Direct Test for Determinism in a Time Series

Daniel T. Kaplan and Leon Glass
Department of Physiology, McGill University, 3655 Drummond Street, Montreal, Quebec, Canada H3G 1Y6
(Received 8 August 1991)

A direct test for deterministic dynamics can be established by measurement of average directional vectors in a coarse-grained d-dimensional embedding of a time series. Theoretical analysis of the statistical properties of a random time series using the same embedding technique is possible by consideration of classical results concerning random walks in d dimensions. Examples are given to show the clear differences between deterministic dynamics, such as may be generated by chaotic systems, and stochastic dynamics.

PACS numbers: 05.45.+b, 05.40.+j, 87.10.+e

Complex time series are ubiquitous in nature and in man-made systems, and a variety of measures have been proposed to characterize them. For example, power spectra [1] are particularly suitable for analysis of linear systems, where their interpretation is often transparent, whereas the dimension [2,3] and Lyapunov number [4] have been used to study geometrical and temporal properties of chaotic dynamics. However, none of these measures can be readily applied directly to determine if the dynamics are generated by a deterministic, rather than a stochastic, process. Here we propose a novel method to characterize a time series that is directed towards the analysis of whether the time series is generated by a deterministic system. The method is based on the observation that the tangent to the trajectory generated by a deterministic system [5] is a function of position in phase space, and therefore all the tangents to the trajectory in a given region of phase space will have similar orientations.

The method is illustrated in Fig. 1(a). The left-hand panel shows the x component of the deterministic chaotic Lorenz attractor [6] embedded in a two-dimensional phase space with the abscissa given by $x(t)$ and the ordinate by $x(t-\tau)$. This phase plane is coarse grained into a 16×16 grid, and each pass k of the trajectory through a box j of phase space generates a vector of unit length, called the *trajectory vector* $v_{k,j}$, whose direction is determined by the vector between the point where the trajectory enters the box and the point where it leaves the box, that is, the average direction of that pass of the trajectory in the box. The resultant vector from the vector addition of all the passes through each box, normalized by the number of passes n_j through the box, is $V_j=\sum_k v_{k,j}/n_j$. This gives a coarse caricature of the dynamics. In regions of the phase space where the vectors are well aligned in the 2D embedding, the resultant vector is almost of unit length. In the regions of phase space where trajectories cross, the length of the resultant vector is reduced. Now consider the results of performing the same procedure in a 3D phase space with coordinates of $x(t)$, $x(t-\tau)$, $x(t-2\tau)$. The phase space is coarse grained into a $16\times16\times16$ grid, and the resultant vector average V_j is determined and plotted as a projection onto the $x(t)$-$x(t-\tau)$ plane. Now the crossings of the trajectories are resolved and the geometry of the flow on the attractor is well approximated. In contrast, in Fig. 1(b), we consider embedding a synthesized random signal [7] that has an autorcorrelation function $\Psi(T)$ identical to that of the Lorenz attractor. Now there is a spaghetti mess that is not resolved in the 2D or 3D embeddings.

A statistical characterization of the dynamics can be generated by first determining the resultant vector V_j in each box j in phase space. The average value of this sum for a d-dimensional embedding of the dynamics is denoted $\bar{L}_n^d=\langle|V_j|\rangle_{n_j=n}$ where the angular brackets denote an average over all boxes containing n passes of the trajectory. Figure 2 shows $\bar{L}_n^d$ for the data displayed in Fig. 1. The strong correlations between the directions of the vectors in each box for the Lorenz attractor are reflected in the high values of $\bar{L}_n^d$ that are close to the maximal value of 1, for the 3D embedding. In contrast, for the randomized signal, $\bar{L}_n^d$ decreases approximately as $n^{-1/2}$.

