*Review Article*
# Chaotic Time Series Analysis

## Zonghua Liu

*Institute of Theoretical Physics and Department of Physics, East China Normal University,
Shanghai 200062, China*

Correspondence should be addressed to Zonghua Liu, zhliu@phy.ecnu.edu.cn

Chaotic dynamical systems are ubiquitous in nature and most of them does not have an explicit
dynamical equation and can be only understood through the available time series. We here briefly
review the basic concepts of time series and its analytic tools, such as dimension, Lyapunov
exponent, Hilbert transform, and attractor reconstruction. Then we discuss its applications in a
few fields such as the construction of differential equations, identification of synchronization and
coupling direction, coherence resonance, and traffic data analysis in Internet.

## 1. Introduction

Chaotic dynamical systems are ubiquitous in nature such as the tornado, stock market,
turbulence, and weather. Their functions are different in different situations. For example,
in the case of tornado, the chaotic behavior is harmful to human beings and need to be
avoided or controlled. But in the case of the activities in human brain, the chaotic behaviors
are useful and necessary to sustain the normal functions of brain. Thus, it is an important task
to understand chaos and let it serve human society better.

The chaos has been studied for a long time. It is usually believed that Poincaré is the
first one who studied chaos. In the later 19th century, Poincaré studied the restricted three-
body problem where one body is negligible small, compared to the other two. Poincaré found
that the solution of this simple system is very complicated and cannot be given precisely. Then
in 1963, Lorenz revealed the "butterfly effect" in studying the weather prediction and is thus
recognized as the father of chaos. But the formal use of chaos is from the paper of "Period
three implies chaos" by Li and Yorke in 1975. After that, chaos have been widely studied and a
lot of important concepts has been introduced, such as the dimensions, Lyapunov exponents,
Fourier transform and Hilbert transform, and attractor reconstruction.

The most striking feature of chaos is the unpredictability of its future. This feature is usually called as the "sensitivity dependence on initial conditions" or "butterfly effect." For example, if two initial conditions have a small difference $\delta x$, their difference after time $t$ will be $\delta x e^{\lambda t}$ with $\lambda > 0$, that is, exponential separation. Thus, a tiny difference or even a cutoff error will be blown up quickly and results in a big difference in the near future. "Similar causes have similar effects" is invalid in chaotic systems except for short periods. A system is called chaotic system, provided that its maximum Lyapunov exponent is positive.

Mathematically, chaos can be produced by both discrete and continuous equations [1–6]. The discrete systems can be expressed as

$$\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n) \tag{1.1}$$

such as the Logistic map, Henon map, standard map, tent map, circle map, and Ikeda map. The continuous systems can be expressed as differential equation

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}(\mathbf{x}(t)) \tag{1.2}$$

with three or more degrees of freedom $\mathbf{x}(t) = [x_1(t), x_2(t), x_3(t), \ldots, x_m(t)]$. Typical chaotic flows are the Lorenz equation, Rössler equation, Duffing's equation, Chua's circuit, and so forth. The discrete map and continuous flow are not unrelated but have a close relationship. The discrete map (1.1) can be considered as a projection of flow (1.2) on a specific Poincaré surface of section. Letting $\delta t = 1$ and taking limit, (1.1) will be transformed into (1.2). On the other hand, (1.2) can be transformed into (1.1) by integration, that is, $\mathbf{x}_{n+1} = \mathbf{x}_n + \int_t^{t+T_n} \mathbf{F}(\mathbf{x}(t)) dt$ with $T_n$ being the returning time to the Poincaré section. A common point between the discrete and continuous systems is that both are deterministic. That is, each initial condition uniquely determines the time evolution. Thus, the chaos is usually called "deterministic chaos" and can be considered as some kind of randomness produced by deterministic system. A necessary condition for (1.1) and (1.2) to show chaos is that the functions $\mathbf{f}$ and $\mathbf{F}$ must be nonlinear.

There are infinite unstable periodic orbits in chaos, which form invariant sets [1, 2]. An invariant set is the image of itself under time evolution. Once a trajectory is located in these invariant sets, it will stay there forever. Comparing with the dense chaotic orbits, the invariant sets are seldom and its measure is zero. In experimental situations or in numerical simulations, as there is always noise or cutoff error, an arbitrary trajectory will never stay at these invariant sets. The invariant sets are thus not observed in general. Given a long enough time, a trajectory will lose memory on its initial condition and eventually settle on a restricted geometry, called stable attractor. Once a trajectory enters the attractor, it will stay in it forever if there is no external perturbation. The trajectory will go through all the phase points there with its nature measure except the unstable periodic orbits. That is, the attractor is also an invariant set. This property is called ergodicity in chaotic attractor.

