissipative diffeomorphism of a two-manifold with boundary into itself has its topological entropy given by the length growth rate of its boundary.

2. THE LENGTH GROWTH ALGORITHMS

2.1. The First Length Growth Algorithm

In this section we shall describe the first algorithm to estimate the growth of the length of curves. We initially consider the Hénon family of polynomial mappings of the plane into itself.[7] This is the family of mappings

$$f_{a,b}(x, y) = (a - x^2 + by, x)$$

where a and b are parameters. Numerically, for $a = 1.4$ and $b = 0.3$, $f_{a,b}$ has a strange attractor Λ which is the ω-limit set of the point $(0, 0)$. That is, the forward orbit of $(0, 0)$ accumulates on Λ. The mapping has two fixed points (x_1, y_1) and (x_2, y_2) given by $x_1 = -1.583\ldots$, $y_1 = -1.583\ldots$, $x_2 = 0.883\ldots$, $y_2 = 0.883\ldots$. In this case it can be shown that if γ_0 is a small interval about (x_2, y_2) in the unstable manifold of (x_2, y_2), then $h(f) = G(\gamma_0)$. Thus, to compute the entropy, one should compute the quantity $G(\gamma_0)$. In practice, it turns out that one does not need to know the curve γ_0 exactly. In fact any curve which is C^1 near a piece of the orbit of γ_0 will work. Such a curve γ will be obtained by iterating a unit vector at $(0, 0)$ INIT number of times until it gets close to Λ and then choosing γ to be short line segment centered at $f^{\text{INIT}}(0, 0)$ in the direction of the final unit vector. Then, $G(\gamma)$ will be estimated by computing (approximately of course) the positive part of the logarithm of the length of $f^n(\gamma)$ for $n = 1$ to $n = \text{NUMIT}$ and taking a least squares slope.

We use the following procedure:

1. Let $\mathbf{x}_0 = (0, 0)$ and $\mathbf{v}_0 = (1, 0)$.
2. Given $\mathbf{x}_n$ and $\mathbf{v}_n$ for $n \geqslant 0$, let

$$\mathbf{x}_{n+1} = f(\mathbf{x}_n) \qquad \text{and} \qquad \mathbf{v}_{n+1} = \frac{T_{\mathbf{x}_n} f(\mathbf{v}_n)}{|T_{\mathbf{x}_n} f(\mathbf{v}_n)|}$$

That is, iterate the unit vector and rescale to obtain a new unit vector. We iterate a certain number of times until $\mathbf{x}_n$ is near Λ. Presumably, the vector $\mathbf{v}_n$ will then be near some unstable direction in Λ. At this stage, lay down a line segment γ centered at $\mathbf{x}_n$ which is completely in the basin of attraction of Λ. This is the

curve whose length growth rate we will estimate. We assume that γ is parametrized by a small interval I, and its tangent vector ξ is a unit vector at each point.

3. For $n=1$ to NUMIT, let L_n equal the positive part of the logarithm of the length of $f^n(\gamma)$.
4. Plot L_n against n and take the best least squares fit. The slope of this fit is an estimate of $G(\gamma)$.

The procedure to compute L_n requires rescaling since the numbers are generally growing exponentially. We use the following procedure. For simplicity of notation, let us write log for $\log^+$ in what follows. Thinking of L_n as $\log \int_I |Tf^n\xi|\, dx$, we take points x_k in the interval I and approximate $\int_I |Tf^n\xi|\, dx$ by

$$\frac{1}{N}\sum_{k=0}^{N-1} e^{b_{nk}}$$

where

$$b_{nk} = \log |T_{x_k} f^n \xi|$$

Let $b_n = \max_k b_{nk}$; then

$$\log\left(\frac{1}{N}\sum_{k=0}^{N-1} e^{b_{nk}}\right) = \log \sum_{k=0}^{N-1} e^{b_n + b_{nk} - b_n} - \log N$$

$$= b_n + \log\left(\sum_{k=0}^{N-1} e^{b_{nk} - b_n}\right) - \log N$$

The numbers b_{nk} are the same ones that would be obtained in computing Lyapunov exponents along the orbit of x_k with the initial vector ξ. They are computed recursively as in a Lyapunov exponent calculation in the following way. Let $v_{0,k} = \xi$, $b_{0k} = 1$. For $n \geqslant 0$, let

$$v_{n+1,k} = \frac{T_{f^n x_k} f(v_{n,k})}{|T_{f^n x_k} f(v_{n,k})|}$$

and let

$$a_{n+1,k} = |T_{f^n x_k} f(v_{n,k})|$$

Then, let

$$b_{n+1,k} = \frac{n}{n+1} b_{n,k} + \frac{1}{n+1} \log(a_{n+1,k})$$

We get reasonable estimates for $G(\gamma)$ which agree to a certain extent with rigorously obtainable numbers in several cases. Figure 1 contains a plot of L_n against n in the case $a = 1.4$, $b = 0.3$.

In Fig. 1, n runs from MINIT = 5 to MAXIT = 300 in increments of 5, and INIT was 1000. If we let MAXIT be much larger, then the L_n vs. n plot falls off slightly, so that the least squares slope decreases. This occurs because our discrete approximation to γ has too few points to accurately estimate the length of $f^n \circ \gamma$ for n large. Thus, there is an inherent tradeoff in our procedure.

Assuming

$$h(f) = \lim_{n \to \infty} \frac{1}{n} \log^+ |f^n \circ \gamma|$$

we want n large in order to approximate $h(f)$ by $(1/n) \log^+ |f^n \circ \gamma|$ [or the least squares slope of $\{(q, L_q)\}$]; but for large n, we need many points and much computer time to estimate L_n. In practice, we examine a plot of L_n vs. n to see which values to use in the least squares fit. The problem of varying MINIT, MAXIT, the number of data points (ISIZE), and the initial starting vector is heavily dependent on the map or differential equation under consideration. We will not discuss this in detail here.