In general, we wish to embed the signal in d dimen-

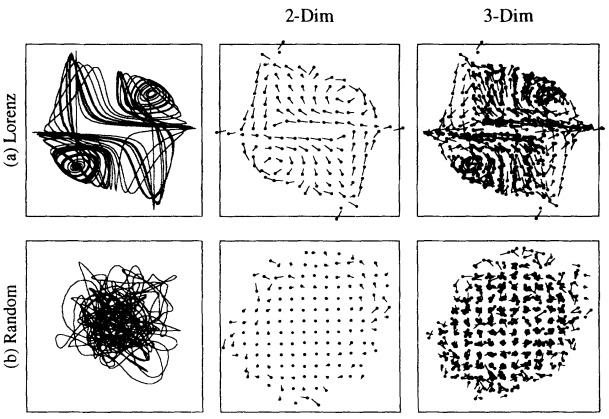

FIG. 1. An embedding plot $x(t-\tau)$ vs $x(t)$ and coarse-grained flow averages for the x component of Lorenz equations and for a random signal with the identical power spectrum. Embedding lag $\tau=0.75$. The length of the arrows shows the alignment of the trajectory vectors in the corresponding box, and the direction shows the direction of mean flow. Scales: (a) $[-20,20]$ and (b) $[-30,30]$. The signals shown in the trajectory plots cover 50 time units; the other plots cover 655 time units.

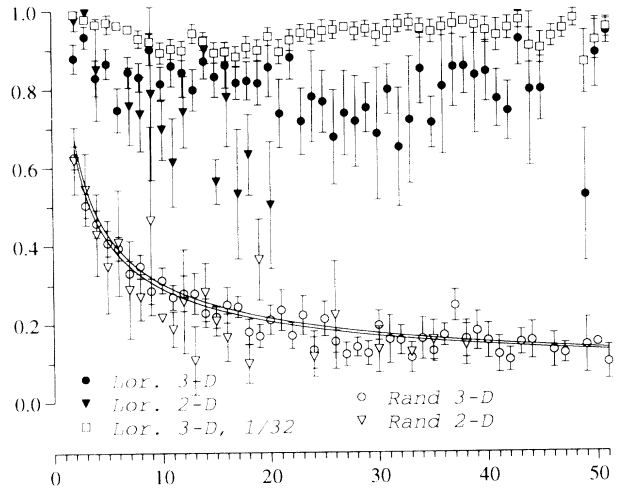

FIG. 2. $\bar{L}_n^d$ vs n for $d=2,3$ for cases shown in Fig. 1 for a signal length of 655 time units and gridding of 16^d. Error bars show the standard error in the estimate of the mean. The lines show the theoretical values computed from Eq. (1) for random flights in 2D and 3D. The open squares represent a 3D embedding with grid 32^3. $\tau=0.75$.

sions and to develop statistical criteria to distinguish deterministic dynamics from dynamics that involve stochastic processes. To carry out the statistical analysis, we consider random walks in d dimensions [8–11]. A random walk consists of n steps of unit length in d dimensions where the angle from each step to the next is chosen randomly. In this Letter we consider the average displacement per step, $\bar{R}_n^d$, which, for large n, is [11]

$$\bar{R}_n^d=\left(\frac{2}{nd}\right)^{1/2}\frac{\Gamma((d+1)/2)}{\Gamma(d/2)}, \tag{1}$$

where Γ is the gamma function. Thus, $\bar{R}_n^d=c_d n^{-1/2}$, where c_d equals $\pi^{1/2}/2$, $4/(6\pi)^{1/2}$, $3\pi^{1/2}/32^{1/2}$ for $d=2,3,4$, respectively, and $\lim_{d\to\infty}c_d=1$. Although Eq. (1) is valid only for the asymptotic limit when n is large, it agrees with the analytical values for $n=2$ to within about 3% [12]. In Fig. 2, the solid curves show the theoretical values of Eq. (1) for $d=2$ (lower curve) and $d=3$ (upper curve), in close agreement with the randomized time series.

To analyze a time series it is necessary to set three parameters: the time lag τ of the embedding, the number of dimensions d of the embedding, and the edge length ϵ of a box. For a system with an m-dimensional attractor, a necessary condition to embed the attractor is $d\geq m$, although hints of deterministic structure may be seen at smaller d. As ϵ decreases, longer data sets must be used in order to generate multiple passes through individual boxes of phase space. For high-dimensional attractors, the necessity of long data sets and the memory size of computers provide practical limitations on the utility of the method.

