Limitations of delay reconstruction for chaotic dynamical systems

George G. Malinetskii, Alexey B. Potapov, and Alexey I. Rakhmanov
M. V. Keldysh Institute of Applied Mathematics, Miusskaya sq. 4, Moscow, 125047, Russia
and Department of Applied Mathematics, Russia's Open University, Moscow, Russia
(Received 16 September 1992)

We consider the problems arising in the application of algorithms of fractal-dimension measurements to data obtained from numerical integration of equations describing Rayleigh-Bénard convection. It is shown that in some cases the delay reconstruction of an attractor from a scalar time series requires much more data than does processing in the original phase space. If the length of the time series is too small, then the result resembles that for the case of large noise.

PACS number(s): 05.45.+b, 47.27.−i

I. INTRODUCTION

The invention of the method of delay reconstruction of strange attractors from scalar time series provided a rich perspective for new techniques of data processing. For the first time, we believe, this method was published in a widely circulated journal in 1980 by Packard *et al.*, and the word "theorem" was associated with it in the famous paper by Takens in 1981 [1]. Since then the application of the methods of dynamical systems theory (measurement of fractal dimensions, Lyapunov exponents, etc.) became a standard procedure of data processing such as spectral or correlation analysis.

An application of the reconstruction technique is as follows. Let there be a series of values of some observable (velocity, temperature, etc.) measured in sequential moments of time $t_i = i\Delta t$: $x_i = x(t_i)$, $i = 1,\ldots,N$. Then it is supposed that (1) the series $x(t)$ is a "projection" of a trajectory $\mathbf{x}(t)$ of a certain n-dimensional dynamical system with continuous $[\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})]$ or discrete $[\mathbf{x}(t+1) = \mathbf{F}(\mathbf{x}(t))]$ time, $x(t) = \Phi(\mathbf{x}(t))$; (2) this trajectory belongs to the system's attractor $\mathcal{A}$ and is dense on it, that is, eventually (for $t \to \infty$) it will pass arbitrary close to any given point of the attractor; and (3) the total observation time $T_{\text{obs}} = N\Delta t$ and the number of data points N are large enough for a trajectory to display all important details of the attractor studied. The Takens theorem also requires that attractor must belong to a smooth d-dimensional manifold. Then one can obtain the m-dimensional reconstruction $\mathcal{A}_R$ of $\mathcal{A}$ as a set of vectors $\mathbf{z}(t)$ in $\mathbb{R}^m$ for $m \geq 2d+1$:

$$\mathbf{z}(t) = \Lambda_m(\mathbf{x}(t)) \equiv \{x(t), x(t+\tau), \ldots, x(t+(m-1)\tau)\} . \quad (1)$$

The theorem guarantees that this reconstruction defines the mapping $\Lambda_m\ \mathcal{A} \to \mathcal{A}_R$, which is smooth and invertible on $\mathcal{A}_R$ for almost all τ. Hence, at least for very large N, we obtain a set of $\mathbf{z}$ vectors that may be processed instead of the original (and often unknown) $\mathbf{x}$ vectors.

But in practice N is always finite, and the parameters of reconstruction turn out to be vital: for a good reconstruction one can obtain more information than from a bad one using the same amount of data. We encountered this problem when analyzing the numerical data from modeling of convection in a fluid layer: the results were very sensitive to the choice of m and τ. We decided to investigate this problem on a quantitative basis, and it proved that the delay reconstruction may introduce various geometrical distortions; to overcome them and to "reach" the structure of attractor itself one needs N larger than in the case of processing $\mathbf{x}$ vectors. Moreover, in some cases it proves that there may be no delay reconstruction for which $\mathcal{A}_R$ is close to $\mathcal{A}$. In such cases, processing in $\mathbf{x}$ space may be preferential compared to the use of the reconstruction (1). It also proves (at least approximately) that the most important parameter is the window length covered by the $\mathbf{z}$ vector, we shall denote it by $w = (m-1)\tau$.

Below we shall characterize reconstructions in terms of measurement of one of the fractal dimensions—the correlation exponent ν. First, because it is natural to use a geometrical characteristic to describe geometrical distortions. Second, it was during the measurement of ν that this problem became important for us. But certainly, all basic conclusions are also valid for measurements of Lyapunov exponents, etc., because other methods use the same delay reconstructions.

II. FRACTAL-DIMENSION MEASUREMENTS

It is well known now that when the behavior of a dynamical system is chaotic, its attractor turns out to be a fractal set of noninteger dimension. Many different dimensions have been proposed to describe them, their values typically being close to each other and characterizing different details of the whole complex structure [2]. Numerically, the set of generalized dimensions D_q can be measured most reliably [3,4]. One of them—the correlation dimension D_2, sometimes denoted by ν—was proposed earlier than most others, and its definition is a bit simpler. Here we shall restrict ourselves to ν, but generalization can be made for any other D_q.

All the necessary definitions can be found, e.g., in [3,5–8,18]; here we shall use the one most close to practical estimation of ν from a set of N vectors. It may be

both original vectors $\mathbf{x}\in\mathcal{A}$ or reconstructed vectors $\mathbf{z}$, but first we shall consider only $\mathbf{x}$. Let us denote the number of attractor points $\mathbf{x}_j$ within the ϵ ball centered at $\mathbf{x}_i$ by $k_i(\epsilon)$. Then the probability that an attractor point would fall into this ball (the measure of the ball) can be approximated by $P_i(\epsilon,N)=k_i(\epsilon)/N$. This probability averaged on the attractor is called the correlation integral and can be approximated by $C(\epsilon,N)=N^{-1}\sum_i P_i(\epsilon,N)$ [in fact, $C(\epsilon,N)$ is just the ratio of the number of pairs of points with the distance $\|\mathbf{x}_i-\mathbf{x}_j\|<\epsilon$ to the total number of pairs]. If we denote the mean number of ϵ neighbors of a point by $k(\epsilon)=\sum_i k_i(\epsilon)/N$, then $C(\epsilon)=k(\epsilon)/N$. [Below we shall omit the argument N for $C(\epsilon)$.] The definition of the correlation exponent ν is based upon the scaling properties of this averaged probability,

$$\nu=\lim_{\epsilon\to 0}\{\ln[\lim_{N\to\infty} C(\epsilon)]/\ln\epsilon\}\ , \tag{2}$$

that is, for ϵ small $C(\epsilon)$ scales as ϵ^{ν}.

