PHYSICAL REVIEW E **66**, 021102 (2002)

Anisotropic diffusion and correlation analysis

Jacopo Bellazzini
Dipartimento di Ingegneria Aerospaziale dell'Universitá di Pisa, Via Caruso, 56100 Pisa, Italy
(Received 22 April 2002; published 9 August 2002)

A method of statistical analysis of the correlation between two given scale invariant sequences is proposed. The relation between the fractal dimension of a two-dimensional random walk, generated with jumps derived from the signals, and the scaling exponents of the sequences is investigated, and a well-defined relation is found in the case of statistically independent signals. The method of analysis, whose performance is described for the case of an intermittent map, might represent a new tool for the study of the correlation between coupled complex systems.

DOI: 10.1103/PhysRevE.66.021102 PACS number(s): 05.40.−a, 02.50.−r, 89.75.Da

I. INTRODUCTION

It is now well known that even very simple coupled nonlinear equations can generate random behavior and self-organized stuctures such as, for instance, vortices in fluids [1]. The diffusion approach, based on the pioneering contribution of Hurst and Mandelbrot (see, e.g., [2]), is a method to analyze such complex systems and to measure the long-range memory of a given sequence. This method allows the physical origin of fluctuations having finite variance to be explained and is able, in the case of steady systems, to detect the single or multiple scaling exponents and hence to recognize strange kinetics and anomalous diffusion [3,4].

The discovery of the widespread existence, both in Hamiltonian and dissipative systems [5,6], of stable processes with infinite variance (Lévy processes) encouraged the development of time-series analyses that do not involve the mean-square displacement. A specific method to deal with these systems is based on the evaluation of the scaling exponents via the Shannon entropy [7,8]. In the case of complex systems described by a couple of sequences $\{x_i\}_{i=1,N}, \{y_i\}_{i=1,N}$, it is important for the real understanding of the dynamics to analyze not only the average behavior of the coupling between the two sequences, but also its time evolution. In fact, the simple evaluation of the mean correlation function is clearly misleading in the presence of, for example, intermittent features. On the other hand, in the case of chaotic sequences dominated by a whole spectrum of frequencies, it is very difficult to detect time features using only tools based on the Fourier transform, such as the wavelet analysis (see, e.g., [9,10]). The aim of the present work is to overcome these limitations by deriving a new function of correlation C_q in the case of sequences characterized by a monoscaling property. This function, which is equal to zero in the case of statistical independence, estimates the correlation distribution between $\{x_i\}_{i=1,N}$ and $\{y_i\}_{i=1,N}$ by evaluating the inhomogeneity of the two-dimensional trajectory $(z_x(i),z_y(i))$ defined as

$$z_x(i+1)=z_x(i)+x_i,$$

$$z_y(i+1)=z_y(i)+y_i.$$

Specifically, we show that, if the signals are statistically independent, with scaling exponents δ_x and δ_y, and with finite mean value, the two-dimensional trajectory $(z_x(i),z_y(i))_{i=1,N}$ is monofractal with fractal dimension

$$D=\frac{\delta_x+\delta_y}{2\,\delta_x\delta_y}. \tag{1}$$

The new correlation function C_q is able to recognize the correlation evolution even in the case of chaotic maps or very noisy data where standard methods usually fail. As an example, we study a two-dimensional intermittent chaotic map, having a mean correlation $\langle x_i y_i\rangle$ close to zero.