As a method of checking our estimates, we may consider the case $b = 0$. In this case, we have the map $(x, y) \to (a - x^2, x)$. Its orbit structure is determined by the one-dimensional unimodal map f_a: $x \to a - x^2$. We may as well assume that f_a is defined on the interval $[-2, 2]$. Here it is known that if f_a has a single periodic sink, then the entropy can be calculated in terms of the orbit of the sink. The procedure is as follows. Let $z_0, z_1, \ldots, z_{\tau-1}$ be the successive iterates of the sink, $f^i(z_0) = z_i$. Place them

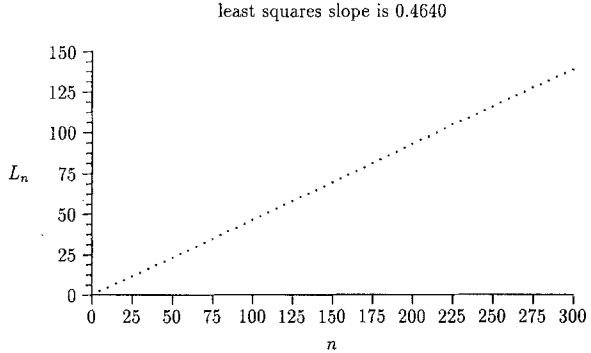

Fig. 1. Hénon map with $a = 1.4$, $b = 0.3$.

in the interval $[-2, 2]$ and relabel them so that they form an increasing sequence: $z_{i_0}, z_{i_1}, \ldots, z_{i_{\tau-1}}$.

Consider the $\tau - 1$ intervals $I_j = [z_{i_j}, z_{i_{j+1}}]$ with $0 \leqslant \tau - 1$. Define a $(\tau-1)\times(\tau-1)$ matrix A of 0's and 1's by

$$A_{kl} = 1 \qquad \text{if and only if} \quad f(I_k) \supseteq (I_l)$$

Let $\lambda(A)$ be the unique real eigenvalue of A of largest modulus. Then $h(f_\alpha) = \log \lambda(A)$.

We give two examples.

1. $a = 1.76$. Here there is a period-3 sink approximately equal to 1.76. The intervals I_0, I_1 are $I_0 = [-1.33, -0.023]$, $I_1 = [-0.023, 1.76]$. The matrix A is $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, and the entropy is $\log[(1+\sqrt{5})/2] \approx 0.48121$.

2. $a = 1.6280$. Here there is a period-5 sink approximately at 1.625. The matrix A is

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

and the entropy is approximately 0.4140.

Figure 2a shows a plot of the output of algorithm 1 with INIT = 1, MAXIT = 50 for $a = 1.76$, while Fig. 2b contains the output for $a = 1.628$. Figure 2c shows the output of algorithm 1 with n running from 150 to 600. Note that the right side of the graph begins to decrease. This indicates that there is a transient chaotic set on which the entropy lives, and that most points eventually settle down to the period-five sink.

As one can see, there is good agreement with the precise theoretical results.

Typically, in a smooth system, there may be several regions with positive entropy, and the entropy of f will be the maximum of those entropies. For instance, suppose one has several strange attractors with positive entropy. One could of course apply the above algorithm in each basin separately, but it is interesting to take a line segment crossing several basins and to try to see how its length grows in a particular subregion. This suggests a slight modification of algorithm 1 which we shall call algorithm 1a. Let $R \supset R_1$ be regions in the plane and let γ be a line segment in R meeting R_1. For each integer $n \geqslant 0$, let $W^s(R_1, n) = \{x\colon f^j x \in R_1$ for $0 \leqslant j \leqslant n\}$. The set $\gamma^{-1}(W^s(R_1, n))$ is the set of points x in the domain of

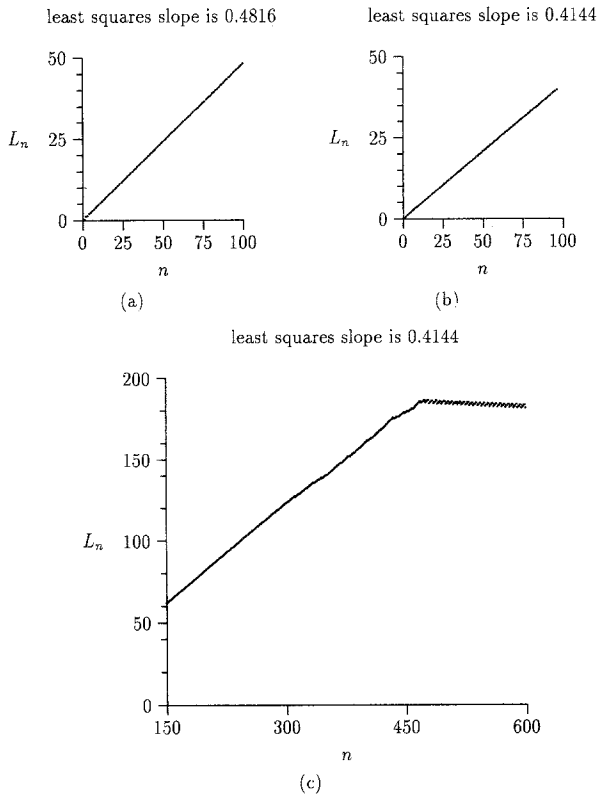

Fig. 2. Hénon map with $b = 0$. (a) $a = 1.76$, (b) $a = 1.628$, (c) $a = 1.628$.

