Symbolic Analysis of Chaotic Signals and Turbulent Fluctuations

M. Lehrman and A. B. Rechester
Institute of Nonlinear Science Applications, Sharon, Massachusetts 02067

R. B. White
Plasma Physics Laboratory, Princeton, New Jersey 08543
(Received 26 December 1995; revised manuscript received 6 September 1996)

The symbolic analysis introduced in this paper allows quantitative description of dynamical coupling between different time signals. In order to demonstrate how this method works we applied it to the explicit examples of chaotic signals. Our results appear to be quite robust when external noise is added. [S0031-9007(96)01857-1]

PACS numbers: 47.27.Eq, 05.40.+j, 05.45.+b, 47.11.+j

The purpose of this paper is to introduce the method for identification of fluctuations governed by the same dynamics. The methods of correlation functions and Fourier transform that are often used for the analysis of fluctuations in fluids, plasmas, etc., are inadequate for this purpose. In order to demonstrate how our method works we shall use the explicit examples of chaotic signals. The first example is presented in Fig. 1. These are the time records of $X(t)$ and $Z(t)$ generated by the Lorenz model [1] and are related to fluid velocity and temperature fluctuations in a simplified model of Bénard thermal convection. These two signals look quite different, but as we know they are representations of the one attractor, governed by the same dynamics. Very little could be learned from the study of the cross-correlation function $C_{xz}(\tau) = \langle X(t + \tau)Z(t)\rangle$ (here the averaging is done over the time t) or Fourier transforms of $X(t)$ and $Z(t)$. Using the symbolic analysis presented in this paper we will be able to demonstrate that signals $X(t)$ and $Z(t)$ correspond to the same dynamical process. There also exists a geometrical method which allows one to reconstruct the phase space dynamics from a single variable [2]. The advantages of our method are that it does not require the determination of dynamical system dimension and it works well in the presence of noise. A different method of symbolic analysis of noisy chaotic signals is presented in Ref. [3].

We begin with discretization of our signals,

$$X_n = X\,(t_0 + n\tau),\quad Z_n = Z\,(t_0 + n\tau). \tag{1}$$

Here $n = 0, 1, 2, \ldots.$ In the computations below we choose $\tau = 1$ because around this value our main result presented at Fig. 2 appears to be the most pronounced.

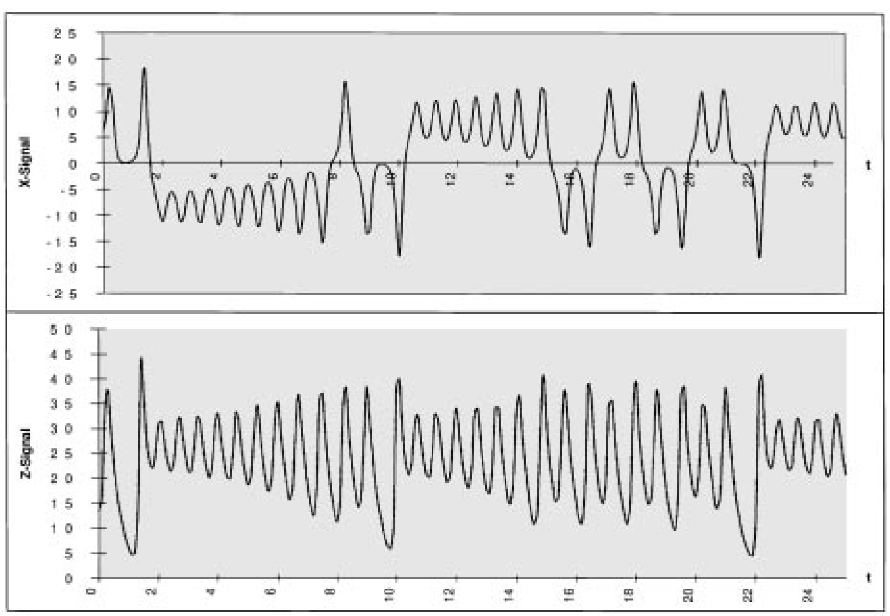

FIG. 1. Time records of $X(t)$ and $Z(t)$ generated by the Lorenz model.