Mathematically, there are a lot of measures such as the Lebesgue measure, Hausdorff measure, counting measure, and natural measure. Suppose that the relative probability for a continuous variable $x$ to appear at some location is described by the probability density

$P(x)$. The measure $\mu(D)$ is the probability to find $x$ in an area $D$, that is, $\mu(D) = \int_D P(x)dx$. The average of a function $f(x)$ is

$$\langle f(x) \rangle = \int f(x)P(x)dx = \int f(x)d\mu(x). \tag{1.3}$$

When the measure $\mu$ is an invariant measure and the average of function $f(x)$ on $\mu$ equals the average on time, $\mu$ becomes an ergodic measure with

$$\langle f(x) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T f(x)dt = \int f(x)d\mu(x). \tag{1.4}$$

This measure does not include those unstable periodic orbits in the attractor and is called as natural measure. To use this measure, one waits until all transients have been wiped off and looks for an invariant probability measure describing the distribution of typical orbits. Thus, in the natural measure, we say that the chaotic trajectories are ergodic. Based on this property, many quantities are more conveniently defined as averages over the natural measure in phase space. Because of the nonperiodicity of chaotic trajectory, the attractor is usually called as "strange attractor," in contrast to the attractors of fixed point and periodic orbits.

Practically, we usually do not have the dynamical equations (1.1) and (1.2), like the earthquake, brain and stock market, and so forth, thus the details of the system equations in the phase space and the asymptotic invariant set that determines what can be observed through experimental probes are unknown. What we can obtain is some time series from one or at best a few of the dynamical variables of the system, which poses a big challenge to characterize the chaotic systems. An interesting question is how one go from one or a few time series to the multivariate state or phase space which is required for chaotic motions to occur?

Fortunately, this problem has been intensively studied in the past decades and has now formed a subject called *data mining or data analysis* [7–11]. The basic assumption is that the measured time series comes from the attractor of the unknown system with ergodicity, which contains the information of the attractor. Then one can use the measured time series to figure out the properties of attractor such as its dimension, its dynamical skeleton, and its degree of sensitivity on initial conditions. The delay-coordinate embedding technique established by Takens [12] provides a practical solution to this task.

Suppose that an experiment is conducted and one time series $s_n = s(n\Delta t)$ is measured, where $\Delta t$ is the sampling interval. Such a time series can be, for instance, a voltage signal from a physical or biological experiment, or the concentration of a substance in a chemical reaction, or the amount of instantaneous traffic at a point/node in the Internet, and so on. In principle, the measured time series comes from an underlying dynamical system with (1.1) and (1.2), with or without the influence of noise. Most commonly, the time series is a consequence of scalar measurements of some quantity which depends on the current state of the system, taken at a fixed sampling time:

$$s_n = s(\mathbf{x}(n\Delta t)) + \eta_n, \tag{1.5}$$

where, more often than not, the measurement function $s$ is unknown as $\mathbf{F}$, $\mathbf{x}$ is the $m$-dimensional variable and $\eta_n$ is the measurement noise. Let us neglect the effect of noise at this level of presentation. Equation (1.5) becomes

$$s_n = s(\mathbf{x}(n\Delta t)), \quad n = 1, \ldots, N. \tag{1.6}$$

The total recording time is then $T = N\Delta t$. In general, $T$ should be sufficiently large so that the full dynamics are exhibited in the time series. This is what we have in hand. In the following we will show how to extract the information on the unknown attractor from (1.5) or (1.6).

## 2. Basic Concepts and Analytic Tools of Chaotic Time Series

To characterize the attractor, some basic concepts have been introduced such as dimensions and Lyapunov exponents [1, 2]. These quantities are invariant under the evolution operator of the system and thus are independent of changes in the initial conditions of the orbit, and both are independent of the coordinate system in which the attractor is observed. Therefore, we can evaluate them from experimental data. The advantage of these quantities is that each one of them will turn a sequence of data into a single number, which is convenient for comparison between different time series.

As an attractor is generally located in a finite area, its trajectory shows some kinds of oscillatory behavior, that is, recurrence. To show the oscillatory behavior from a given scalar time series, a convenient way is to transform it into a 2-dimensional vector by the Hilbert transform. To obtain the attractor of the corresponding time series, we have to reconstruct an auxiliary phase space by an embedding procedure, that is, the delay embedding technique.