γ such that $f^j(\gamma(x)) \in R_1$ for $j \in [0, n)$. The length of $f^{n-1}(\gamma^{-1}(W^s(R_1, n)))$ is

$$\int_{\gamma^{-1}(W^s(R_1, n))} |T_{\gamma(t)} f^n \circ \gamma(t)| \, dt$$

and will be denoted $l(f^n \circ \gamma, R_1)$. We set

$$\bar{G}(\gamma, R_1) = \limsup_{n \to \infty} \frac{1}{n} \log^+ l(f^n \circ \gamma, R_1)$$

and

$$G(\gamma, R_1) = \lim_{n \to \infty} \frac{1}{n} \log^+ l(f^n \circ \gamma, R_1)$$

Algorithm 1a is the same as algorithm 1 except that we use INIT = 0 and we compute $|Tf^k \circ \gamma(t)|$ only at points in $W^s(R_1, k)$. This algorithm gives

useful information if most of the points on γ lie in $W^s(R_1, \text{MAXIT})$. Peculiar output from algorithm 1a suggests that the relations between γ, R_1, the entropy-producing sets in R_1, and the parameters like INIT, MAXIT, etc., should be investigated more closely.

2.2. The Second Length Growth Algorithm

In this section we present a second algorithm for the estimation of length growth, which is based on the following observations.

Table I. Hénon Map with $1.4 \leqslant a \leqslant 1.7$ and $0 \leqslant b \leqslant 0.07$[a]

a \ b	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
1.40	0.0261	0.1152	0.1614	0.1985	0.2155	0.1152	0.2368	0.2617
1.41	0.0988	0.1437	0.1906	0.2095	0.2300	0.1437	0.2558	0.2771
1.42	0.1247	0.1826	0.2013	0.2242	0.2338	0.1827	0.2698	0.2947
1.43	0.1545	0.1993	0.2161	0.2342	0.2387	0.1996	0.2832	0.3030
1.44	0.1898	0.2100	0.2264	0.2323	0.2556	0.2101	0.2971	0.3215
1.45	0.1984	0.2226	0.2314	0.2491	0.2697	0.2262	0.3125	0.3561
1.46	0.2140	0.2303	0.2375	0.2616	0.2877	0.2331	0.3467	0.3678
1.47	0.2286	0.2303	0.2575	0.2799	0.2998	0.2350	0.3643	0.3773
1.48	0.2303	0.2492	0.2739	0.2914	0.3162	0.2534	0.3743	0.3825
1.49	0.2391	0.2622	0.2883	0.3089	0.3546	0.2657	0.3821	0.3889
1.50	0.2570	0.2810	0.3010	0.3305	0.3670	0.2872	0.3879	0.3970
1.51	0.2708	0.2918	0.3203	0.3635	0.3764	0.2969	0.3951	0.4031
1.51	0.2893	0.3109	0.3569	0.3731	0.3821	0.3159	0.4013	0.4109
1.53	0.3009	0.3446	0.3685	0.3822	0.3900	0.3460	0.4088	0.4132
1.54	0.3235	0.3652	0.3780	0.3873	0.3971	0.3651	0.4136	0.4141
1.55	0.3588	0.3734	0.3831	0.3932	0.4035	0.3745	0.4141	0.4157
1.56	0.3700	0.3820	0.3910	0.4008	0.4109	0.3821	0.4140	0.4203
1.57	0.3794	0.3888	0.3968	0.4101	0.4136	0.3884	0.4202	0.4275
1.58	0.3845	0.3967	0.4060	0.4159	0.4138	0.3968	0.4278	0.4324
1.59	0.3930	0.4031	0.4129	0.4145	0.4173	0.4034	0.4306	0.4369
1.60	0.3995	0.4109	0.4141	0.4161	0.4261	0.4109	0.4373	0.4410
1.61	0.4086	0.4141	0.4148	0.4256	0.4316	0.4140	0.4440	0.4413
1.62	0.4138	0.4147	0.4230	0.4304	0.4386	0.4143	0.4432	0.4419
1.63	0.4137	0.4194	0.4304	0.4375	0.4410	0.4201	0.4456	0.4418
1.64	0.4180	0.4288	0.4360	0.4416	0.4444	0.4283	0.4454	0.4413
1.65	0.4275	0.4350	0.4416	0.4482	0.4514	0.4355	0.4447	0.4437
1.66	0.4329	0.4418	0.4468	0.4529	0.4553	0.4416	0.4513	0.4437
1.67	0.4409	0.4465	0.4533	0.4565	0.4581	0.4461	0.4593	0.4587
1.68	0.4433	0.4533	0.4609	0.4633	0.4606	0.4521	0.4632	0.4618
1.69	0.4524	0.4617	0.4652	0.4462	0.4625	0.4609	0.4614	0.4695
1.70	0.4613	0.4671	0.4681	0.4693	0.4694	0.4678	0.4678	0.4787

[a] The data were produced with ISIZE = 512000.