In practice N is always finite and both limits $\epsilon\to 0$ and $N\to\infty$ in (2) become senseless. The usual way to estimate ν from a finite set of vectors is to find scaling for ϵ "not too small," and for this purpose a linear dependence of $\ln C(\epsilon)$ vs $\ln\epsilon$ is searched for, the slope of corresponding graph being taken as a ν estimate. On this log-log plot as a rule there are three characteristic scales: (1) the size of the attractor $\epsilon_G=\max\|\mathbf{x}_i-\mathbf{x}_j\|$ [for $\epsilon>\epsilon_G$, $C(\epsilon)\equiv 1$] (global scale); (2) the upper end of the linear part ϵ_F, where the fractal scaling begins with "good accuracy" (fractality scale). Usually the loss of linearity for $\epsilon>\epsilon_F$ is considered as an influence of attractor edges [3,8], where the number of ϵ neighbors is less than in the "middle" of the attractor; (3) the lower end of the linear part ϵ_U, below which the structure of the attractor remains unresolved (unresolved structure scale).

Our experiments and the $\ln C(\epsilon)$ plots published by other authors enable one to conclude that the value of ϵ_F is characteristic of a specific dynamical system and the geometry of its attractor. Experiments show that for many known systems ϵ_F is rather big and $\approx 0.1\epsilon_G$ (e.g., for the Lorenz model).

The accuracy of dimension measurements essentially depends on the length of linear part, i.e., $l=\ln\rho$, $\rho=\epsilon_F/\epsilon_0$. Experts (see [3,8–10] and references therein) suppose that ρ cannot be less than $\rho_{\min}\cong 2-3$ (for reliable results ρ must be ≈ 10), and propose the estimates of the minimal number of data points for the correct determination of ν. The heuristic arguments differ slightly from paper to paper, but all of them are based upon some assumptions about "homogeneity" of attractor. Real attractors almost never are homogeneous, but nonetheless these estimates are useful and give at least an qualitative picture of the situation. They can be set as follows.

There are three important characteristics of a data set: total observation time $T_{\rm obs}$, the number of data points N, and the sampling rate Δt. They are related by the equality $T_{\rm obs}=N\Delta t$. They must be considered together, because one of them is not enough: large $T_{\rm obs}$ with small N may be as bad as small $T_{\rm obs}$ with large N.

The possibility of studying attractor properties from data about only one trajectory is based upon the well-known property of chaotic attractors: a trajectory beginning in a point $\mathbf{X}$ will for some time wander along the attractor, but after time interval T_R it will return into the ϵ neighborhood of $\mathbf{X}$ [6–8]. The less ϵ is, the more T_R must be. For the same ϵ, T_R is usually different for different attractor points, but nonetheless some mean value can be introduced. To measure ν, for $\epsilon\in[\epsilon_U,\epsilon_F]$ the attractor points must have enough ϵ neighbors; it is clear that the dimension ν cannot be obtained if the number of neighbors is less than ν. For different points the number of ϵ neighbors is different, but it is natural to suppose that ϵ_U corresponds to the situation when on average an attractor point would have about one neighbor within the ϵ_U ball (or box), i.e., when $k(\epsilon_U)\cong 1$. The corresponding $T_{\rm obs}$ must be such that for most points the trajectory will return into their ϵ_U neighborhood 1–2 times. Thus, if we cover the attractor $\mathcal{A}$ by ϵ_U boxes (their number we shall denote by M), most cubes will be visited by the trajectory 1–2 times.

For the given $T_{\rm obs}$ there is an optimal number of data points which is close to M (it is necessary to have 1–2 points in every box). Because the speed $|\dot{\mathbf{x}}|$ varies along the trajectory, for homogeneous t_i ordering the density of $\mathbf{x}_i$ will not be homogeneous, but we shall neglect it. If $N\ll M$, the scale resolution will be limited by N instead of $T_{\rm obs}$, and in case of $N\gg M$ on small scales $\epsilon<\epsilon_U$ the algorithm will measure the dimension of trajectory rather than attractor, unless special precautions are taken [3].

Consequently, to resolve the scale ϵ_U we need $N\cong M$, which corresponds to the optimal sampling rate Δt. The total length of the trajectory corresponding to the value of $T_{\rm obs}$ must be close to $M\epsilon_U$, and taking the mean speed $v=\langle|\dot{\mathbf{x}}|\rangle$ we get an estimate $\Delta t\approx\epsilon_U/\langle|\dot{\mathbf{x}}|\rangle$. Below it is supposed that the sampling rate Δt is close to the optimal, and thus all estimates of N and $T_{\rm obs}$ are equivalent.

M can be roughly estimated as a ratio of ν-dimensional volume of the attractor V_ν to that in one box (ϵ_U^ν): $M\approx V_\nu/\epsilon_U^\nu$. The minimal N corresponds to the minimal possible value of $\rho_{\min}$: $\epsilon_U=\epsilon_F/\rho_{\min}$, so

$$N_{\min}\cong V_\nu/\epsilon_U^\nu=(V_\nu/\epsilon_F^\nu)\rho_{\min}^\nu\ . \tag{3}$$

If we denote $N_0=V_\nu/\epsilon_F^\nu$, which is nothing but the number of ϵ_F cubes needed to cover the attractor, we have

$$N_{\min}=N_0\rho_{\min}^\nu\ . \tag{4}$$

It is clear that on average a point will have $N/N_0=k(\epsilon_F)$ ϵ_F neighbors and thus $C(\epsilon_F)=N_0^{-1}$.

Note that the estimate (4) is based upon the assumption that $k(\epsilon_U)\approx 1$, and thus $C(\epsilon_U)\cong N^{-1}$. Experiments show that sometimes the linear part extends to smaller ϵ values, where $k(\epsilon_U)\ll 1$. Usually variations of slope for these ϵ become rather large and the corresponding ν estimate may not be reliable. But sometimes it is possible to use this ϵ interval, and there are different approaches for estimating $N_{\min}$ that take it into account. For example, Eckmann and Ruelle [10] propose the estimate based upon the assumption that the $\ln C$ plot is linear down to minimal possible ϵ, where $C(\epsilon)\approx N^{-1}$, $k(\epsilon)\approx N^{-1}$, which gives

$$N_{\min}^2 > N_0 \rho_{\min}^{\nu} . \quad (5)$$

Estimates such as (4) are said to be valid for algorithms of calculations of Lyapunov exponents. Certainly, there may be intermediate estimates between (4) and (5). In what follows we shall restrict ourselves to (4), but the main results will hold in case of any such $N_{\min}$ estimate.

III. QUALITY OF A DELAY RECONSTRUCTION

The properties of delay reconstructions were studied in several papers. For example, in [11] it was characterized by homogeneity of measure on attractor, and from this criterion the method of time delay choice based upon mutual information was proposed. Later, the redundancy criterion for the choice of m was proposed on this basis (we must add new dimensions until this gives us new information) [12].