 0031-9007/96/78(1)/54(4)$10.00

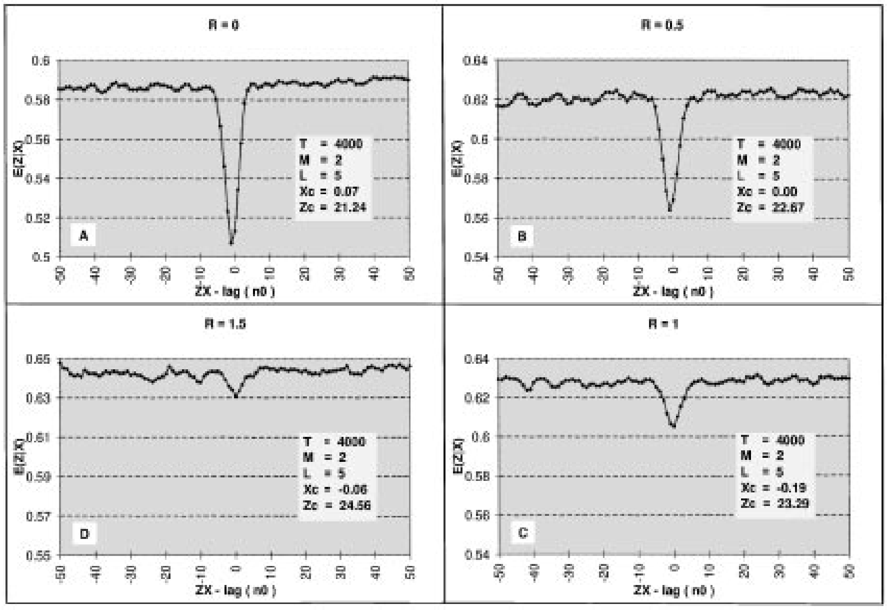

FIG. 2. Conditional entropy as a function of shift parameter n_0. Here $\tau = 1$ and noise ratio parameter $R = 0$ (a), $R = 0.5$ (b), $R = 1$ (c), and $R = 1.5$ (d).

In order to recognize time patterns in complex dynamical processes we need to have a language in which to expressexpress these patterns. That is, one has to substitute actual signals X_n, Z_n with their symbolic representation [4]. We will be using a simple example of symbolic dynamics, defined by the following symbols:

$$S_n = \begin{cases} \text{A (0) } X_{\min} < X_n < X_{C_1} \\ \text{B (1) } X_{C_1} < X_n < X_{C_2} \\ \text{C (2) } X_{C_2} < X_n < X_{C_3} < X_{\max} \\ \vdots \end{cases}, \tag{2}$$

The range of the variable X has been divided into domains, separated by the critical points X_{C_i}. In the computations below we will often use integer numbers instead of symbols. The coarseness of this representation makes it clear that a very small choice for the time step Δt would not make this number series contain more information; it would simply make each number repeat several times before changing. The number of critical points necessary to obtain a faithful symbolic representation of the dynamics depends on the system studied. We describe below how this is determined. The symbolic dynamics for variable Z is defined in a similar manner, with critical points $Z_{C_1}, Z_{C_2}, \ldots$.

The resulting long symbolic series we partition into short sequences of a given length L ($L = 5$ in the example below),

B | C | A B C | B | B | A B....

For ease of reference and identification it is convenient to identify every short sequence uniquely by just one integer [5,6].

$$\ell = \sum_{i=1}^{L} M^{L-i} S_i \,. \tag{3}$$

Here M is a number of different symbols, which we defined as integers according to Eq. (2).