Assume γ is given. Let $V_0 = |\gamma|$ and $V_n = |f^n \circ \gamma|$. Then,

$$\frac{1}{n}\sum_{k=0}^{n-1} \log \frac{V_{k+1}}{V_k} = \frac{1}{n}(\log V_n - \log V_0)$$

Thus, $(1/n)(\log V_n - \log V_0)$ is the average of the quantities $\log(V_{k+1}/V_k)$ for $0 \leqslant k \leqslant n-1$. The quantity V_{k+1}/V_k equals

$$\frac{1}{V_k}\int |Tf(\gamma_k(t))| \cdot |\gamma_k'(t)|\, dt$$

Table II. Hénon Map with $1.71 \leqslant a \leqslant 2$ and $0 \leqslant b \leqslant 0.07$[a]

a \ b	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
1.71	0.4689	0.4701	0.4699	0.4682	0.4762	0.4701	0.4779	0.4827
1.72	0.4730	0.4727	0.4735	0.4678	0.4816	0.4748	0.4875	0.4882
1.73	0.4772	0.4792	0.4758	0.4719	0.4852	0.4761	0.4929	0.4893
1.74	0.4804	0.4782	0.4773	0.4809	0.4875	0.4780	0.4984	0.5030
1.75	0.4813	0.4781	0.4822	0.4904	0.4963	0.4788	0.5036	0.5124
1.76	0.4811	0.4808	0.4880	0.4967	0.5026	0.4828	0.5172	0.5264
1.77	0.4844	0.4848	0.4953	0.5010	0.5119	0.4879	0.5236	0.5371
1.78	0.4839	0.4903	0.5054	0.5115	0.5141	0.4914	0.5350	0.5411
1.79	0.4822	0.4973	0.5078	0.5176	0.5193	0.4947	0.5437	0.5498
1.80	0.4870	0.5022	0.5152	0.5205	0.5284	0.5022	0.5483	0.5556
1.81	0.4982	0.5067	0.5207	0.5261	0.5379	0.5053	0.5565	0.5588
1.82	0.5108	0.5137	0.5225	0.5354	0.5473	0.5141	0.5617	0.5663
1.83	0.5212	0.5230	0.5257	0.5411	0.5541	0.5235	0.5652	0.5740
1.84	0.5295	0.5291	0.5356	0.5493	0.5612	0.5280	0.5727	0.5863
1.85	0.5394	0.5372	0.5440	0.5579	0.5681	0.5358	0.5818	0.5954
1.86	0.5427	0.5457	0.5551	0.5635	0.5702	0.5458	0.5940	0.6117
1.87	0.5475	0.5562	0.5639	0.5705	0.5714	0.5562	0.6017	0.6701
1.88	0.5587	0.5657	0.5710	0.5735	0.5792	0.5659	0.6186	0.6555
1.89	0.5669	0.5722	0.5801	0.5775	0.5865	0.5724	0.7077	0.6841
1.90	0.5574	0.5812	0.5849	0.5877	0.5996	0.5811	0.6822	0.7623
1.91	0.5849	0.5894	0.5883	3.5976	0.6076	0.5896	0.6379	0.7069
1.92	0.5940	0.5967	0.5994	0.6099	0.6268	0.5965	0.6378	0.7834
1.93	0.6028	0.6030	0.6082	0.6186	0.7254	0.6030	0.6241	0.6918
1.94	0.6095	0.6122	0.6209	0.6381	0.6854	0.6125	0.6380	0.6415
1.95	0.6140	0.6217	0.6310	0.6930	0.7092	0.6215	0.6748	0.6269
1.96	0.6256	0.6331	0.6510	0.7046	0.7017	0.6333	0.7226	0.6214
1.97	0.6342	0.6414	0.7176	0.7042	0.7269	0.6413	0.7367	0.7008
1.98	0.6495	0.6621	0.7589	0.7257	0.7164	0.6620	***	0.6936
1.99	0.6629	0.6980	0.7202	0.7267	0.7152	0.6856	***	0.7356
2.00	0.6931	0.7348	0.7297	0.7244	0.7315	0.6943	***	0.7156

[a] The data were produced with ISIZE = 512000.

where $\gamma_k(t) = f^k(\gamma(t))$. This can be thought of as the integral of $u \to |Tf(u)|$ with respect to the normalized arc length measure on $f^k \circ \gamma$. After discretization this yields the formula for algorithm 2. For each k, a weighted average of the expansion of Tf along γ_k is evaluated. This calculation hides the large numbers that appear in algorithms 1 and 1a, since one renormalizes at each iterate. Using algorithm 2 with γ the line segment joining $(-1, 0)$ and $(1, 0)$, and taking the region R_1 to be the strip $-3 \leqslant x \leqslant 3$, we obtained estimates of $h(f_{a,b})$ for various a, b in the Hénon mappings described previously. Tables I and II present the results in matrix

Table III. Variances Corresponding to Table II[a]