But the estimates of $N_{\min}$ give a new tool for evaluating the quality of reconstruction. It is well known that a bad choice of reconstruction parameters results in geometrical distortions of the reconstructed set. For example, too small τ makes the $\mathcal{A}_R$ points concentrate near the diagonal $z=\{1,1,\ldots,1\}$, yielding an underestimated ν value; too big τ also may create problems.

The two factors in the estimate (4) have an obvious sense. The second is responsible for the processing of the fractal structure, and cannot change for any reconstruction, while the first one corresponds to the amount of "efforts" that are necessary to reach the fractality level. The more complex the global structure of an attractor is, the greater the prefactor N_0 will be, and thus it may be used as a quality measure.

Below we try to analyze quantitatively how "bad" or "good" a reconstruction is by comparing the minimal number of points required to obtain a correct result in original $\mathbf{x}$ space $N_{\min}(\mathbf{x})$ and reconstructed $\mathbf{z}$ space $N_{\min}(\mathbf{z})$. Our main result is that for chaotic systems with "too big" chaoticity (entropy) or too wide spectrum a delay reconstruction may be impractical even for optimal τ or w; in any case $N_{\min}(\mathbf{z}) \gg N_{\min}(\mathbf{x})$. If the measured time series data has low precision, then a delay reconstruction may be totally inapplicable, no matter what N or τ is chosen.

IV. EXAMPLE OF PROBLEMS WITH DELAY RECONSTRUCTION: ATTRACTOR OF CONVECTIVE FLOW

We encountered the problems with applying delay reconstructions during processing the numerical data obtained by Rodichev in modeling convection in a plane horizontal fluid layer heated from below and rotating about the vertical axis $0z$ with the angular velocity ω [13]. The set of equations describing it in the dimensionless form is

$$\frac{\partial \mathbf{V}}{\partial t} + (\mathbf{V}\cdot\nabla)\mathbf{V} = -\nabla p + \mathrm{Pr}\Delta\mathbf{V} + \mathrm{Pr}\,\mathrm{Ra}T\mathbf{k} + \mathrm{Ta}\mathbf{k}\times\mathbf{V} ,$$

$$\frac{\partial T}{\partial t} + (\mathbf{V}\cdot\nabla)T = \Delta T + (\mathbf{V}\cdot\mathbf{k}) ,$$

$$\mathrm{div}\mathbf{V} = 0 .$$

Here $\mathbf{V}$ is the fluid velocity, T is the temperature deviation, $\mathbf{k}$ is the unit vector of the z axis and Pr, Ra, and Ta, are Prandtl, Rayleigh, and Taylor numbers, respectively. The boundary conditions for $z=0$ and $z=1$ were $T=V_z=\partial V_x/\partial z=\partial V_y/\partial z=0$, and the solutions periodic in x and y were searched for. The solution was represented in the form

$$Q_i(t,x,y,z) = \sum_{n,m,l=1}^{L} A_{inml}(t)\sin(\beta_i + n\pi z) \times \sin(\gamma_i \pi m \alpha_1 x + \pi l \alpha_2 y) .$$

Here Q_i for $i=1,2,3,4$ stands for the three components of $\mathbf{V}$ and T, and $\beta_3=\beta_4=\gamma_1=\gamma_2=0$, $\beta_1=\beta_2=\gamma_3=\gamma_4=\pi/2$. With the help of the Galerkin method the system of differential equations was obtained, which was integrated numerically. For the case described below $L=2$ and the total number of Fourier modes used was 23: 7 for T and 6, 6, and 4 for x, y, and z components of $\mathbf{V}$, respectively.

The purpose of the computations was the study of transition to turbulence at small Prandtl numbers. The values of the parameters were $\mathrm{Pr}=0.025$, $\mathrm{Ta}=30$, while Rayleigh number Ra was the bifurcation parameter. At $\mathrm{Ra}=\mathrm{Ra}^c\cong 800$ the stationary solution $\mathbf{V}=0$, $T=0$ loses stability and convective rolls appear, which correspond to four nonzero fixed points in the system of ordinary differential equations (ODE's). For larger Ra a limiting cycle arose, and then a 2-torus. In the example below Ra was chosen slightly above the point of 2-torus collapse, $\mathrm{Ra}/\mathrm{Ra}^c=1.6$; thus the chaotic regime that arose was not expected to be too complex, despite the fact that the attractor "exploded" away from the preexisting torus.

Here we shall not discuss the physical relevance of the results. It will be enough to say that the processed time series was generated by a 23-dimensional dynamical system, and Rodichev supplied us with the recorded values of all 23 variables with the step $\Delta t=0.125$ (the time step of numerical integration was smaller to provide accuracy $\approx 10^{-6}$) and the data array contained $\approx 55\,000$ 23-vectors. The examples of time history and power spectrum for one of the temperature components T_{101} are shown in Fig. 1. The sampling rate seems to be enough to describe all main features of the process.

Some typical results of dimension measurements for $N=20\,000$ are shown in Figs. 2(a) and 2(b). It is seen that the slope at the most linear part of the log-log plot increases as the window length $w=(m-1)\tau$ grows, and eventually reaches values $\approx 8-10$, which seems to be improbable in this case. There were two conjectures about the cause of this phenomenon: "noise" of some nature in the numerical data or quality of delay reconstructions. To verify them we repeated the dimension measurements in $\mathbf{x}$ space using all 23 modes and avoiding delay reconstruction. The result is shown in Fig. 2(c) and yields the dimension estimate close to 5.

It is known that in a case of spectrum like in Fig. 1(b) there may be problems with dimension measurements [3] because the observation time may prove to be too small to resolve slow modes. Indeed, ν estimated from 10 000 sample points ($T_{\mathrm{obs}}=1250$) was ≈ 4.3. But for

$T_{obs}=2500$ and $T_{obs}=7500$ (20 000 points with $\Delta t=0.375$) the estimate in **x** space was close to 5.0–5.2.

Hence, the amount of data seemed to be enough to determine ν in **x** space and probably insufficient to do it with delay reconstruction. So, we made the estimates of N_{min} for different w.

V. RECONSTRUCTION FOR LARGE w

For large values of m, and hence w, the dependence of correlation integral C on the reconstruction parameters m and τ was studied in [14,15] with the purpose of measuring the generalized entropy K_2. The following relation was obtained:

$$\ln C(\epsilon,m,\tau)\cong a+\nu\ln\epsilon-K_2 w\ ,\quad a=\text{const}\ , \tag{6}$$

for ϵ small and $w=(m-1)\tau$ large enough (e.g., because m must be greater than ν and a trajectory during time w must visit a certain minimal number of ϵ partition boxes).