II. THE DIMENSIONAL CORRELATION

We begin by recalling the diffusion entropy method. The main idea of the entropy method is remarkably simple and is based on the generalized central limit theorem [11]. We have to analyze a single signal $\{x_i\}_{i=1,N}$ and to detect, if it exists, the scaling exponent δ which is a measure of the long-range memory of the sequence. Many stationary phenomena are indeed characterized by a single scaling exponent; our analysis will be limited to just this class of time sequences. Let us consider an integer number t, with $1<t<N$, which will be referred to as "time." We thus generate $N-t+1$ "trajectories" considering the following sums:

$$z_j(t)=\sum_{i=j}^{i=j+t} x_i. \tag{2}$$

Therefore, the value $z_j(t)$ has to be considered the final position of a walker which jumped for t times. If $p(z,t)$ is the probability to be in position z after t jumps, we want to observe the "time" evolution of such a distribution. The monoscaling condition means that the probability follows the following rule:

$$p(z,t)=\frac{1}{t^{\delta}}F\left(\frac{z}{t^{\delta}}\right), \tag{3}$$

where F, due to the central limit theorem, is Gaussian- or Lévy-shaped. The diffusion entropy method evaluates the scaling exponent δ by calculating the Shannon entropy of the probability $p(z,t)$,

$$S(t)=-\int_{-\infty}^{+\infty}p(z,t)\ln(p(z,t))dz. \tag{4}$$

Indeed, for a distribution having scaling properties, it is very easy to demonstrate, by substituting Eq. (3) into Eq. (4), that $S(t)=A+\delta\ln t$, where A is the Shannon entropy of the generating function F. The diffusion entropy method is very useful to understand the dynamics when $\{x_i\}_{i=1,N}$ are N statistically independent random numbers with infinite variance. In fact, if the probability density function $p(x)$ of the random numbers goes to infinity as a power law, with $p(x)\approx x^{-\mu}$, the generalized central limit theorem, developed by Paul Lévy in the 1920s and 1930s, allows the relation between μ and the scaling exponent δ to be estimated:

$$\delta=\frac{1}{\mu-1}\quad\text{for}\quad 1<\mu<3,$$

$$\delta=0.5\quad\text{for}\quad\mu>3. \tag{5}$$

In an experimental situation we can evaluate both the power law of the distribution, by observing the statistics of the data, and the scaling exponent δ, by calculating the Shannon entropy. In the presence of a single scaling exponent δ, if the relations (5) are not satisfied, as for the human heartbeat [7], we can assume that the random data are strongly correlated with long-range memory.

From a different point of view, it is possible to put in relation the scaling exponent δ of the time series with the fractal dimension of the random walk,

$$z(i)=\sum_{j=1}^{i}x_j, \tag{6}$$

generated by the sequence. It is easy to show (see, e.g., [2]), by using the definition of moments, that if the signal $\{x_i\}_{i=1,N}$ has a scaling exponent δ, then the generated trajectory $[z(i)]_{i=1,N}$ is monofractal with a fractal dimension $D=1/\delta$. Indeed, if we evaluate the structure functions of the trajectory $[z(i)]_{i=1,N}$, we obtain the moments of the previously defined distribution $p(z,t)$,

$$\langle|z_j(t)|^q\rangle=\int_{-\infty}^{+\infty}|z|^q p(z,t)dz=t^{\delta q}A_q. \tag{7}$$

Remembering that in the fractal literature [12] the linear behavior of the structure functions means that the trajectory is homogeneous and hence monofractal, we can assume that signals $\{x_i\}_{i=1,N}$ with scaling exponents δ generate monofractal trajectories $[z(i)]_{i=1,N}$ with fractal dimension $D=1/\delta$.

Let us now consider two time series $\{x_i\}_{i=1,N},\{y_i\}_{i=1,N}$ whose correlation has to be characterized, and suppose that both of them satisfy individually the scaling equation (3), with scaling exponents δ_x and δ_y. Furthermore, it is assumed that $\langle x\rangle=\langle y\rangle=0$.