a	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
1.71	0.0001	0.0002	0.0001	0.0002	0.0001	0.0000	0.0002	0.0001
1.72	0.0001	0.0000	0.0001	0.0004	0.0001	0.0001	0.0001	0.0002
1.73	0.0000	0.0001	0.0011	0.0002	0.0002	0.0001	0.0001	0.0001
1.74	0.0001	0.0002	0.0004	0.0002	0.0001	0.0002	0.0004	0.0001
1.75	0.0006	0.0008	0.0004	0.0004	0.0003	0.0002	0.0002	0.0004
1.76	0.0002	0.0003	0.0010	0.0000	0.0001	0.0005	0.0015	0.0000
1.77	0.0010	0.0003	0.0002	0.0001	0.0001	0.0001	0.0003	0.0001
1.78	0.0005	0.0000	0.0001	0.0000	0.0001	0.0001	0.0001	0.0000
1.79	0.0007	0.0003	0.0001	0.0000	0.0001	0.0001	0.0000	0.0000
1.80	0.0002	0.0004	0.0007	0.0000	0.0001	0.0001	0.0000	0.0001
1.81	0.0003	0.0002	0.0000	0.0002	0.0001	0.0001	0.0001	0.0001
1.82	0.0001	0.0001	0.0001	0.0001	0.0000	0.0002	0.0003	0.0000
1.83	0.0000	0.0001	0.0002	0.0000	0.0000	0.0001	0.0000	0.0000
1.84	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
1.85	0.0001	0.0001	0.0001	0.0003	0.0000	0.0001	0.0001	0.0000
1.86	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.87	0.0000	0.0000	0.0003	0.0000	0.0001	0.0000	0.0000	0.0448
1.88	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0275
1.89	0.0000	0.0000	0.0000	0.0001	0.0001	0.0000	0.0233	0.0125
1.90	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0095	0.0514
1.91	0.0000	0.0000	0.0000	0.0001	0.0000	0.0001	0.0156	0.0197
1.92	0.0000	0.0000	0.0009	0.0000	0.0000	0.0000	0.0147	0.0725
1.93	0.0000	0.0000	0.0000	0.0000	0.0223	0.0000	0.0139	0.0117
1.94	0.0000	0.0000	0.0000	0.0000	0.0064	0.0000	0.0111	0.0329
1.95	0.0000	0.0000	0.0000	0.0070	0.0193	0.0000	0.0250	0.0298
1.96	0.0000	0.0000	0.0001	0.0097	0.0308	0.0000	0.0150	0.0292
1.97	0.0000	0.0000	0.0127	0.0269	0.0325	0.0000	0.0195	0.0506
1.98	0.0000	0.0000	0.0467	0.0378	0.0215	0.0000	0.0195	0.0409
1.99	0.0000	0.0046	0.0097	0.0519	0.0562	0.0002	0.0195	0.0466
2.00	0.0000	0.0117	0.0242	0.0272	0.0376	0.0007	0.0195	0.0285

[a] The data were produced with ISIZE = 512000.

format. The value of a is fixed across rows, while b is fixed going down columns. Entries with asterisks indicate poor convergence, which we ascribe to the fact that most points eventually escape the region R_1.

We also estimated $h(f_{a,b})$ with γ a line segment of unit length centered at the rightmost fixed point of $f_{a,b}$ in the direction of the expanding eigenvector. We use ISIZE = 32000 and MAXIT = 50. The results are fairly consistent with the previous ones.

In each case when ambiguities appear, going to higher ISIZE or specifying different line segments seems to resolve them. As an empirical measure of the adequacy of ISIZE in algorithm 2, we used the variance

$$\frac{1}{\Delta(\mathrm{IT})} \sum_{k=\mathrm{MINIT}}^{\mathrm{MAXIT}} \log\left(\frac{V_{k+1}}{V_k}\right)^2 - \left(\frac{1}{\Delta(\mathrm{IT})} \sum \log \frac{V_{k+1}}{V_k}\right)^2$$

where $\Delta(\mathrm{IT}) = \mathrm{MAXIT} - \mathrm{MINIT}$. When this variance was larger than 10^{-3}, an increase in ISIZE seemed to be necessary. This is evident in Table III, which shows the value of the variance in each of the cases corresponding to the entropy estimates in Table II.

2.3. Applications to Differential Equations

Next we present an example of algorithm 1 applied to differential equations. The Lorenz system is the set of ordinary differential equations

$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = -xz + rx - y$$
$$\dot{z} = xy - bz$$

where σ, r, and b are real parameters.[9] Fixing $b = 8/3$ and $\sigma = 10$, one sees the following. For $0 < r < 1$, the origin (0, 0, 0) is a globally attracting fixed point. When $1 < r < 24.74$, the origin becomes unstable and the two fixed points $(\pm[b(r-1)]^{1/2}, \pm[b(r-1)]^{1/2}, r-1)$ become locally attracting. For $r > 24.74$, the chaotic behavior known as the Lorentz attractor appears. We estimated the entropy of the time-one map of the Lorentz system for various parameter values. The result using algorithm 1 with $r = 28$ is shown in Table IV.

We used the algorithm to estimate the entropy for some other parameters in the Lorentz system as well. Figure 3 shows some outputs for $r = 22$. Note the clear presence of chaotic transients in this case. Figure 4 shows some outputs for $r = 24$, 26.

Table IV. Lorenz System with $\sigma = 10$, $r = 28$, $b = 2.6$[a]

N	L	E
100	0.894741E+02	0.894741E+00
200	0.180095E+03	0.900476E+00
300	0.270800E+03	0.902666E+00
400	0.361462E+03	0.903654E+00
500	0.452383E+03	0.904765E+00
600	0.543644E+03	0.906073E+00
700	0.633327E+03	0.904753E+00
800	0.727047E+03	0.908808E+00
900	0.815244E+03	0.905827E+00
1000	0.905984E+03	0.905984E+00
1100	0.995469E+03	0.904972E+00
1200	0.108633E+04	0.905272E+00
1300	0.117884E+04	0.906797E+00
1400	0.127094E+04	0.907817E+00
1500	0.136123E+04	0.907484E+00

[a] N = number of iterates of time-one map, L = log of length, E = estimate of entropy = L/N. ISIZE = 2500, MAXIT = 1500.

3. THE DIRECT METHOD

3.1. The Algorithm

For comparison, topological entropy was calculated using what will be referred to as the *Takens method.* Beginning with a single time series of

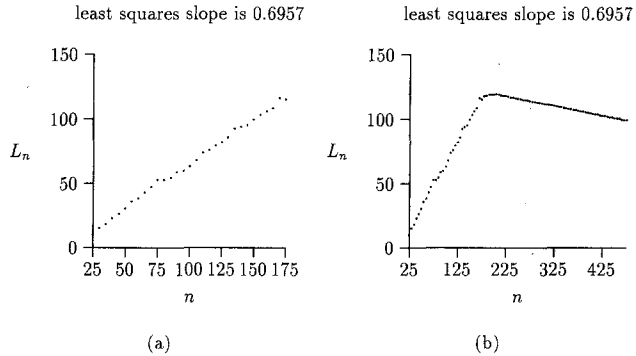

Fig. 3. Lorenz equation with $r = 22$.