Let us denote ϵ_U for **x** and **z** space by $\epsilon_U(\mathbf{x})$ and $\epsilon_U(\mathbf{z})$, respectively. Because $\ln C(\epsilon_U)=-\ln N$ both in **x** and **z** space, we have for $\epsilon=\epsilon_U$

$$\nu\ln\epsilon_U(\mathbf{z})=-\ln N-a+K_2 w\ ,$$

$$\nu\ln\epsilon_U(\mathbf{x})=-\ln N-b\ ,$$

or

$$\nu\ln\epsilon_U(\mathbf{z})=b-a+\nu\ln\epsilon_U(\mathbf{x})+K_2 w\ ,$$

that is, $\epsilon_U(\mathbf{z})$ grows exponentially as

$$\epsilon_U(\mathbf{z})\sim\epsilon_U(\mathbf{x})\exp(K_2 w/\nu)\ . \tag{7}$$

As for ϵ_F, as well as ϵ_G, our experiments for the Lorenz model and the Henon mapping [16] and the results from other papers (e.g., see figures in [15]) show that it remains almost the same for all reconstructions provided w is large enough [i.e., $\epsilon_F(\mathbf{z})\cong\epsilon_F(\mathbf{x})$]. In fact, $\epsilon_F(\mathbf{z})$ and $\epsilon_G(\mathbf{z})$ grow as $\approx m^{1/2}$, which can be neglected in comparison with exponential growth of ϵ_U.

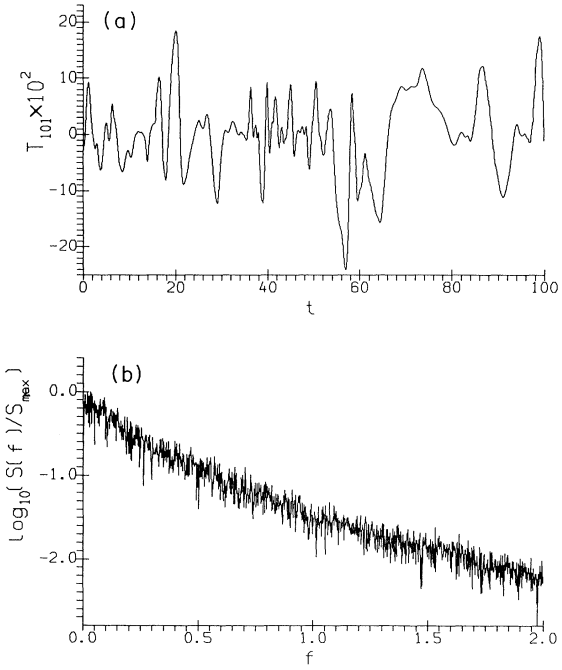

FIG. 1. The examples of (a) time history and (b) power spectrum for the numerical data from convection modeling (Fourier mode T_{101}).

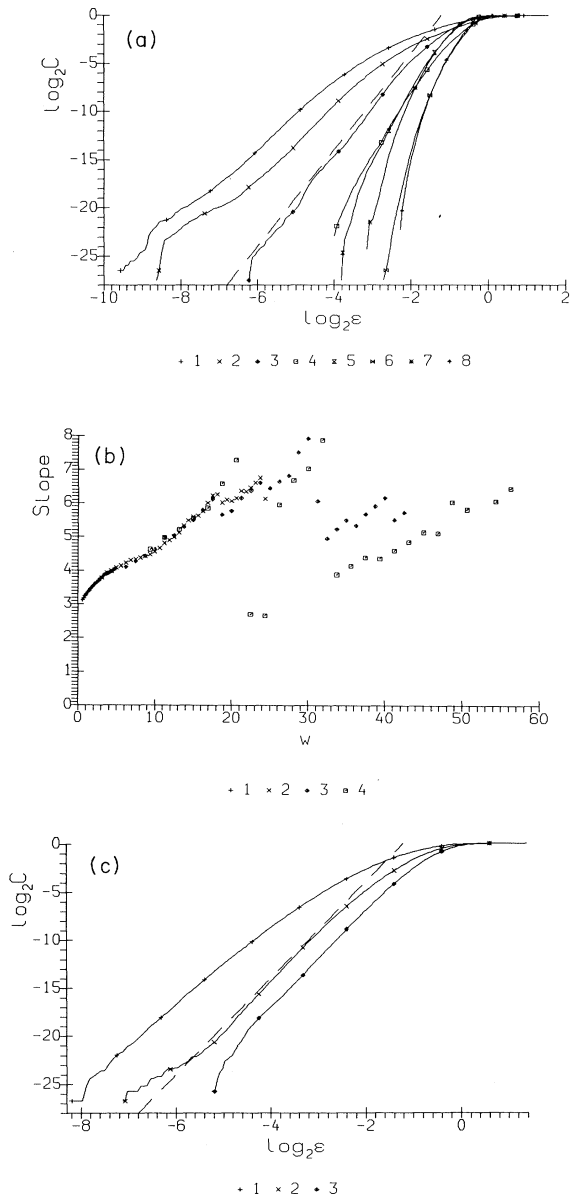

FIG. 2. The results of application of dimension algorithm to convection data. (a) Typical examples of log$C$-vs-logϵ plots; parameters: $m=20$ and 30 for every $\tau=0.125$, 0.625, 1.25, 3.75. (b) Dependence of slope vs w for the same τ values; for every τ, m takes 35 values from 6 to 40. (c) Calculations in **x** space using 11, 17, and 23 modes, respectively. The dimension value is estimated as $\nu\approx5$ (the slope of the dashed line). In all cases $\Delta t=0.25$ and $N=20\,000$.