### 2.1. Dimensions

Dimension quantifies the self-similarity of a geometrical object [7, 8]. For a homogeneous object, its dimension is a fixed number; while for a heterogeneous object, its different parts may have different dimensions and need to be characterized by the multifractal dimension. We here focus on the homogeneous objects. In this situation, a finite collection of points is zero dimensional, lines have dimension one, surfaces two and bulks three, and so forth. A chaotic attractor usually has a fractal dimension, that can be characterized by several ways. Three of them are convenient for numerical calculation, which are the box-counting dimension, information dimension, and correlation dimension.

For calculating the box-counting dimension $D_0$, we divide the attractor into spheres/boxes of radius $\epsilon$ and ask how many boxes do we need to cover all the points in the data set. If we evaluate this number $N(\epsilon)$ as a function of $\epsilon$ as it becomes small, then we define the box-counting dimension by

$$D_0 = \lim_{\epsilon \to 0} \frac{\ln N(\epsilon)}{\ln(1/\epsilon)}. \tag{2.1}$$

Take Henon map $x_{n+1} = y_n + 1 - 1.4x_n^2$, $y_{n+1} = 0.3x_n$ as an example. Its attractor is shown in Figure 1(a), where the square lattice has length $\epsilon$. We count the number of square lattice which contains at least one point of the trajectory and then obtain one point in Figure 1(b).

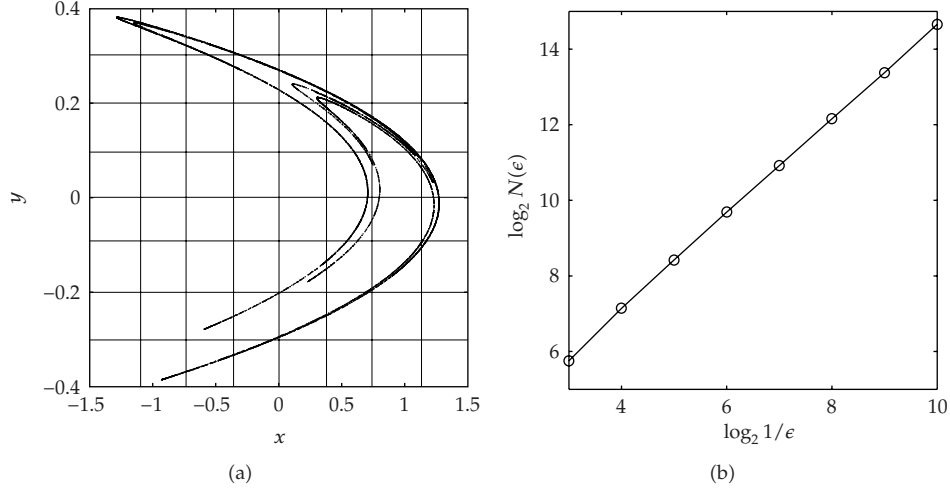(a)                                                                      (b)

**Figure 1:** (a) The attractor of Henon map; (b) the slope of the straight line represents the box-counting dimension $D_0$.

Then we change the $\epsilon$ in Figure 1(a) into $\epsilon/2$ and do the same process to obtain another point in Figure 1(b). In this way, we get Figure 1(b) in which the slope of the straight line is $D_0 \approx 1.27$.

The shortage of box-counting dimension is that it treats all the boxes with points of the trajectory as the same, no matter how many points in each box. Information dimension overcomes this shortage and is defined in terms of the relative frequency of visitation of a typical trajectory by

$$D_1 = \lim_{\epsilon \to 0} \frac{H(\epsilon)}{\ln(1/\epsilon)}, \tag{2.2}$$

where

$$H(\epsilon) = -\sum_{i=1}^{N(\epsilon)} P_i \ln P_i. \tag{2.3}$$

$P_i$ is the relative frequency with which a typical trajectory enters the $i$th box of the covering. Comparing with the definition of $D_0$, we see that the box-counting dimension is weight-free while the information dimension is weighted on the visiting frequency in the specific boxes. When $P_i$ is a constant, $D_1$ will become $D_0$.