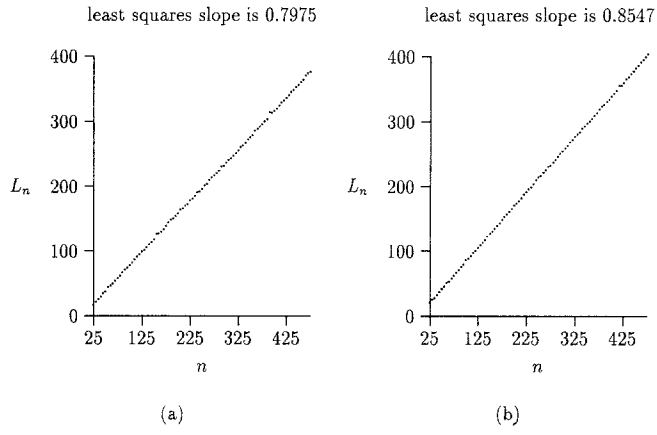

Fig. 4. Lorenz equation with $r =$ (a)24, (b)26.

data $\{x_i\}_{i=1}^{N}$, the underlying attractor is reconstructed using time-delay coordinates.[12] This means a sequence of n-dimensional vectors

$$S_n = \{(x_i, x_{i+1}, \ldots, x_{i+n-1})\}_{i=1}^{N-n+1}$$

is used to represent points on the attractor, M, in $\mathbf{R}^n$. So long as $n \geqslant 2m+1$ (where m is the actual dimension of M), Takens proved that, in the ideal case when $N = \infty$ and certain generic assumptions are made, there is a one-to-one correspondence between the positive limit set of the dynamical system producing the attractor and the set of limit points of the sequence S_n.[15] In other words, there is a one-to-one correspondence between the asymptotic behavior of the real and reconstructed systems. This fact has been used to calculate the dimension of M, both for numerically generated time series and for those from physical experiments (e.g., ref. 14). The limitations of these methods for calculating dimension have been explored[6,8] and similar difficulties can be expected in calculating topological entropy along these lines.[13]

In the case of entropy, an approximate (n, ε)-spanning set $S_{n,\varepsilon}$ is obtained from the sequence of vectors S_n in the following way. For $n > 2m+1$, the n-dimensional vectors represent points on the embedding of M into $\mathbf{R}^{2m+1}$ together with a segment of their orbits of length $n-(2m+1)$. Choosing a subset $S_{n,\varepsilon} \subset S_n$ whose vectors all differ by at least ε from each other in some coordinate and such that all other vectors in S_n lie within ε of some vector in $S_{n,\varepsilon}$ gives an $(n-2m-1, \varepsilon)$-spanning set. For large n, this is our approximate (n, ε)-spanning set. Again, there is a

rigorous foundation to saying that the topological entropy of the dynamical system restricted to the positive limit set is

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{\log C_{n,\varepsilon}}{n}$$

where $C_{n,\varepsilon}$ is the cardinality of $S_{n,\varepsilon}$.[15] In practice, $C_{n,\varepsilon}$ is calculated as a function of increasing n, over a range of small values of ε, with the maximum value of n as large as computationally feasible. Then $\log C_{n,\varepsilon}$ is plotted against n for fixed ε, with the asymptotic slope giving the entropy as ε tends to zero. Looking at slopes rather than ratios helps to minimize the effects of slowly vanishing nonasymptotic terms (a standard procedure used in dimension calculations).

3.2. Numerical Results

Transients are avoided by skipping a number of iterations or integration steps before forming the time series. The idea is that the calculations should begin on or near the attractor. The first computational hurdle is the length N of the time series that is needed to create a reasonable reconstruction of the attractor. This can be determined by plotting $C_{n,\varepsilon}$ vs. N. Once N is sufficiently large so that S covers M, $C_{n,\varepsilon}$ should stop increasing as a function of N, as in Fig. 5. Since this frequently occurred only after N had reached several million, least squares fits of the data to a constant minus a multiple of $1/N$, $1/N^2$, and $1/\sqrt{N}$ were performed in an attempt to predict the asymptotic value of $C_{n,\varepsilon}$ (i.e., the constant in the fit). In general, this did not prove useful, since a million or more data points were still needed to capture the asymptotics. Also, there was no clear-cut best fit. The

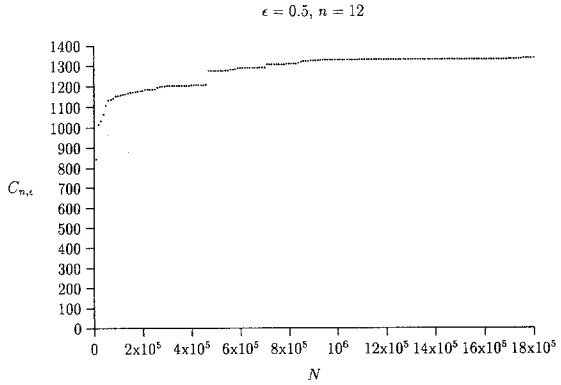

Fig. 5. Lorenz system with $\sigma = 10$, $r = 28$, $b = 2.6$.

three cases were almost evenly divided as being the best fit for given sets of data ($1/N$ was slightly favored overall). In all cases, each of the correlation coefficients was greater than or equal to 0.90 in absolute value, with the best fit case usually being 0.99.