It may be shown (see the Appendix for details) that if the two series are statistically independent, the trajectory $[z_x(i),z_y(i)]_{i=1,N}$ is monofractal with fractal dimension

$$D=\frac{\delta_x+\delta_y}{2\,\delta_x\delta_y}. \tag{8}$$

This result shows that the fractal dimension of the anisotropic trajectory coincides with the reduced mass of a two-body system with masses given by the scaling exponents. Let us now apply the previous result to evaluate the fractal dimension of a two-dimensional trajectory originating from normal Brownian motion ($\delta=0.5$), fractional Brownian motion with Hurst exponent H ($\delta=H$), and Lévy flights with $2<\mu<3$ (because the random numbers should have finite mean value). It is clear that in the case of isotropic diffusion, when $\delta_x=\delta_y=\delta$, we obtain $D=1/\delta$. In the other cases we have

normal-fractional:

$$(\delta_x,\delta_y)=(0.5,H)\rightarrow D=1+\frac{1}{2H}, \tag{9}$$

normal-Lévy:

$$(\delta_x,\delta_y)=\left(0.5,\frac{1}{\mu-1}\right)\rightarrow D=\frac{\mu+1}{2}, \tag{10}$$

fractional-Lévy:

$$(\delta_x,\delta_y)=\left(H,\frac{1}{\mu-1}\right)\rightarrow D=\frac{\mu-1}{2}+\frac{1}{2H}, \tag{11}$$

Lévy-Lévy:

$$(\delta_x,\delta_y)=\left(\frac{1}{\mu_x-1},\frac{1}{\mu_y-1}\right)\rightarrow D=\frac{\mu_x+\mu_y}{2}-1. \tag{12}$$

However, the real interest of our result resides in the case in which we detect a deviation from the theoretical value, i.e., when the two time series are not independent. Each time series could, singularly taken, generate a diffusion process with scaling exponent δ_x,δ_y, but, due to the existence of a strong correlation between the two sequences, they can produce a trajectory in the two-dimensional Euclidean space with fractal dimension different from the theoretical one. Furthermore, it is also possible that the trajectory $[z_x(i),z_y(i)]_{i=1,N}$ is multifractal, see [12,13], due to the nonconstant behavior of the correlation. In the case of inhomogeneous fractals, the deviation from the homogeneity is given by the evaluation of the Renyi dimensions D_q, which estimate the moments of the distribution of the local scaling exponents [14]. Then we propose to evaluate the function

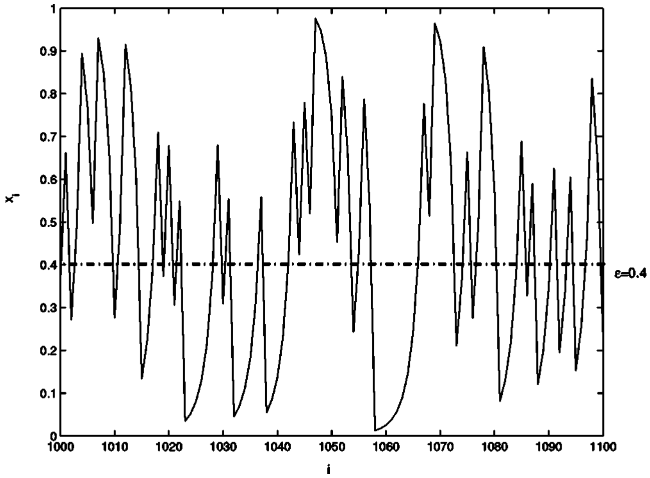

FIG. 1. The sequences x_i generated by the Manneville map with $z=1.2$. The value of ϵ is established when the two-dimensional signal is not correlated.

$$C_q=\frac{\delta_x+\delta_y}{2\,\delta_x\delta_y}-D_q\,, \tag{13}$$

where D_q are the Renyi dimensions, which can be evaluated numerically from the trajectory $[z_x(i),z_y(i)]_{i=1,N}$. In the case of two statistically independent signals, the function C_q is identically equal to zero for every value of the parameter q. When C_q is a constant function of q (monofractal trajectory) and different from zero, we have a homogeneous correlation between the sequences; indeed, the constancy of the function C_q means that the correlation between the sequences is not time-dependent. In the more interesting case in which the correlation between the sequences is time-dependent, we can distinguish different situations, such as, for instance, the presence of an intermittent correlation, by observing the behavior of the function C_q.