From (4) we can relate $\epsilon_U(\mathbf{x})$ with n as

$$\epsilon_U(\mathbf{x})=\epsilon_F(\mathbf{x})(N_0/N)^{1/\nu}\cong\epsilon_F(\mathbf{z})(N_0/N)^{1/\nu} .$$

Substituting it into (7) we get

$$\epsilon_U(\mathbf{z})\sim\epsilon_F(\mathbf{z})(N_0/N)^{1/\nu}\exp(K_2 w/\nu)$$

or

$$N\approx N_0[\epsilon_F(\mathbf{z})/\epsilon_U(\mathbf{z})]^{\nu}\exp(K_2 w)$$

and thus the minimal N for the $\mathbf{z}$ space is

$$N_{\min}(\mathbf{z})\approx N_0\rho_{\min}^{\nu}\exp(K_2 w)=N_{\min}(\mathbf{x})\exp(K_2 w) . \quad (8)$$

As we suppose that the sampling rate Δt is optimal, the same holds for the observation times,

$$T_{\text{obs}}(\mathbf{z})\approx T_{\text{obs}}(\mathbf{x})\exp(K_2 w) . \quad (9)$$

This effect has the following consequence: when the true interval of linearity $[\ln\epsilon_U,\ln\epsilon_F]$ becomes too small or vanishes, one may search for linearity on the interval $[\ln\epsilon_F,\ln\epsilon_G]$, its length remaining almost constant. The slope $\mathcal{S}$ of the $\ln C$ plot on this interval must grow with w as

$$\mathcal{S}\approx S_0+K_2 w/(\ln\epsilon_G-\ln\epsilon_F) . \quad (10)$$

In fact, it is very probable that in cases of too small data samples the estimated value relates to the attractor's entropy rather than dimension (see also [17]).

All that has been said is shown in Fig. 3 for the Lorenz attractor ($\Pr=10$, $r=28$, $b=\frac{8}{3}$). For $w>6$ the remnants of the linear part $[\log_2\epsilon_U,\log_2\epsilon_F]$ are situated in the domain where $k(\epsilon)<1$. Here the largest Lyapunov exponent $\lambda\cong 0.9$, and K_2 is most probably close to it.

The growth of $\epsilon_J(\mathbf{z})$ can be explained from dynamical arguments. If we consider a pair of close points $\mathbf{x}_1$ and $\mathbf{x}_2$, the distance between them grows with time: $|\Delta\mathbf{x}(t)|\approx|\Delta\mathbf{x}(0)|\exp(\mu t)$. As w grows, so does the distance between reconstructed vectors $\mathbf{z}_1$ and $\mathbf{z}_2$: $|\Delta\mathbf{z}(t,w)|\approx|\Delta\mathbf{x}(t)|\exp(\mu w)$. The concrete μ value may change from point to point, but on average there must be exponential growth of $\epsilon_U(\mathbf{z})\cong\epsilon_U(\mathbf{x})\exp(Mw)$. Most probably, M must not exceed λ (after the largest Lyapunov exponent), and our experiments show that for moderate w M may be close to λ. More detailed consideration for larger w must give the value $M\cong K/\nu$.

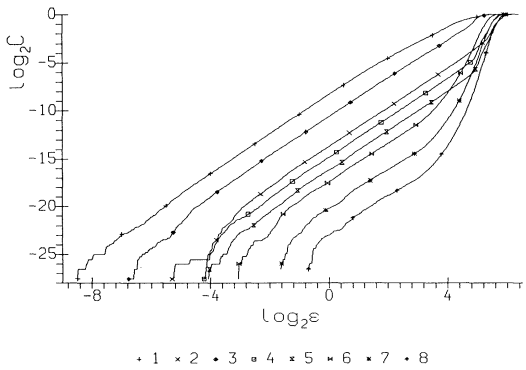

FIG. 3. Examples of correlation integral plots for the Lorenz attractor. The parameters of reconstruction are $m=4$ and 20 for $\tau=0.1$; $m=6,16,20$ for $\tau=0.2$; and $m=12,16,20$ for $\tau=0.5$, respectively. $\Delta t=0.1$, $N=10\,000$.

Increasing N, we can diminish $\epsilon_U(\mathbf{x})$ and compensate growth of $\epsilon_U(\mathbf{z})$. But if the measured data are superimposed by noise of amplitude δ, then it is useless to make $\epsilon_U(\mathbf{x})<\delta$ (unless one wants to determine δ itself). Thus the compensation capabilities are limited by δ, and if $\epsilon_U(\mathbf{z})$ grows more than ϵ_F/δ times then the compensation is impossible. Hence, for large w the structure of the attractor would be hidden by noise, so there is the natural limit for w: $K_2 w<\nu\ln(\epsilon_F/\delta)$. If this inequality cannot be satisfied, the delay reconstruction will be *inapplicable*.

VI. RECONSTRUCTION FOR SMALL w

It may seem that it is better to use small w values keeping T_{obs} the same. But if w becomes significantly less than some characteristic time w_0, the dimension measurement also requires $T_{\text{obs}}\gg T_{\text{obs}}(\mathbf{x})$ due to geometrical distortions caused by reconstruction—"condensation" near the diagonal $\{1,1,\ldots,1\}$.

Let us suppose that $x(t)$ is a smooth function and let w_0 be maximal for which the set of inequalities

$$\max_t[(w_0/2)^k|x^{(k)}(t)|/k!]\le x_A , \quad k=0,1,\ldots,m$$

holds, where $x_A=\max_t|x(t)-\langle x\rangle|$ is the "amplitude" of $x(t)$. For example, if $x(t)=\cos(\omega t)$, then $w_0=2/\omega=T/\pi$, for the Lorenz attractor (parameters as in Fig. 3) $w_0\approx 0.27$. Below we shall consider the case $w<w_0$.

The set of variables $\{\chi_k(t)\}$,

$$\chi_k(t)=(w_0/2)^k x^{(k)}(t)/k! , \quad (11)$$

usually can be considered as an alternative representation for a dynamical system. For example, for the Lorenz system

$$\chi_1=x , \quad \chi_2=(w_0/2)\Pr(y-x) ,$$

$$\chi_3=(w_0/2)^2\Pr[(\Pr+r-z)x-(\Pr+1)y]/2 .$$

At least in some cases this representation proves to be as good as the original $\mathbf{x}$ representation [16].

Let us introduce the set of linearly independent vectors, which may be considered as the values of u^k, $k=0,1,\ldots,m-1$, $u\in[-1,1]$, taken at m points on the grid with the step $h=2/(m-1)$:

$$Z_2=\{1,1,1,\ldots,1\} ,$$

$$Z_2=\{-1,(-1+h),(-1+2h),\ldots,1\} ,$$

$$Z_3=\{1,(-1+h)^2,(-1+2h)^2,\ldots,1\} ,$$

$$\vdots$$

$$Z_m=\{(-1)^{m-1},(-1+h)^{m-1},(-1+2h)^{m-1},\ldots,1\} .$$

Orthonormalization of $\{Z_k\}$ gives an orthonormal basis $\{\mathbf{v}_k\}$ in $\mathbf{z}$ space, and the components of $\mathbf{v}_k$ for m large

enough would be close to the values of the Legendre polynomials $P_{k-1}(u)$ in the corresponding grid points. In this basis the $\mathbf{z}$ vector will have the components

$$z_k(t)=(\mathbf{v}_k,\mathbf{z}(t))=\sum_i x(t_i)v_{ki} \approx \sum_i x(t+u_i w/2)P_{k-1}(u_i) .$$