When dimension is high, the calculation of both $D_0$ and $D_1$ will be time consuming and a convenient dimension for this case is the correlation dimension $D_2$. Comparing with the box-counting dimension $D_0$ and the information dimension $D_1$, the correlation dimension $D_2$ is much easier to be calculated and is defined by

$$D_2 = \lim_{\epsilon \to 0} \frac{\ln \sum_{i=1}^{N(\epsilon)} P_i^2}{\ln \epsilon}. \tag{2.4}$$

The sum in the numerator of (2.4) can be expressed as the following correlation sum [9]:

$$C(\epsilon) = \frac{2}{N(N-1)} \sum_{j=1}^{N} \sum_{i=j+1}^{N} \Theta\big(\epsilon - |\mathbf{x}_i - \mathbf{x}_j|\big), \tag{2.5}$$

where $\Theta(\cdot)$ is the Heaviside function, $\Theta(x) = 1$ for $x \geq 0$ and $0$ otherwise, and $|\mathbf{x}_i - \mathbf{x}_j|$ stands for the distance between points $\mathbf{x}_i$ and $\mathbf{x}_j$. Thus the sum just counts the pairs $(\mathbf{x}_i, \mathbf{x}_j)$ whose distance is smaller than $\epsilon$. Then (2.4) becomes

$$D_2 = \lim_{\epsilon \to 0} \frac{\ln C(\epsilon)}{\ln \epsilon}. \tag{2.6}$$

As (2.5) is very easy to be calculated, the correlation dimension $D_2$ has been widely used in time series analysis.

In practice, the sum in the equation of $C(\epsilon)$ also depends on the embedding dimension $m$ as $x(t)$ depends on the embedding space. The correlation dimension can be calculated in two steps. First one has to determine the correlation sum $C(\epsilon)$, for the range of $\epsilon$ available and for several embedding dimensions $m$. Then we inspect $C(m, \epsilon)$ for the signatures of self-similarity. If these signatures are convincing enough, we can compute a value for the dimension. Grassberger and Procaccia demonstrated that if the embedding dimension and the number of data points are large enough and if the data are sufficiently noise-free, then the function $\ln C(\epsilon)$ versus $\ln \epsilon$ has a linear region, called the scaling region [13, 14]. The slope of the function in that linear region is $D_2$. Due to such reasons, the correlation dimension $D_2$ usually is estimated by examining the slope of the linear portion of the plot of $\ln C(\epsilon)$ versus $\ln \epsilon$ for a series of increasing values of $m$.

### 2.2. Lyapunov Exponents

Lyapunov exponent is the most important quantity to chaotic systems as a positive maximal Lyapunov exponent is a strong signature of chaos. In the contrary, a zero maximal Lyapunov exponent denotes a limit cycle or a quasiperiodic orbit and a negative maximal Lyapunov exponent represents a fixed point. An $m$-dimensional system has $m$ Lyapunov exponents with $\lambda_1, \lambda_2, \ldots, \lambda_m$ in descending order.

Consider a dynamical system described by (1.2). Taking variation of both sides of (1.2) yields the following equation governing the evolution of the infinitesimal vector $\delta\mathbf{x}$ in the tangent space at $\mathbf{x}(t)$:

$$\frac{d\delta\mathbf{x}}{dt} = \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \cdot \delta\mathbf{x}. \tag{2.7}$$

Solving for (2.7) gives

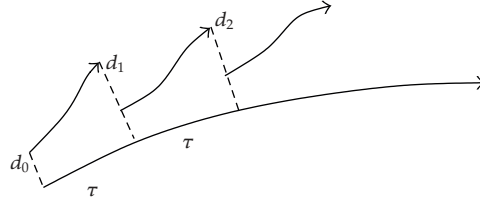$$\delta\mathbf{x}(t) = \mathbf{A}^t \delta\mathbf{x}(0), \tag{2.8}$$

**Figure 2:** Schematic illustration of how to numerically calculate the maximal Lyapunov exponent.

where $\mathbf{A}^t = \exp(\int(\partial\mathbf{F}/\partial\mathbf{x})dt)$ is a linear operator that evolves an infinitesimal vector at time 0 to time $t$. The mean exponential rate of divergence of the tangent vector is then given by

$$\lambda[\mathbf{x}(0), \delta\mathbf{x}(0)] = \lim_{t\to\infty} \frac{1}{t} \ln\left|\frac{\delta\mathbf{x}(t)}{\delta\mathbf{x}(0)}\right|. \tag{2.9}$$

The Lyapunov exponents are given as

$$\lambda_i \equiv \lambda[\mathbf{x}(0), \mathbf{e}_i], \tag{2.10}$$

where $e_i$ is an $m$-dimensional basis vector. Obviously, each Lyapunov exponent is an average of the local divergence rates over the whole attractor. For chaotic system, values of $\lambda_i$ do not depend on the choice of the initial condition $\mathbf{x}(0)$ because of the ergodicity.