We will show an example of a two-dimensional chaotic map, in which the mean correlation $\langle x_i y_i\rangle$ is close to zero. In this map, the correlation between the sequences is highly intermittent and we show that our function C_q can measure such intermittency. We consider a two-dimensional map based on the Manneville map [15]. The Manneville map is an example of a discrete dissipative dynamical system with intermittency, an alternation between long regular phases, called laminar, and short irregular phases, called turbulent. The map, which is defined on the interval $I=[0,1]$, by the relation

$$x_{n+1}=x_n+x_n^z \quad (\mathrm{mod}1) \quad z>1, \tag{14}$$

has been used as a simple model displaying complicated behavior, as it may appear in fluid dynamics or in DNA sequences [16]. We analyze the correlation of a two-dimensional intermittent map where the variable x_i is ruled by the Manneville map and the variable y_i is a Gaussian distributed random number if $x_i>\epsilon$ and $y_i=x_i$ elsewhere. It is clear that the correlation between x_i and y_i is ruled by $\epsilon \in[0,1]$, which establishes the length of the laminar-correlated phase, see Fig. 1. Indeed, when the variable x_i is greater than ϵ, the variable y_i is a random number and hence is not correlated with x_i. In our specific example, by taking $z=1.2$, $\epsilon=0.4$, and $N=30\,000$, the mean correlation coefficent is $\langle x_i y_i\rangle/\sigma_x\sigma_y=-0.07$. In practice, the standard mean correlation analysis suggests that the sequences are, on average, not correlated. We show that our method is able to recognize the strong intermittent correlation. First we evaluate the scaling exponent δ_x,δ_y of the signals $\{x_i\}_{i=1,N},\{y_i\}_{i=1,N}$. Figure 2 shows that the scaling exponents are $\delta_x=\delta_y=0.5$. We have to evaluate now the function C_q of the trajectory $(z_x(i),z_y(i))_{i=1,N}$, see Fig. 3. We limit ourselves to evaluating the correlation dimension of the two-dimensional trajectory (the Renyi dimension D_2) via the Grassberger-Procaccia algorithm [17]. Figure 4 shows that the correlation dimension D_2 is equal to 1.74 ± 0.02 so that $C_2=0.26$, which is well removed from the theoretical value $C_2=0$ corresponding to statistical independence. The correlation dimension D_2 of the trajectory $(z_x(i),z_y(i))_{i=1,N}$ shows unambiguously that the sequences are strongly corre-

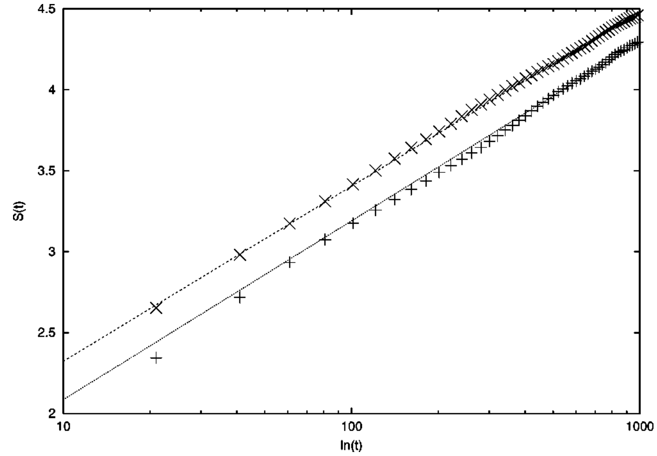

FIG. 2. Evaluation of the scaling exponents $\delta_x=0.5$ (bottom) and $\delta_y=0.5$ (top) of the intermittent map via the diffusion entropy method.