The information content in the symbolic series can be quantified through the introduction of the entropy [7].

$$E = -\frac{1}{L} \sum_{\ell} P_\ell \ln P_\ell \,. \tag{4}$$

Here P_ℓ is the probability of finding a particular sequence ℓ that is the number of times this sequence can be found in the long symbolic time series divided by the number of all short sequences. The coding ability of the trial language defined by Eq. (2) depends strongly on the values of X_{C_i}. We performed a computer search to find out the optimum language which maximizes the entropy.

To maximize the entropy we introduce a given number of critical points (one, two, etc.) and study the entropy as a function of the placement of these critical points, by calculating the entropy for a given random placement of the X_{C_i}. A placement of these points is readily found which maximizes the entropy for a given number of critical points. If the number of critical points is too small, the symbolic language will, however, not contain the full information content of the signal. This is apparent by the fact that the entropy will increase substantially when a new critical point is added. When sufficient critical points are present, the symbolic signal will contain essentially all the information contained in the original signal, and this is reflected by the fact that the addition of another critical point does not increase the entropy. At this point the set X_{C_i} can be said to give a faithful symbolic representation of the original signal, and the symbolic language can be called optimum. As we will see later, only a rough approximation to the optimum language is required.

The sequences ℓ can be used for symbolic coarse graining of the phase space of the dynamical system [5,6]. Namely, to every sequence ℓ corresponds a cell volume of the phase space Δ_ℓ. In the case of degeneracy there can be several cells corresponding to one sequence ℓ. We found numerically that the better approximations to the optimum language correspond to the less degenerate coarse grainings.

Now we are ready to do comparative analysis of the X_n and Z_n data, presenting them as symbolic states $\ell_x(n)$, $\ell_z(n)$. If $X(t)$ and $Z(t)$ fluctuations are governed by different dynamics, then the evolution of $\ell_x(n)$ and $\ell_z(n)$ states is not correlated. Namely, every time the variable X occupies the state ℓ_0, the variable Z could occupy any of the states available to it. On the other hand, if $X(t)$ and $Z(t)$ are governed by the same dynamics, then we will observe the following relationship between $\ell_x(n)$ and $\ell_z(n)$: Everytime the variable X occupies the state ℓ_0, the variable Z can occupy only neighboring states. This is due to the fact that these states are just different symbolic coarse grainings of the same orbit [5,6]. We can easily destroy such a correlation by time shifting: $\ell_x(n), \ell_z(n + n_0)$. In order to check this effect we have computed the conditional entropy defined as

$$E(Z|X) = -\frac{1}{N_\ell} \sum_{\ell_x} \frac{1}{L} \sum_{\ell_z|\ell_x} P(\ell_z|\ell_x) \ln P(\ell_z|\ell_x). \quad (5)$$

Here $P(\ell_z|\ell_x)$ is a conditional probability for the variable Z to occupy state ℓ_z while the variable X occupies state ℓ_x, N_ℓ is the total number of different ℓ_x sequences; the first summation in Eq. (5) is done over all dynamically accessible ℓ_z states and fixed ℓ_x states. The result of these computations is presented in Fig. 2(a). The sharp minimum of the conditional entropy as a function of a shift parameter n_0 is a clear demonstration of the effect described above. For these computations we maximize entropy (4) with only one critical point ($X_c = 0.07, Z_c = 21.24$) while the full optimization requires approximately three critical points. On the other hand, without optimization, if we choose some symbolic language with a relatively low entropy, then the minimum in Fig. 2(a) becomes much less pronounced or disappears.

We have also studied the effect of external noise. The equation of motion for Lorenz's model [1] in the presence of additive noise can be written as

$$dX/dt = -aX + aY + \delta X, \quad (6)$$

$$dY/dt = -XZ + rX - Y + \delta Y, \quad (7)$$

$$dZ/dt = XY - bZ + \delta Z. \quad (8)$$

Here $a = 10$, $b = 8/3$, and $r = 28$ and the terms δX, δY, and δZ correspond to additive noise. We have solved Eqs. (1)–(3) by fourth-order Runge-Kutta numerical procedure with $\Delta t = 0.005$, $X_0 = 6.0$, $Y_0 = 6.0$, and $Z_0 = 13.5$. Every time step Δt we added to $X(t)$, $Y(t)$, and $Z(t)$ a random variable with a Gaussian distribution. The variances of all 3 random variables were equal and parametrized in the following way:

$$\sigma_x = \sigma_y = \sigma_z = R\sqrt{\langle(\Delta Y)^2\rangle},$$

$$\sqrt{\langle(\Delta X)^2\rangle} = 0.212; \quad \sqrt{\langle(\Delta Y)^2\rangle} = 0.325; \quad (9)$$

$$\sqrt{\langle(\Delta Z)^2\rangle} = 0.392.$$

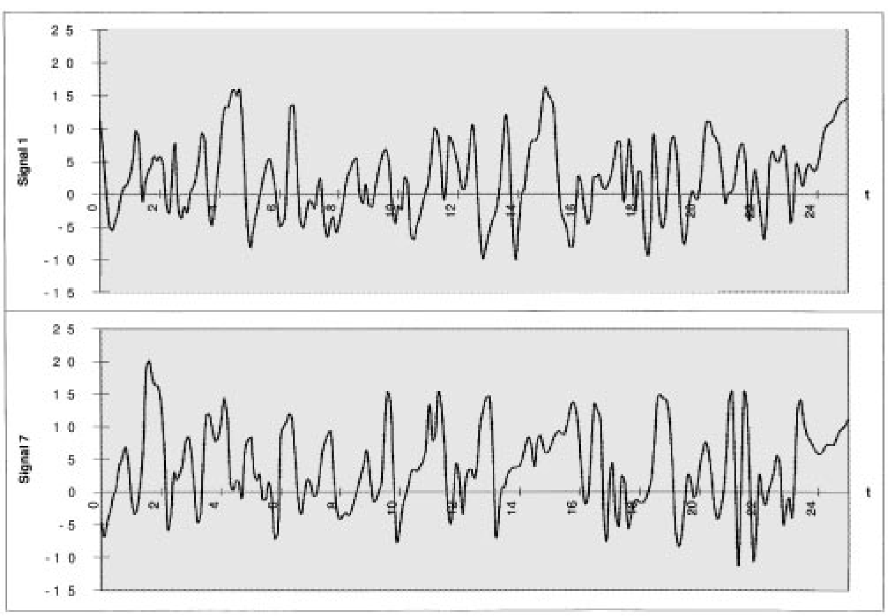

FIG. 3. Time records of signals 1 and 7 for the high dimensional Lorenz model $K = 10$, $M = 15$.

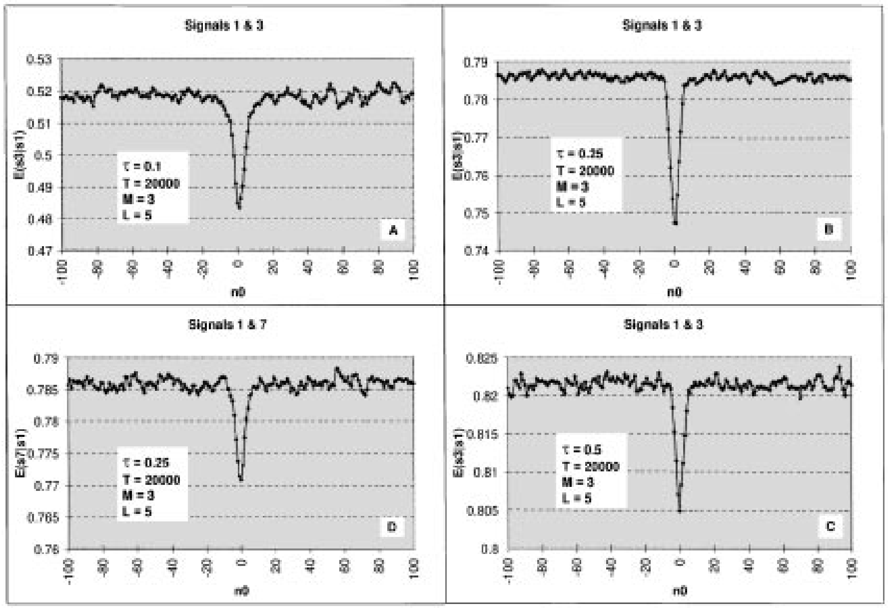

FIG. 4. Conditional entropy as a function of shift parameter n_0 for the high dimensional Lorenz model.