The next complication is the tradeoff between memory and CPU time. Calculations are done in double precision and the length n of the vectors being compared is taken to be as large as possible (a maximum of 21 in the calculations described here). Since the number of vectors is large, sorting and storing become prohibitive, and computations become CPU intensive.

The analysis was performed on several different models, with varying degrees of success. Total automorphisms and unimodal one-dimensional maps, for which the topological entropy is known rigorously, were used as test cases. These yielded reasonably accurate results using the Takens method.

Recall that entropy is estimated by the slope of the straight line between successive points on the graph $\log C_{n,\varepsilon}$ vs. n. For the Hénon map with $a = 1.4$ and $b = 0.3$, the results are summarized in Table V. They are in agreement with those using the length growth methods, giving approximately 0.46 for the entropy. It should be noted that the estimates for large n and/or small ε may be less accurate because these are precisely the cases when the difficulties described in connection with Fig. 5 arise.

Although calculations for the Lorenz system with $r = 28$ are still underway, the results do not seem promising. Because of the large number of points needed in the time series, the time-one-tenth map was used.[3] This

[3] The equations were integrated using a fourth-order Runge-Kutta scheme with a time step of 10^{-2}. The time series was formed by keeping every tenth point.

Table V. Hénon Map with $a = 1.4$ and $b = 0.3$

Epsilon	Entropy estimate = slope between successive values of n in parentheses								
0.3	0.45 (6–11)	0.44 (11–16)	0.44 (16–21)						
0.2	0.47 (4–5)	0.45 (5–6)	0.43 (6–7)	0.45 (7–9)	0.46 (9–11)	0.45 (11–15)			
0.1	0.49 (4–5)	0.43 (5–6)	0.46 (6–7)	0.49 (7–8)	0.48 (8–9)	0.45 (9–10)	0.46 (10–12)	0.46 (12–14)	0.46 (14–16)
0.05	0.44 (4–5)	0.44 (5–6)	0.47 (6–7)	0.48 (7–8)	0.47 (8–9)	0.45 (9–10)	0.45 (10–12)	0.47 (12–15)	
0.025	0.44 (4–5)	0.45 (5–6)	0.47 (6–7)	0.47 (7–8)	0.47 (8–10)	0.46 (10–12)	0.45 (12–14)		

Table VI. Lorenz System with $\sigma = 10$, $r = 28$, $b = 2.6$

Time-one-tenth map	
Values of n	Entropy estimate
4–6	0.274
6–8	0.284
8–10	0.267
10–12	0.249
12–14	0.221
14–16	0.190
16–18	0.168
18–20	0.161

means the results should be multiplied by a factor of 10 in order to compare them with the estimates obtained in the previous section. Table VI shows an example of what happens. The numbers are decreasing more slowly than our patience.

4. POSSIBILITIES FOR FURTHER STUDY

1. It would be nice to devise a length growth algorithm which could be applied to time series. Something similar to the way in which Lyapunov exponents are calculated from such data[4] might work.
2. As noted at the end of Section 2.1, the algorithms described here are sensitive to the particular situation in which they are being applied. It would be helpful to develop methods which would be generally applicable.
3. Several notions of entropy, exponents, etc., exist (for example, ref. 1). A better understanding of how they are related may lead to useful algorithms, as well as deeper mathematical insight. A similar remark can be made regarding the relation of topological entropy to Hausdorff dimension.[5]

5. SOME MATHEMATICAL COMPLEMENTS

In certain situations it is possible to estimate the difference between

$$\frac{1}{n}\log^+ |f^n \circ \gamma|$$

and $h(f)$. As we shall see, the typical error is $O(1/n)$.

Suppose f is a C^∞ diffeomorphism of a compact manifold M, and Λ is a compact, isolated, invariant, uniformly hyperbolic set which is topologically transitive. An isolating neighborhood U of Λ is a compact neighborhood such that $\bigcap_{n\in Z} f^n U = \Lambda$. It is known that such neighborhoods always exist. Given a set A and a positive integer n, let $W^s(A, n) = \{x \in M: f^j x \in A \text{ for } j \in [0, n)\}$. Thus, $W^s(A, n)$ is the set of points which remain in A for time 0 through $n-1$. The quantity $V_n = |[f^{n-1} \circ \gamma | \gamma^{-1}(W^s(U, n))]|$ is the volume of the $(n-1)^{\text{th}}$ iterate of the points of γ which are in $W^s(U, n)$.

Theorem 3. With f and Λ as above, suppose γ is a C^∞ disk of the same dimension as the unstable manifolds of Λ, which meets some stable manifold of Λ transversely at some point. Then there are constants $C_1 > 0$ and $C_2 > 0$ such that for every positive integer n

$$\frac{C_1}{n} \leqslant \left| \frac{1}{n} \log^+ V_n - h \right| \leqslant \frac{C_2}{n}$$

The proof uses Markov partitions and is fairly elementary. Here is a sketch.

Let us use the notation $a_n \sim b_n$ to mean that there are constants $C_1 > 0$, $C_2 > 0$ independent of n such that $C_1 < a_n/b_n < C_2$ for all n. After a finite number of iterates, the disk γ is close to a disk W in the unstable manifold of a point x in Λ. Let $S_n = |[f^{n-1} \circ W | W^{-1}(W^s(U, n))]|$. Then, $V_n \sim S_n$. If $\mathscr{R}$ is a Markov partition for Λ,[3] then the elements of $\mathscr{R}$ (rectangles) form the symbols of a finite-state Markov shift which exhibits Λ as a finite-to-one factor. The number q_n of words of length n starting at a fixed symbol satisfies $q_n \sim e^{hn}$. We may assume that the orbit of x never hits the stable or unstable boundaries of the elements of $\mathscr{R}$. Then, if q_n is the number of words of length n which begin at the rectangle which contains x, it is not difficult to show that $S_n \sim q_n \sim e^{hn}$, which is the statement of the theorem.