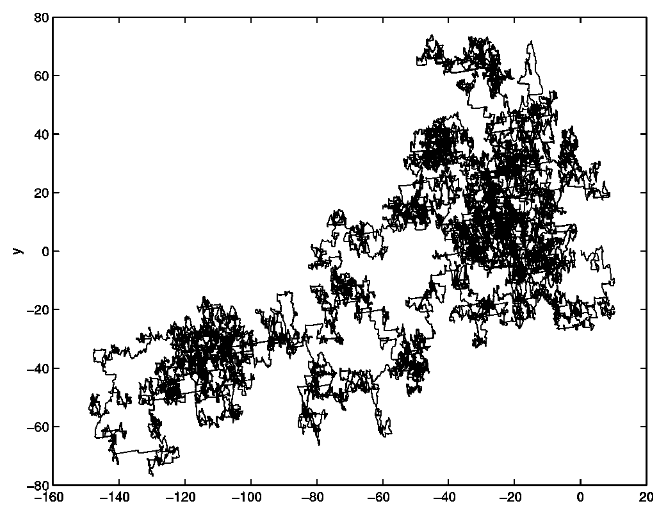

FIG. 3. Trajectory $(z_x(i),z_y(i))_{i=1,N}$ of the intermittent map with $N=30\,000$, $\epsilon=0.4$, and $z=1.2$.

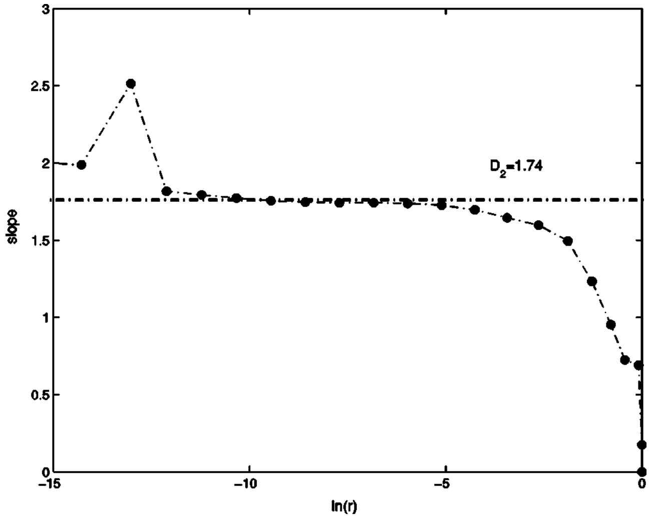

FIG. 4. Evaluation of the correlation dimension D_2 of the trajectory $(z_x(i),z_y(i))_{i=1,N}$ via the Grassberger-Procaccia algorithm. We obtain $D_2=1.74\pm 0.02$ and $C_2=0.26$.

lated, providing more complete information with respect to the standard mean correlation coefficient. A full characterization of the intermittency of this correlation is not the aim of the present example and might be obtained by measuring further Renyi dimensions and by analyzing the shape of the function C_q. The previous application shows that there exists a wide class of two-dimensional maps, and more in general of two-dimensional signals, for which a statistical approach based on the multifractal distribution of the two-dimensional random walk may be applied to overcome some limitations of the standard tools. The proposed method, although still unable to exactly localize a particular feature due to its statistical approach, may be useful to characterize the correlation distribution of this kind of complex system.

APPENDIX

We have to prove the relation described by Eq. (8). Let us consider two statistically independent sequences $\{x_i\}_{i=1,N}$ and $\{y_i\}_{i=1,N}$, which individually satisfy the scaling equations with scaling exponents δ_x and δ_y. Furthermore, it is assumed that both sequences have a null mean value. We generate two new sequences

$$z_x(i)=\sum_{j=1}^{i} x_j\,,$$

$$z_y(i)=\sum_{j=1}^{i} y_j\,,$$

and we define $p((i,j),l)$ as the cumulative probability that

$$|z_x(i)-z_x(j)|<l,$$

$$|z_y(i)-z_y(j)|<l,$$

and $p_x((i,j),l)$, $p_y((i,j),l)$ the probabilities that the previous relations hold for the single components. The probability to find points inside a box of side l centered in $(z_x(i),z_y(i))$ in the case of independent time series is given by the relation