To check whether a time series is chaotic or not, one needs to calculate the $\lambda_1$, that is, the maximal Lyapunov exponent $\lambda_{\max}$. This task is much easier than the calculation of all the $\lambda_i$. The reason is that a chaotic trajectory will automatically go to its maximum expending direction. This property can be conveniently used to calculate the $\lambda_{\max}$. Similarly, a chaotic trajectory will automatically go to its maximum contracting direction if we let it do backward evolution with $t \to -t$, which can be used to calculate the smallest Lyapunov exponent $\lambda_m$. Numerically, one can calculate the $\lambda_{\max}$ as follows. Choose two very close initial points and let their distance be $d_0 \ll 1$. After integrating the dynamical system for a small time interval $\tau$, their distance will become $d_i$. Considering that the attractor has a finite size, it is very easy for the trajectory to reach the boundary of attractor. Once it happens, the distance $d_i$ will not be exponentially growing. A convenient way to overcome this problem is to do renormalization, which makes the evolution begin at a small distance again. In detail, we choose a new point at the end of one trajectory and let their distance be $d_0$. Doing the integration again, we can get another $d_i$. In this way we have

$$\lambda_{\max} = \lim_{n\to\infty} \frac{1}{n\tau} \sum_{i=1}^{n} \ln\frac{d_i}{d_0}. \tag{2.11}$$

Figure 2 shows its schematic illustration.

The approach shown in Figure 2 can be also applied to the case of discrete systems. Consider an $m$-dimensional map $x(i)$ ($i = 1, 2, \ldots, m$). Let $\{x_n(i)\}$ be its state at time $n$. We make a new state $\{x'_n(i)\}$ by adding a small perturbation $\{dx(i)\}$. That is,
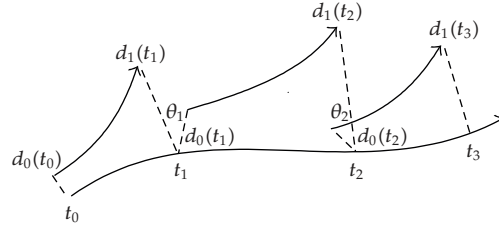
**Figure 3:** Schematic illustration of how to calculate the maximal Lyapunov exponent from time series or experimental data.

$\{x'_n(i) = x_n(i) + dx(i)\}$. Let both $\{x_n(i)\}$ and $\{x'_n(i)\}$ evolve for a time period to obtain their distance $\{dx_n(i)\}$. Then the maximal Lyapunov exponent $\lambda_{\max}$ can be given by

$$\lambda_{\max} = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \ln \frac{\|dx_n\|}{\|dx_{n-1}\|}, \tag{2.12}$$

where $\|dx_n\| = \sqrt{\sum_{i=1}^{m} [dx_n(i)]^2}$.

Practically, we cannot use the above approach to experimental data or time series as there may not always have a pair of points with distance $d_0$ at time $t = n\tau$. Thus we need to do some modification on it. The detailed process is as follows. Because of the recurrence of chaotic trajectory, one can always find two close points with distance $d_0(t_0)$ at time $t_0$. Then follow their trajectory to time $t_1$ and measure their distance $d_1(t_1)$. The evolution time is $t_1 - t_0$. After that, we do renormalization to find a new beginning point. Considering that the candidates for choosing are not sufficient, we choose one from them by the following two rules: (1) the distance $d_0(t_1)$ must be small; (2) to keep the direction of the maximal expending, the angle between $d_0(t_1)$ and $d_1(t_1)$ must be small. Figure 3 shows its schematic illustration. In this way we have

$$\lambda_{\max} = \lim_{N \to \infty} \frac{1}{t_N - t_0} \sum_{k=1}^{N} \ln \frac{d_1(t_k)}{d_0(t_{k-1})}. \tag{2.13}$$

Experimental data are generally contaminated by noise. Its influence can be minimized by using an averaging statistics when computing Lyapunov exponents. The concrete process can be taken as follows. Choose a point $\mathbf{x}(0)$ and select all its neighbors with distance smaller than $\epsilon$. The size of the neighborhood should be as small as possible, but large enough such that on average each reference point has at least a few neighbors. Compute the average over the distances of all neighbors to the reference part of the trajectory as a function of the relative time. Substitute this average into (2.11) or (2.13) to get the maximal Lyapunov exponent $\lambda_{\max}$. To calculate more Lyapunov exponents from data except the $\lambda_{\max}$, see [15, 16] for details.