For large m this sum can be approximated by the integral

$$z_k(t)\cong \int_{-1}^{1} x(t+uw/2)P_{k-1}(u)du . \quad (12)$$

Using the formula

$$P_n(u)=A_n\frac{d^n}{dx^n}[(u^2-1)^n] , \quad A_n=(n+\tfrac{1}{2})^{1/2}/(2^n n!)$$

and integrating (12) $k-1$ times by parts we obtain

$$\begin{aligned} z_k(t)&\cong (w/2)^{k-1}A_{k-1} \\ &\quad\times\int_{-1}^{1}(1-u^2)^{k-1}x^{(k-1)}(t+uw/2)du \\ &=\vartheta^{k-1}(w_0/2)^{k-1}A_{k-1}x^{(k-1)}(t^*) \\ &\quad\times\int_{-1}^{1}(1-u^2)^{k-1}du \\ &=\vartheta^{k-1}B_k\chi_k(t^*) , \end{aligned} \quad (13)$$

where $\vartheta=w/w_0$, $B_k=O(1)$, $t^*\in[t-w/2,t+w/2]$, and for w small $\chi_k(t^*)\cong\chi_k(t)$.

Thus the delay reconstruction $\mathcal{A}_R$ can be considered as an affine transformation of the attractor in χ representation, its size in the direction $\mathbf{v}_k$ being multiplied by ϑ^{k-1}.

If we suppose that the $\mathbf{x}$ and χ representations are equivalent and give similar estimates for $N_{\min}$, attractor volume, etc., then estimates for $\mathbf{z}$ reconstruction also can be made. The transformation of "ν-dimensional volume" V_ν would be

$$V_\nu(\mathbf{z})\approx V_\nu(\chi)\prod_{k(\le\nu-1)}\vartheta^k\approx V_\nu(\mathbf{x})\vartheta^\alpha , \quad \alpha=\nu(\nu-1)/2 . \quad (14)$$

But the minimal scale to be resolved diminishes even more because the small-w reconstruction decreases ϵ_F. Indeed, if in original representations the size of $\mathcal{A}$ in all directions was $\approx\epsilon_G$, and the influence of edges saturated for all directions on the same scale $\epsilon_F(\mathbf{x})$, then the smallest attractor size would be $\approx\vartheta^{\nu-1}$ times less and $\epsilon_F(\mathbf{z})\approx\epsilon_F(\mathbf{x})\theta^{\nu-1}$. The ϵ_U must be $\approx\rho_{\min}$ times less, and the ν-dimensional volume per ϵ_U cube is

$$\epsilon_U(\mathbf{z})^\nu\approx\epsilon_F(\mathbf{z})^\nu\rho_{\min}^{-\nu}\approx\epsilon_F(\mathbf{x})^\nu\rho_{\min}^{-\nu}\theta^{\nu(\nu-1)} =\epsilon_U(\mathbf{x})^\nu\vartheta^{\nu(\nu-1)}=\epsilon_U(\mathbf{x})^\nu\vartheta^{2\alpha} . \quad (15)$$

Thus for $w<w_0$ we get the following estimate:

$$N_{\min}(\mathbf{z})\approx N_{\min}(\mathbf{x})(w/w_0)^{-\alpha} , \quad \alpha=\nu(\nu-1)/2 \quad (16)$$

and

$$T_{\text{obs}}(\mathbf{z})\approx T_{\text{obs}}(\mathbf{x})(w/w_0)^{-\alpha} . \quad (17)$$

If the χ representation is worse than the $\mathbf{x}$ representation, this estimate may be even worse.

For $\epsilon>\epsilon_F$ the slope of the $\ln C$ plot would usually be less than ν. If ϑ is very small, then the dimensions on the $\ln C$ plot wold "turn on" sequentially. For $\vartheta\epsilon_G<\epsilon<\epsilon_G$ the attractor looks one-dimensional (1D), for $\vartheta^2\epsilon_G<\epsilon<\vartheta\epsilon_G$, 2D, and so on, until ϵ_F has been reached. If N is too small, ϵ_F may never be reached, and ν will be underestimated.

Consequently, w cannot be much less than w_0, especially for high-dimensional attractors; otherwise ν will be seriously underestimated. If the data are superimposed by noise with amplitude δ and $(w/w_0)^{\nu-1}<\delta$, the correct reconstruction again will be impossible for any N and T_{obs}.

VII. PROPERTIES OF RECONSTRUCTIONS AND FOURIER BASIS

The results of the preceding section can be interpreted from the point of view of spectral analysis if we use another natural orthonormal basis instead of $P_k(u)$: $y_k(u_n)=\exp(i\pi k u_n)$, $n,k=0,\ldots,m-1$. The projections of $\mathbf{z}(t)$ onto $\mathbf{y}_k$ are

$$\zeta_k(t)=(\mathbf{z}(t)\cdot\mathbf{y}_k) \cong\int_{-1}^{1}x(t-uw/2)\exp(i\pi k u)du . \quad (18)$$

This convolution works as a bandpass filter. Applying Fourier transform, we obtain

$$Z_k(\omega)=X(\omega)S_k(\omega) , \quad (19)$$

where

$$\begin{aligned} S_k(\omega)&=\int_{-1}^{1}\exp[i(\pi k-\omega w/2)u]du \\ &=2\sin(\omega w/2-\pi k)/(\omega w/2-\pi k) . \end{aligned}$$

Thus the kth component of $\mathbf{z}$ is formed mainly from the frequency band $[\pi(k-1)/w,\ \pi(k+1)/w]$ with the half width $\omega_B=\pi/w$. For w small the bands become very wide and 1–2 of them may cover the essential part of the spectrum $X(\omega)$. If the latter is concentrated within the limits $\omega_0=\pi/w_0$ (no processes faster than w_0) and $w<w_0$, then the signal is projected only onto the first component, others being very small. But for correct reconstruction of the attractor we need from ν to $2\nu+1$ different components, so we will have to use that with small amplitudes and to resolve very small scales, which requires large T_{obs}. In the presence of noise the small-amplitude projections will be formed mainly by it, and correct reconstruction becomes impossible.

VIII. APPLICABILITY OF DELAY RECONSTRUCTION

Combining (9) and (17) we get

$$T_{\text{obs}}(\mathbf{z})\approx T_{\text{obs}}(\mathbf{x})[(w/w_0)^{-\alpha}+\exp(K_2 w)] . \quad (20)$$

This function must have a minimum near w_0 and $1/K_2$, which corresponds to the best w choice. If $K_2 w_0\le 1$, then a good delay reconstruction can be obtained, which is true, e.g., for the Lorenz attractor ($K_2 w_0\cong 0.3$) and

some other low-dimensional chaotic systems.