The next result generalizes one in ref. 11.

A diffeomorphism is called *dissipative* on a set A if, in some metric, the absolute value of the Jacobian determinant of its derivative is less than one at each point in A. Given A and a curve γ, we let

$$G(\gamma, A) = \limsup_{n \to \infty} \frac{1}{n} \log^+ |[f^n \circ \gamma | \gamma^{-1} W^s(A, n)]|$$

Thus, $G(\gamma, A)$ is the logarithmic growth rate of the part of the iterates of γ which remain in A for $0 \leqslant j < n$.

Theorem 4. Suppose f is a dissipative diffeomorphism of a two-dimensional manifold M, and $\mathscr{R}$ is a compact smooth submanifold with nonempty boundary $\partial\mathscr{R}$. Let $\Lambda = \bigcap_{n\geqslant 0} f^n\mathscr{R}$ be the largest f-invariant set in $\mathscr{R}$. Then

$$h(f \mid \Lambda) = G(\partial\mathscr{R}, \mathscr{R})$$

Remarks. 1. The previous theorem has applications to various forms of polynomial diffeomorphisms of the plane $\mathbf{R}^2$. If $f(x, y) = (g(x) + by, x)$, where $g(x)$ is an even polynomial in the variable x and $|b| < 1$, then there is a rectangle $\mathscr{B}$ such that $h(f \mid \bigcap_{n\in\mathbf{Z}} f^n\mathscr{B}) = G(\gamma, \mathscr{B})$, where γ is the top of $\mathscr{B}$. In many cases, it may be true that the entropy of the largest invariant set in $\mathscr{B}$ equals $G(\gamma, \mathscr{B})$, where γ is a small line segment in the unstable manifold of one of the fixed points of f.

2. At present we do not have a rigorous useful error bound on the difference between the estimates of $(1/n)\log^+|f^n\circ\gamma|$ given by algorithms 1 and 2 and the actual value of this quantity. Nevertheless, empirically, we seem to obtain reasonable estimates for the entropy in many cases.

3. Similarly, we have no meaningful error estimates for the Takens method.

ACKNOWLEDGMENTS

The work of S.N. is partially supported by DARPA. T.P. is an Alfred P. Sloan Research Fellow, partially supported by grants NSF-DMS-8603355 and NSF-DMS-880-7141.

REFERENCES

1. D. Bessis, G. Paladin, G. Turchetti, and S. Vaienti, Generalized dimensions, entropies, and Liapunov exponents from the pressure function for strange sets, *J. Stat. Phys.* **51**:109–134 (1988).
2. L. Block, J. Keesling, S. Li, and K. Peterson, An improved algorithm for computing topological entropy, *J. Stat. Phys.* **55**:929–939 (1989).
3. R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Springer-Verlag, Berlin, 1975).
4. J.-P. Eckmann, S. Oliffson-Kamphorst, D. Ruelle, and S. Ciliberto, Liapunov exponents from time series, *Phys. Rev. A* **34**:4971–4979 (1986).
5. A. Fathi, Expansiveness, hyperbolicity and Hausdorff dimension, *Commun. Math. Phys.* **126**:249–262 (1989).
6. H. S. Greenside, A. Wolf, J. Swift, and T. Pignataro, Impracticality of a box-counting algorithm for calculating dimensionality of strange attractors, *Phys. Rev. A* **25**:3453–3456 (1982).
7. M. Hénon, A two-dimensional mapping with a strange attractor, *Commun. Math. Phys.* **50**:69–77 (1976).

8. E. J. Kostelich and H. L. Swinney, Practical considerations in estimating dimension from time series, in *Chaos and Related Nonlinear Phenomena*, I. Procaccia and M. Shapiro, eds. (Plenum Press, New York, 1987).
9. E. N. Lorenz, Deterministic nonperiodic flow, *J. Atmos. Sci.* **20**:130–141 (1963).
10. S. Newhouse, Entropy and volume, *Ergod. Theory & Dynam. Syst.* **8**:283–299 (1988).
11. S. Newhouse, Continuity properties of entropy, *Ann. Math.* **129**:215–235 (1989).
12. N. Packard, J. Crutchfield, J. Farmer, and R. Shaw, Geometry from a time series, *Phys. Rev. Lett.* **45**:712–716 (1980).
13. T. Pignataro, Feasibility of calculating dimension and topological entropy, in *Dynamical Systems and Chaos* (Springer-Verlag, Berlin, 1983).
14. J. C. Roux, A. Rossi, S. Bachelart, and C. Vidal, Representation of a strange attractor from an experimental study of chemical turbulence, *Phys. Lett.* **77A**:391–393 (1980).
15. F. Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence, Warwick 1980* (Springer-Verlag, Berlin, 1981).
16. S. Vaienti, Computing pressure for Axiom A attractors by time series and large deviations for the Lyapunov exponent, *J. Stat. Phys.* **56**:403–413 (1989).
17. A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, Determining Lyapunov exponents from a time series, *Physica D* **16**:285–317 (1985).
18. Y. Yomdin, c^k-resolution of semialgebraic mappings. Addendum to volume growth and entropy, *Israel J. Math.* **57**:301–317 (1987).
19. Y. Yomdin, Volume growth and entropy, *Israel J. Math.* **57**:285–300 (1987).