### 2.3. Fourier Transform and Hilbert Transform

Fourier spectral analysis is traditionally the time series analysis method and very powerful in revealing the periodicity of time series. Its basic idea is that most signals, and all engineering

signals, can be represented as a sum of sine waves. By Fourier transform, a time series is transformed into a frequency spectrum, that is, from real space to frequency domain or Fourier space. The Fourier transform establishes a one-to-one correspondence between the signal at certain times (time domain) and how certain frequencies contribute to the signal. Thus, instead of describing the statistical properties of a signal in real space one can ask about its properties in Fourier space.

The Fourier transform of a function $s(t)$ is given by

$$\widetilde{s}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} s(t) e^{2\pi i \omega t} dt \tag{2.14}$$

and that of a finite, discrete time series by

$$\widetilde{s}_k = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} s_n e^{2\pi i k n / N}. \tag{2.15}$$

Here, the frequencies in physical units are $\omega_k = k/N\Delta t$, where $k = -N/2, \ldots, N/2$ and $\Delta t$ is the sampling interval.

A necessary condition for the Fourier analysis to be meaningful is that the time series should be piecewise stationary. However, a chaotic time series has a broad-band power spectra for which the Fourier spectrum gives no indication about the deterministic origin, let alone the fundamental invariants of the underlying dynamical system. For chaotic time series, a powerful approach is the Hilbert transform [10].

Consider a time series $x(t)$. One performs a mathematical transform to obtain the corresponding imaginary part $\widetilde{x}(t)$, yielding the following complex signal $\psi(t)$:

$$\psi(t) = x(t) + i\widetilde{x}(t) = A(t) e^{i\phi(t)}. \tag{2.16}$$

$\psi(t)$ is called *analytic signal* and the value of $\widetilde{x}(t)$ is obtained from $x(t)$ through a transform called *Hilbert transform*:

$$\widetilde{x}(t) = \text{P.V.} \left[ \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(t')}{t - t'} dt' \right], \tag{2.17}$$

where P.V. stands for the Cauchy principal value for the integral.

The Hilbert transform is closely related to the Fourier transform and their relationship can be seen clearly as follows. Observe that if the real signal $x(t)$ has a Fourier transform $S(\omega)$, then the complex signal, $\psi(t)$, the spectrum of which is composed of positive frequencies of $S(\omega)$ only, is given by the inverse transform of $S(\omega)$, where the integration goes only over the positive frequencies:

$$\psi(t) = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} S(\omega) e^{i\omega t} d\omega. \tag{2.18}$$

The factor of 2 is inserted so that the real part of the analytic signal is $x(t)$, not one half of that. The explicit form of $\psi(t)$ can then be obtained in terms of the real signal $x(t)$. Because

$$S(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt, \tag{2.19}$$

the complex signal can be written as

$$\psi(t) = \frac{1}{\pi} \int_{0}^{\infty} \int_{-\infty}^{\infty} x(t') e^{-i\omega(t-t')} dt' d\omega. \tag{2.20}$$

The following mathematical identity [17]:

$$\int_{0}^{\infty} e^{i\omega\tau} d\omega = \pi\delta(\tau) + \frac{i}{\tau} \tag{2.21}$$

gives

$$\psi(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} x(t') \left[ \pi\delta(t-t') + \frac{i}{t-t'} \right] dt', \tag{2.22}$$

which yields

$$\psi(t) = x(t) + i\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(t')}{t-t'} dt', \tag{2.23}$$

which is the analytic signal corresponding to the real signal $x(t)$. The imaginary part of (2.23) is just the Hilbert transform (2.17).

The analytic signal $\psi(t)$ corresponds geometrically to a rotation with amplitude and phase as follows:

$$A(t) = \sqrt{x(t)^2 + \tilde{x}(t)^2}, \qquad \phi(t) = \arctan\left(\frac{\tilde{x}(t)}{x(t)}\right). \tag{2.24}$$

By the phase variable $\phi(t)$ we can obtain an instantaneous frequency $\omega(t)$

$$\omega(t) = \frac{d\phi(t)}{dt} = \frac{x(t)\dot{\tilde{x}}(t) - \dot{x}(t)\tilde{x}(t)}{A^2(t)}, \tag{2.25}$$

where $\dot{\tilde{x}}(t)$ and $\dot{x}(t)$ denote the derivatives of $\tilde{x}(t)$ and $x(t)$ to $t$, respectively. Note that the instantaneous frequency $\omega(t)$ is fundamentally different from the concept of frequency in the Fourier transform defined in the base of simple harmonic functions. Here $\omega(t)$ measures the rate of rotation in the complex plane of analytic signal and is very useful in detecting the phase synchronization.