The relation (18) is good when there is one characteristic time w_0, i.e., spectrum is "concentrated" near $\omega_0=2/w_0$. But when slow processes with characteristic time $w_1 \gg w_0$ substantially contribute to the attractor dimension [for example, if $x(t)=x_0(t/w_0)+x_1(t/w_1)$], the dependence of minimal $T_{\text{obs}}(\mathbf{z})$ on w may be more complex. It follows from the previous sections that for $w_1 \gg w \gg w_0$ the components associated with x_0 will be reconstructed satisfactory, while those associated with x_1 will contribute noticeably only to 1–2 **z** projections. Thus, until $w \ll w_1$, the delay reconstruction would cause distortions and require $T_{\text{obs}}(\mathbf{z}) \gg T_{\text{obs}}(\mathbf{x})$, but not so large as for $w < w_0$ [this intermediate region could probably be effectively described by a relation of the form $T_{\text{obs}} \cong \sum_k (w/w_k)^{-\alpha_k}$]. If $K_2 w_1 \gg 1$, then for $w \leq 2/K_2$ the small-w distortions will be large, and when they become small for $w \approx w_1$ the chaoticity causes large-w distortions of another nature. Thus $T_{\text{obs}}(\mathbf{z})$ must always be significantly greater than $T_{\text{obs}}(\mathbf{x})$, even for optimal w. It is possible, that there are some modifications of the delay reconstruction (1) that enable one to avoid such effects, but it is a separate problem which is beyond the frame of the present paper.

What would be observed had we possessed a time series with T_{obs} less than necessary for a correct delay reconstruction? If we increase w, then for small w the slope of the log$C$-vs-logϵ plot will grow because the slower and slower processes will "revive" and contribute to the calculated dimension. For $w > 1/K_2$ the slope will increase due to chaoticity. Thus we observe the permanent dimension increase, as shown in Fig. 2(b). To check this conjecture we have processed the time series

$$x(t)=x_L(t)+\sin(\omega_1 t)+\sin(\omega_2 t)\ , \tag{21}$$

$$\omega_1=2\pi/(4\times 2^{1/4})\ ,\quad \omega_2=2\pi/(16\times 3^{1/4})\ ,\quad \nu\cong 4.06\ ,$$

where $x_L(t)$ is the x variable for the Lorenz system, normalized such that $\max|x_L(t)|=1$. We processed a time series with $N=10\,000$, $\Delta t=0.1$. The results in Fig. 4 resemble that of Fig. 2(b). At the same time calculations in **x** space give the definite estimate $\nu\cong 4.01$, which is not bad for this N.

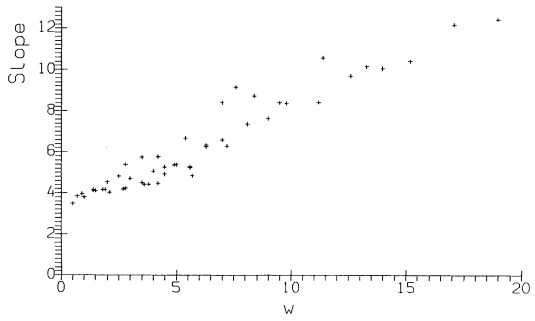

FIG. 4 The slope of the log$C$ plot for the model system (21). τ takes 10 values from 0.10 to 1.0, and for every τ, $m=6,8,10,15,20$. $\Delta t=0.10$, $N=10\,000$,

IX. ESTIMATE OF K_2 FROM ϵ_U

The dependence of ϵ_U on w has not only negative sides, it may serve as a source of information about chaotic properties of a dynamical system, such as $\ln C(\epsilon,w)$ for ϵ fixed does [14,15]. We tried to analyze this information for several systems described above. For the rigorous definition of ϵ_U for this purpose we used the relation $C(\epsilon_U)=1/N$. The plots of $\log_2\epsilon_U$ vs w for the Lorenz attractor (cf. Fig. 3), experimental series (22) (cf. Fig. 4), and the convection computations [cf. Figs. 2(a) and 2(b)] are shown in Figs. 5(a)–5(c), respectively.

In Fig. 5(a) it is clearly seen that for $w<1$ the slope of the plot (transformed from bits/sec to natural logarithms) gives the value $\mu\approx 1$, which is close to the value of largest Lyapunov exponent λ (the solid line with larger slope), and for larger w the slope gives the value close to 0.5,

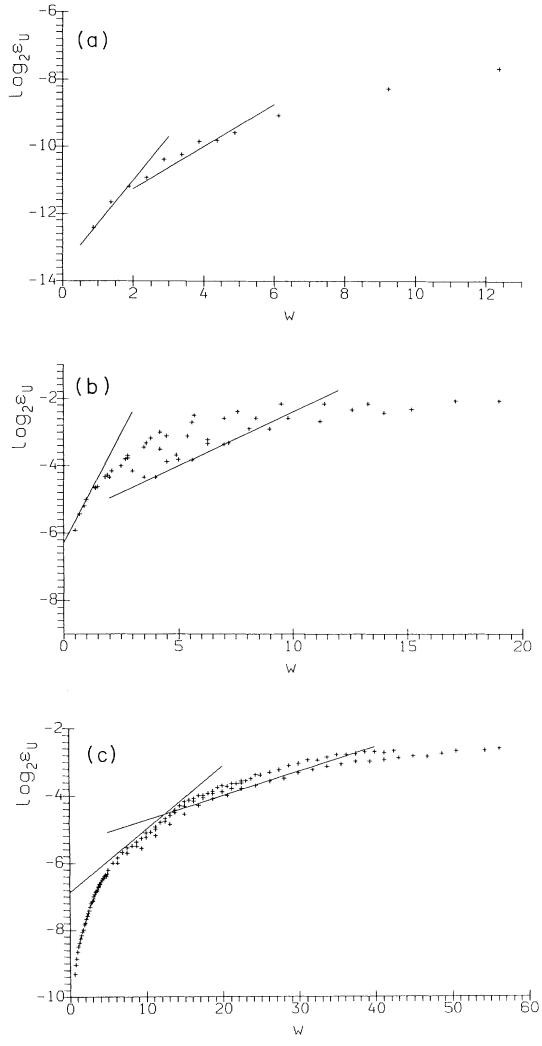

FIG. 5. The dependence of $\log_2\epsilon_U$ on w for (a) Lorenz system, the slopes of the solid lines $s_1=\lambda/\log 2$, $s_2=\lambda/(\nu\log 2)$; (b) the model system (21), $s_1=\lambda/\log 2$, $s_2=\lambda/(\nu\log 2)$; (c) convection data, $s_1=(0.15/\log 2)$, $s_2=(0.26/\log 2)/5.2$. The conjectured values $\lambda\approx 0.15$, $K_2\approx 0.26$.

which is in good agreement with the ratio $\lambda/\nu \cong K_2/\nu$ (another solid line).