Choosing the value of the time lag in the embedding is a subtle issue that has been considered in the related question of the determination of the dimension of attractors. In this earlier work, suggestions for the time lag include a fraction of the first zero of the autocorrelation function [13] and the minimum of the mutual information function [14]. In the present case, there is an interplay between the autocorrelation function $\Psi(T)$ and the mean direction in each box. For Gaussian random processes [7] the structure of the coarse-grained embedded vector field can be approximated in terms of $\Psi(T)$. We consider the covariance matrix [15] of the $2d$-dimensional vector

$$\xi=[x(t),x(t-\tau),\ldots,x(t-(d-1)\tau),$$
$$\Delta x(t),\Delta x(t-\tau),\ldots,\Delta x(t-(d-1)\tau)],$$

where $\Delta x(t)=x(t+b)-x(t)$ and $b>0$ is the time scale for passage through a box. The first d components of ξ describe position in the phase space; the second d components describe the direction of trajectories through that position. The covariance matrix is written $\langle\xi^T\xi\rangle$, where the angular brackets denote the average over time, and can be written in terms of $\Psi(T)$. The off-diagonal elements

$$\langle\Delta x(t-j\tau)x(t-k\tau)\rangle=\Psi((j-k)\tau-b)-\Psi((j-k)\tau), \tag{2}$$

$$\langle\Delta x(t-j\tau)\Delta x(t-k\tau)\rangle$$
$$=2\Psi((j-k)\tau)-\Psi((j-k)\tau-b)-\Psi((j-k)\tau+b) \tag{3}$$

correspond to the first and second finite-difference derivatives of $\Psi(T)$ evaluated at multiples of τ. When they are zero, the directional elements of ξ are independent of each other and the positional elements of ξ. Although $\langle\xi^T\xi\rangle$ provides an accurate description of the statistical dependence between the components of ξ for a Gaussian random process, chaotic and other nonlinear systems have higher-order correlations that are not included in this formulation.

In order to show how the structure of the vector field changes with τ, it is convenient to construct a single number that summarizes the set of $\bar{L}_n^d$. One way of doing this is to construct a weighted average of V_j over all the occupied boxes

$$\bar{\Lambda}=\left\langle\frac{(V_j)^2-(R_{n_j}^d)^2}{1-(R_{n_j}^d)^2}\right\rangle. \tag{4}$$

For a deterministic system $\bar{\Lambda}=1$, while for a random walk $\bar{\Lambda}=0$.

In the Lorenz system (Fig. 3), $\bar{\Lambda}$ falls off slowly with τ—this is due to sensitive dependence on initial conditions destroying the deterministic connection between the

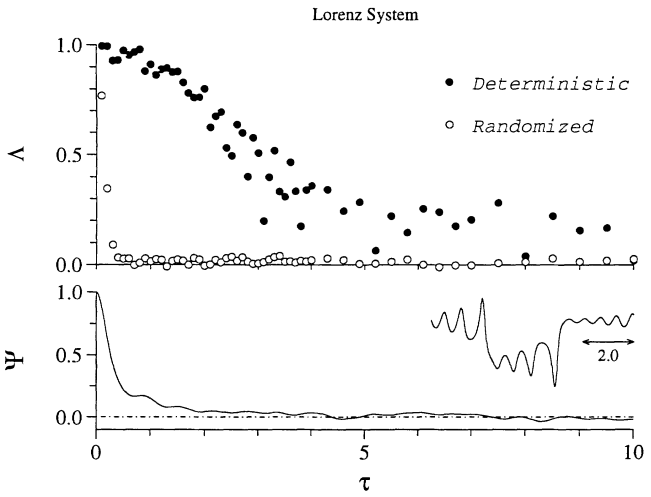

FIG. 3. $\bar{\Lambda}$ vs τ for a gridding of 32^3 for the signals of Fig. 1; $\Psi(\tau)$.