## 2.4. Attractor Reconstruction

Reconstruction of phase space is so important that we can do nothing without doing it. For example, our just introduced dimensions and Lyapunov exponents are very important concepts to the description of chaotic attractor. However, their calculations depend on the trajectory in attractor. For a measured time series, it is just a scalar measurement from one or a few variables of the attractor but not a trajectory. Thus, we need firstly to figure out the trajectory from the given time series. How to do that is a big challenge. Fortunately, the delay-coordinate embedding technique laid by Takens [12] gives the mathematical foundation to solve this problem. He showed that if the dynamical system and the observed quantity are generic, then the delay-coordinate map from a d-dimensional smooth compact manifold $M$ to $R^m$, $m > 2d$, is a diffeomorphism on $M$. That is, under fairly general conditions, the underlying dynamical system can be faithfully reconstructed from time series in the sense that a one-to-one correspondence can be established between the reconstructed and the true but unknown dynamical systems.

Takens' delay-coordinate method goes as follows. From a measured time series $s(k) = s(t_0 + k\Delta t)$ with $\Delta t$ being the sampling interval, the following vector quantity of $m$ components is constructed:

$$\mathbf{u}(t) = \{s(t), s(t + \tau), \dots, s(t + (m - 1)\tau)\}, \tag{2.26}$$

where $t = t_0 + k\Delta t$, $\tau$ is the delay time which is an integer multiple of $\Delta t$, and $m$ is the embedding dimension. Clearly, what time lag $\tau$ to use and what dimension $m$ to use are the central issues of this reconstruction based on delay-coordinate embedding. Let us discuss under what condition of $\tau$ and $m$, the reconstructed vector $\mathbf{u}(t)$ can represent the true trajectory of the unknown attractor.

### 2.4.1. Embedding Dimension

To have a faithful representation of the true dynamical system, the embedding dimension should be large enough. Takens' theorem provides a lower bound for $m$. Let us figure out this bound. We note that in a space of dimension $m$ one subspace of dimension $d_1$ and another subspace of dimension $d_2$ generically intersect in subspaces of dimension

$$d_I = d_1 + d_2 - m. \tag{2.27}$$

If this is negative, then there is no intersection of the two subspaces. Figure 4 shows examples. In Figure 4(a) we have $d_1 = d_2 = 1$ and $m = 2$, thus we obtain $d_I = 0$, which means that the intersection set consists of points, and the intersections are generic because small perturbations cannot remove them. In Figure 4(b) we have $d_1 = d_2 = 1$ and $m = 3$, thus we obtain $d_I = -1 < 0$, which means that two one-dimensional curves do not intersect generally in a three-dimensional space. In Figure 4(c) we have $d_1 = 1$, $d_2 = 2$, and $m = 3$, thus we obtain $d_I = 0$ again, which means that the intersections are generic. To obtain a one-to-one correspondence between points on the invariant sets in the actual and reconstructed phase spaces, self-intersection must not occur. Otherwise, there will be two directions at the self-intersection, which will destroy the one-to-one correspondence.
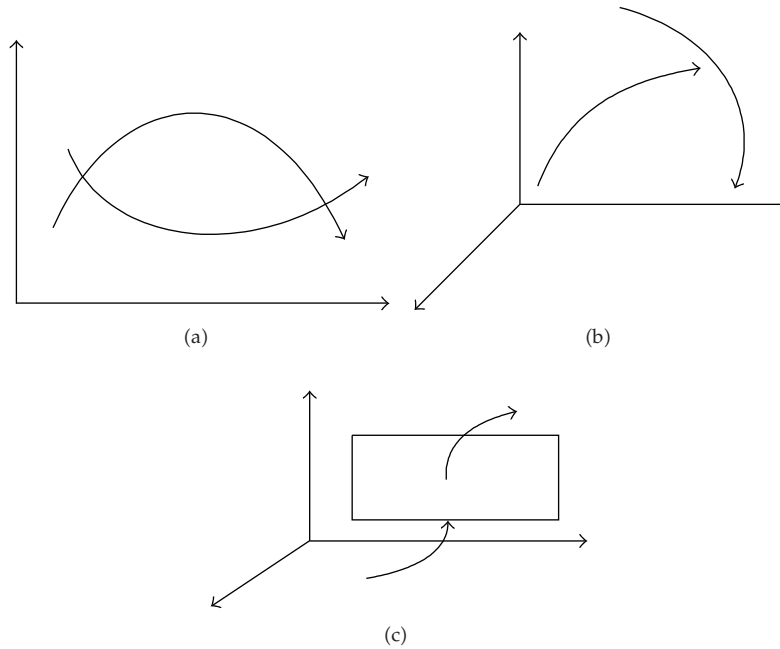
**Figure 4:** Schematic illustration of intersection of the two subspaces. (a) $d_1 = d_2 = 1$ and $m = 2$; (b) $d_1 = d_2 = 1$ and $m = 3$; (c) $d_1 = 1, d_2 = 2$ and $m = 3$.