$$P(i,l)=\frac{1}{N}\sum_j p_x((i,j),l)p_y((i,j),l).$$

The particular choice of $(z_x(i),z_y(i))$ is not important due to the assumption of independency, and we can consider $P(i,l)$ as the dot product between the probability vectors. Remembering that given two arbitrary vectors a,b in the N-dimensional Euclidean space the dot product can be easily evaluated as $\langle a,b\rangle=\|a\|\|b\|\cos\theta$, we obtain

$$P(i,l)=\frac{1}{N}\|p_x(i,l)\|\|p_y(i,l)\|\cos\theta,$$

where θ is the angle between the probability vectors. Now it is sufficient to observe that

$$\frac{1}{N}\|p_x(i,l)\|^2=\frac{1}{N}\langle p_x((i,j),l),p_x((i,j),l)\rangle\approx l^{1/\delta_x},$$

$$\frac{1}{N}\|p_y(i,l)\|^2=\frac{1}{N}\langle p_y((i,j),l),p_y((i,j),l)\rangle\approx l^{1/\delta_y},$$

to conclude that

$$P(i,l)=l^{1/2\delta_x}l^{1/2\delta_x}\cos\theta,$$

and hence

$$P(i,l)=l^{(\delta_x+\delta_y)/2\delta_x\delta_y+\epsilon},$$

where $\epsilon=\log_l(\cos\theta)$. The previous relation is the main result concerning the anisotropic diffusion. In the limit of $l\to 0$, we have $\epsilon\to 0$. Indeed, $\cos\theta$ is a number between 0 and 1 (always different from 0) and l is definitively less than 1; then $\lim_{l\to 0}\log_l(\cos\theta)=0$. Thus we have proved that given two statistically independent sequences with scaling exponents δ_x and δ_y, we can generate an anisotropic trajectory $(z_x(i),z_y(i))_{i=1,N}$ with fractal dimension given by the relation

$$D=\frac{\delta_x+\delta_y}{2\,\delta_x\delta_y}.$$

[1] M. Shlesinger, G. Zaslavsky, and J. Klafter, Nature (London) **361**, 31 (1993).

[2] B. B. Mandelbrot, *The Fractal Geometry of Nature* (W. H. Freeman and Company, New York, 1983).

[3] J. Klafter, M. Shlesinger, and G. Zumofen, Phys. Today **49** (2), 33 (1996).

[4] P. Castiglione, A. Mazzino, P. Muratore, and A. Vulpiani, Physica D **134**, 75 (1999).

[5] G. Zumofen and J. Klafter, Phys. Rev. E **47**, 851 (1993).
[6] J. Klafter and G. Zumofen, Phys. Rev. E **49**, 4873 (1994).
[7] P. Grigolini, L. Palatella, and G. Raffaelli, Fractals **9**, 439 (2001).
[8] N. Scafetta, P. Hamilton, and P. Grigolini, Fractals **9**, 193 (2001).
[9] S. Mallat, *A Wavelet Tour of Signal Processing* (Academic Press, New York, 1997).
[10] M. Farge, Annu. Rev. Fluid Mech. **24**, 395 (1992).
[11] W. Feller, *An Introduction to Probability Theory and Its Application* (Wiley, New York, 1971).
[12] G. Paladin and A. Vulpiani, Phys. Rep. **156**, 149 (1987).
[13] R. Benzi, G. Paladin, G. Parisi, and A. Vulpiani, J. Phys. A **17**, 3521 (1984).
[14] A. Renyi, *Probability Theory* (North-Holland, Amsterdam, 1970).
[15] P. Manneville, J. Phys. (Paris) **41**, 1235 (1980).
[16] M. Buiatti, P. Grigolini, and L. Palatella, Physica A **268**, 214 (1999).
[17] P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983).