In Fig. 5(b) there are similar w intervals, but here for very small $w < \sim 0.5$ the quick growth is observed, which most probably is connected with the small-w distortions (Sec. IV), and for very large w the $\log\epsilon_U$ plot saturates close to the value of $\log\epsilon_F$.

Figure 5(c) resembles 5(b), but here the two different linear parts can be seen more distinctly. Comparing Figs. 5(a) and 5(c), we may suppose that the slope of the first one corresponds to λ, and of the second to K_2/ν; then, using $\nu = 5.2$, we obtain $\lambda \approx 0.15$ and $K_2 \approx 0.26$. This can be interpreted as the presence of at least two positive Lyapunov exponents close to 0.1–0.2. On the other hand, it seems unlikely that there are more than two positive exponents.

We applied the method of Wolf *et al.* [19] to determine the largest Lyapunov exponent for the convection data. The obtained value was $\lambda = 0.15 \pm 0.03$.

Thus, summarizing the results about convection numerical data, we can conclude that (a) the system is chaotic; (b) the dimension of its attractor is between 4 and 6, and most likely $\nu \approx 5$; and (c) the system possesses two positive Lyapunov exponents, their values being close to 0.15; the entropy $K_2 \approx 0.26$.

X. CONCLUSIONS

First we must note that all the estimates of N or T_{obs} given here depend only on w. More accurate ones may (and for m small must) depend on m and τ separately, but the qualitative behavior is essential, which is in good agreement with our experiments.

All estimates for small w were done under the assumption that the attractor geometry in **x** space is optimal for processing. But obviously there may be situations when in **x** space the attractor is organized in a complex manner, and delay reconstruction creates "counterdistortions," which improves the situation. But it is doubtful that such situations occur frequently.

Thus we can conclude that for chaotic systems with rather broad spectra the delay method of attractor reconstruction may turn out to be impractical as requiring too many data. The fluid turbulence is a good candidate for it. To study it by methods of dynamical systems theory it may be better to use reconstructions that are closer to the original **x** representation, i.e., to measure several different observables simultaneously and to form reconstructed vectors from all of them for the same t or in several moments of time. The possibility of such reconstructions is well known; we want to point out their potential *practical* importance.

Here we must say that there may be two kinds of problems with application of dynamical-systems methods to numerical data. (i) One has limited T_{obs} (due to the feature of natural experiment or in case of very expensive computations), but one wants to obtain maximal information. Then it is necessary to measure several observables and to use the original phase space. (ii) One has practically unlimited observational capabilities, and is limited only by the speed of the data processing algorithm. Then delay reconstruction may be used, because there are the highly accelerated box-assisted algorithms that enable one to exclude the large-scale structure of the attractor and process only small scales $\epsilon < \epsilon_F$ (see [20] and references therein).

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

(1) For chaotic systems with broadband spectrum the delay reconstruction may require much more data than processing in original phase space. In any case one must have a very large amount of numerical data. This situation very likely occurs in the case of hydrodynamic turbulence.

(2) If data are superimposed by noise, there are situations when a delay reconstruction may be inapplicable for any time series length N.

(3) When the slope of $\log C$ plot grows as the embedding dimension increases, this effect may not be due to the presence of noise. It may mean that the observation time is too small.

ACKNOWLEDGMENT

We are grateful to E. B. Rodichev for many helpful discussions and the convection numerical data that he kindly gave to us.

[1] N. H. Packard, J. P. Crutchfield, J. D Farmer, and R. S. Shaw, Phys. Rev. Let. **45**, 712 (1980); F. Takens, in *Dynamical Systems and Turbulence,* edited by D. A. Rand and L.-S. Young, Lecture Notes in Mathematics, Vol. 898 (Springer, Berlin, 1981), pp. 336–381.

[2] C. Halsey, M. N. Jensen, L. Kadanoff, I. Procaccia, and B. I. Shraiman, Phys. Rev. A **33**, 1141 (1986); R. Badii and A. Politi, J. Stat. Phys. **40**, 725 (1985).

[3] J. Theiler, J. Opt. Soc. Am. **7**, 1055 (1990).

[4] K. Pawelzik and H. G. Schuster, Phys. Rev. A **35**, 481 (1987).

[5] P. Grassberger and I. Procaccia, Physica D **9**, 189 (1983).

[6] J. D. Farmer, E. Ott, and J. A. Yorke, Physica D **7**, 153 (1983).

[7] J. P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).

[8] G. Mayer-Kress, in *Directions in Chaos* (World Scientific, Singapore, 1987), pp. 122–147.

[9] D. Ruelle, Proc. R. Soc. London, Ser. A **427** 241 (1990); L. A. Smith, Phys. Lett. A **133**, 283 (1988); I. Procaccia, Nature **33**, 498 (1988).

[10] J.-P. Eckmann and D. Ruelle (unpublished).

[11] A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1131 (1986); W. Liebert and H. G. Schuster, Phys. Lett. A **142**, 107 (1989); A. M. Albano, J. Muench, C. Schwartz, A. I. Mees, and P. E. Rapp, Phys. Rev. A **38**, 3017 (1988).

[12] A. M. Fraser, Physica D **34**, 391 (1989).

[13] S. A. Gertsenstein, E. B. Rodichev, and V. M. Smidt,

Dokl. Akad. Nauk SSSR **238**, 545 (1978) [Sov. Phys. Dokl. **23**, 20 (1978)].
[14] P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983).
[15] A. Cohen and I. Procaccia, Phys. Rev. A **31**, 1872 (1985).
[16] P. Berge, Y. Pomeau, and C. Vidal, *Order Within Chaos* (Hermann, Paris, 1984).
[17] J. B. Ramsey and H.-J. Yuan, Nonlinearity **3**, 155 (1990).
[18] G. G. Malinetskii and A. B. Potapov, Journ. Vychisl. Matem. i Matem. Fiz. 28 (1988) 1021 (in Russian); M. Franaszek, Phys. Rev. A **39**, 5440 (1989).
[19] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, Physica D **16**, 285 (1985).
[20] Grassberger, P., Phys. Lett. A **148**, 63 (1990).