Here the average value has been computed along the orbits without noise. The Figs. 2(b)–2(d) correspond to noise ratio $R = 0.5$, 1, and 1.5. Thus we conclude that symbolic analysis presented in this paper appears to be quite robust in the presence of external noise.

In the conclusion of this paper we would like to describe briefly the application of our symbolic method for the analysis of a more complex chaotic signal presented in Fig. 3 and generated by a high dimensional model [8]. The equations of motion are

$$\frac{dX_k}{dt} = -X_{k-2}X_{k-1} + X_{k-1}X_{k+1} - X_k + F\,. \qquad (10)$$

Here k is an integer and variables X_k may be thought of as values of some atmospheric quantity in K sectors of a latitude circle with periodic condition $X_{k+K} = X_k$. Even though Eq. (10) is not much like those of the atmosphere, it models the basic physics of a turbulent system: The external forcing and internal dissipation are simulated by the constant F and linear terms, the quadratic terms simulate advection, which conserves the total energy of the system $X_1^2 + X_2^2 + \cdots X_K^2$. All variables in Eq. (10) have been scaled to reduce the coefficients in derivative, the quadratic, and linear terms to unity. Signals presented in Fig. 3 are the solutions of Eq. (10) for the values of parameters $K = 10$ and $F = 15$. To obtain these solutions we used fourth-order Runge-Kutta numerical procedure with $\Delta t = 0.01$. The results of application of the symbolic analysis to these signals are presented in Fig. 4. It is clear that symbolic method is working for this example as well.

A. B. R. is very thankful to professor Edward N. Lorenz for many useful discussion and encouragement. He would also like to acknowledge useful discussions and suggestions of Dr. Robert S. Granetz, Dr. Todd Evans, and Dr. E. Mazzucato. This work was partially supported by the U.S. Department of Energy under the Contract No. DE-FG0291ER54130.

Note added.—We would also like to report the first application of our method for the analysis of turbulent fluctuations measured in tokamak plasma with microwave reflectometry [9]. Our preliminary results are demonstrating that the method could be used for the analysis of real turbulent data [10].

[1] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).
[2] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).
[3] X. Z. Tang, E. R. Tracy, A. D. Boozer, A. deBrauw, and R. Brown, Phys. Rev. E **51**, 3871 (1995).
[4] A. N. Kolmogorov, Dokl. Akad. Nauk. SSSR **119**, 861 (1958); *ibid.* **124**, 754 (1959); Y. Sinai, Dokl. Akad. Nauk. SSSR **124**, 768 (1959)
[5] A. B. Rechester and R. B. White, Phys. Lett. A **156**, 419 (1991).
[6] A. B. Rechester and R. B. White, Phys. Lett. A **158**, 51 (1991).
[7] C. E. Shannon and W. Weaver, The Mathematical Theory of Communication (University of Illinois Press, Urbana, IL 1949).
[8] This model was kindly suggested to us by Professor Edward N. Lorenz as an example of many dimensional chaotic systems. It was first published in *Predictability, Proceedings of a Seminar at the European Center for Medium Range Weather Forecasts,* (Reading, England, 1995),Vol. I, pp. 1–18.
[9] E. Mazzucato and R. Nazikian, Phys. Rev. Lett. **71**, 1840 (1993).
[10] A. B. Rechester, M. Lehrman, R. S. Granetz, P. Stek, T. Evans, R. B. White, and E. Mazzucato, Bull. Am. Phys. Soc. **41**, 1457 (1996).