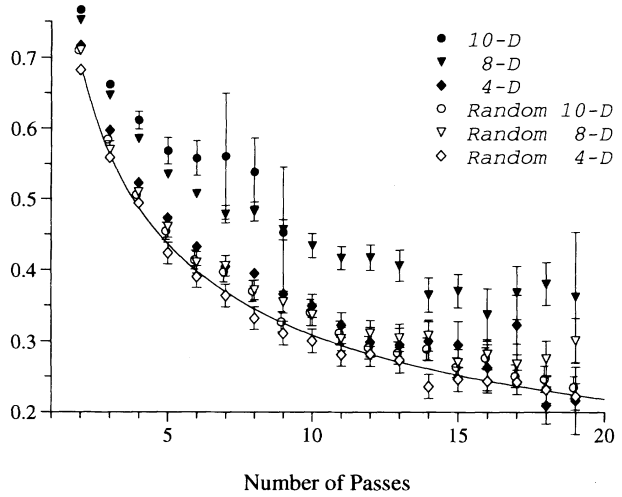

FIG. 4. $\bar{L}_n^d$ vs n for $d=4,8,10$ for $x(t)$ from Eq. (5), and for a random signal with the same $\Psi(T)$. $\tau=75$, near the first zero crossing of $\Psi(T)$. A gridding of 4^d was employed for $d=8,10$ and of 16^4 for $d=4$.

elements of ξ for large τ. For the randomized signal, $\bar{\Lambda}$ falls off similarly to $\Psi(T)$ [16].

We now consider the application of this technique to a high-dimensional system. Previous work on delay-differential equations has shown that, for long time delays, attractors of high dimension are found in the delay equation

$$\frac{dx}{dt}=\frac{ax(t-\delta)}{1+x(t-\delta)^c}-bx(t) \tag{5}$$

that has been proposed as a model for nonlinear feedback control in physiology [17]. For $a=0.2$, $b=0.1$, $c=10$, and $\delta=100$, Eq. (5) has an attractor whose estimated dimension is ~ 7.5 [3].

Figure 4 shows $\bar{L}_n^d$ for $x(t)$ and for a random time series with the same $\Psi(T)$, each of length 1.32×10^5 time units. The box edge length ϵ needs to be large to ensure that many boxes are traversed more than once by the trajectory. The coarse resolution causes $\bar{L}_n^d<1$ for the deterministic system, but even so, $\bar{L}_n^{10}>\bar{L}_n^8>\bar{L}_n^4$, indicating that the higher-dimensional embeddings are untangling the trajectory. In contrast, for the random time series, $\bar{L}_n^{10}\approx\bar{L}_n^8$. At the low resolution of coarse graining used for $d=8,10$, almost all the boxes are at the extremes of the trajectory cloud, where there is a directional bias towards the center even for the random signal. This causes the slight difference between $\bar{R}_n^d$ and the random signal's $\bar{L}_n^d$. Nonetheless, the $\bar{L}_n^d$ for the deterministic system are clearly above those for the randomized time series. This remains true when random noise is added to $x(t)$, even for signal-to-noise ratios as poor as 20 dB (for $d=10$). The statistic $\bar{\Lambda}$ shows a complex dependence on τ due to the delays contained in Eq. (5), but $\bar{\Lambda}$ for the deterministic system stays above $\bar{\Lambda}$ for the random system for τ as large as 200 [18].

The method sketched out in this Letter is directed at establishing whether a time series is generated by a deterministic system. Although in principle $\bar{L}_n^d=1$ in a deterministic system, this is observed only as $\epsilon\to0$. This limit on ϵ requires infinitely long data sets and would be defeated by even small amounts of measurement noise. In a practical situation with finite ϵ, two comparisons can be made: $\bar{L}_n^d$ can be compared to $\bar{R}_n^d$ to establish whether there is any evidence for a deterministic mechanism, and the values of $\bar{L}_n^d$ can be compared to those generated from $\langle\xi^T\xi\rangle$ to establish whether the determinism indicated by $\bar{L}_n^d$ goes beyond that which would be found in a randomly forced linear dynamical system. For the linear system, appropriate choice of τ and ϵ causes $\bar{L}_n^d\to\bar{R}_n^d$, so the two comparisons become equivalent. The fact that the method provides intrinsic estimates of the error in its statistic $\bar{L}_n^d$, coupled with its ability to analyze high-dimensional systems, should make the method useful in the analysis of complex dynamics from diverse sources.

We thank Hiroyuki Ito, Steven Skates, and Michael Marder for useful discussions. This research is supported by grants from the Natural Sciences and Engineering Research Council of Canada, the Heart and Stroke Foundation of Quebec, and les Fonds des Recherches en Santé du Québec. D.T.K. is a Postdoctoral Fellow of the North American Society of Pacing and Electrophysiology.

[1] E. O. Brigham, *The Fast Fourier Transform* (Prentice Hall, Englewood Cliffs, NJ, 1974).
[2] J. D. Farmer, Physica (Amsterdam) **4D**, 366 (1982).
[3] P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983); Physica (Amsterdam) **9D**, 189 (1983).
[4] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, Physica (Amsterdam) **16D**, 285 (1985).
[5] F. Takens, in *Dynamical Systems and Turbulence,*

Warwick, 1980, edited by D. A. Rand and L. S. Young (Springer-Verlag, Berlin, 1981), p. 366.

[6] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963). Parameters: $r=28$, $b=\frac{8}{3}$, $\sigma=10$.

[7] The randomized time series are generated by randomizing the phases of the discrete Fourier transform of the original signal. The resulting signals are examples of Gaussian random processes equivalent to passing Gaussian random noise through a linear filter. D. T. Kaplan and R. J. Cohen, Circ. Res. **67**, 886 (1990); J. Theiler, B. Galdrikian, A. Longtin, S. Eubank, and J. D. Farmer (to be published).

[8] Lord Rayleigh, Philos. Mag. **37**, 321 (1919).

[9] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1–89 (1943).

[10] K. V. Mardia, *Statistics of Directional Data* (Academic, London, 1972); K. V. Mardia, J. Appl. Stat. **15**, 115 (1988).

[11] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume II* (Wiley, New York, 1966). This result follows from the density function on p. 255 and is in accord with results in Refs. [8,9] for lower dimensions.

[12] Computations not shown give $\bar{R}_2^d=2^{d-2}\Gamma(d/2)\Gamma(d/2)/\pi^{1/2}\Gamma(d-\frac{1}{2})$. This equals $2/\pi$, $2/3$, $32/15\pi$ for $d=2,3,4$, respectively. $\bar{R}_2^d$ approaches $1/\sqrt{2}$ for $d\rightarrow\infty$ in accord with earlier calculations: J. M. Hammersley, Ann. Math. Stat. **21**, 447 (1950); R. D. Lord, Ann. Math. Stat. **25**, 794 (1954).

[13] A. I. Mees, P. E. Rapp, and L. S. Jennings, Phys. Rev. A **36**, 340 (1987); A. M. Albano, J. Muench, and C. Schwartz, Phys. Rev. A **38**, 3017 (1988).

[14] A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).

[15] Feller, Ref. [11], pp. 80–87.

[16] The dependence of $\bar{\Lambda}$ on τ is more complex in other situations. For example, dynamical systems such as the Rossler system or a lightly damped harmonic oscillator have $\Psi(T)$ that oscillate with an envelope that falls off only slowly to zero. Random signals with such $\Psi(T)$ show $\bar{L}_n^d\rightarrow\bar{R}_n^d$ when $\Psi(\tau)\rightarrow 0$ and when τ is near a maximum or minimum of $\Psi(T)$, particularly for small ϵ.

[17] M. C. Mackey and L. Glass, Science **197**, 287 (1977); L. Glass and M. C. Mackey, Ann. N.Y. Acad. Sci. **316**, 214 (1979).

[18] All calculations reported here were performed on a SUN SparcStation IPC with 8 Mb RAM. Details on the sorting algorithms that enable analysis of high-dimensional systems will be presented elsewhere. A typical run with a time series of 250000 points and 2^{20} boxes takes $\sim$100 CPU seconds.