When we ask about the intersection of two subspaces of the same dimension $d_A$, namely, the orbit with itself, then intersections fail to occur when $2d_A - m < 0$ or

$$m > 2d_A. \tag{2.28}$$

The question of unfolding a set of dimension $d_A$ from projections to lower dimensions concerns self-intersections of the set with itself as the projections are made. So the relevant criterion for unfolding a single object is $m > 2d_A$. We wish $m$ be an integer and let the lowest $m$ which satisfies the unfolding condition $m > 2d_A$ be the embedding dimension $m_E$. Thus, we have $m_E = \text{int}(2d_A + 1)$, where $\text{int}(\cdot)$ means taking integer. For example, the box counting dimension of the strange attractor for the Lorenz model is $d_A \approx 2.06$ which would lead us to anticipate $m_E = 5$ to unfold the Lorenz attractor.

In some situations, the needed embedding dimension is very large such as the EEG (electroencephalogram) data. As large embedding dimension requires long time series ($10^5 \sim 10^6$ points) and are thus computationally expensive, we hope to relax the condition $m > 2d_A$, especially for calculating some statistical quantities such as Lyapunov exponents and fractal dimension. To reduce the cost of calculation, we may choose $d_A < m < 2d_A$, provided that the self-intersections is neglectable. Schroer et al. show that, provided that the dimension $m$ of measurement space is large than the information dimension $D_1$ of the underlying dynamics, a prediction based on the reconstructed self-intersecting attractor is possible *most of the time* [18].

## 2.4.2. Time Delay

The embedding theorem is silent on the choice of time delay to use in constructing $m$-dimensional data vectors. Considering that the $s(t), s(t + \tau), \ldots, s(t + (m - 1)\tau)$ in (2.26) are serving as independent variables in the reconstruction space, they should be independent from each other and thus the delay time $\tau$ needs to be chosen carefully. In this sense, the chosen of $\tau$ should satisfy three conditions.

(1) It must be some multiple of the sampling time $\tau_s$, since we only have data at those times.

(2) If $\tau$ is too short, the coordinates $s(t)$ and $s(t + \tau)$ will not be independent enough and the reconstructed attractor will be accumulated around the diagonal line. That is, we will not have seen any of the dynamics unfold in that time.

(3) Finally, since chaotic systems are intrinsically unstable, if $\tau$ is too large, any connection between the measurements $s(t)$ and $s(t + \tau)$ is numerically tantamount to being random with respect to each other.

Therefore, the delay time $\tau$ should be large enough so that $s(t)$ and $s(t + \tau)$ are rather independent, but not so large that they are completely independent in a statistical sense. This is a difficult problem and can be solved by the mutual information whose first minimum is the good candidate of $\tau$.

## 2.4.3. Mutual Information $I(\tau)$

For a measured data $s(t)$, we create a histogram for the probability distribution of the data. Denote by $p_i$ the probability that the signal assumes a value inside the $i$th bin of the histogram, and let $p_{ij}(\tau)$ be the probability that $s(t)$ is in bin $i$ and $s(t + \tau)$ is in bin $j$. Then the mutual information for time delay $\tau$ reads

$$I(\tau) = \sum_{i,j} p_{ij}(\tau) \ln p_{ij}(\tau) - 2 \sum_i p_i \ln p_i. \tag{2.29}$$

In the special case of $\tau = 0$ the joint probabilities $p_{ij} = p_i \delta_{ij}$ and the expression yields the Shannon entropy of the data distribution. Apart from this degenerate case, the value of the mutual information is independent of the particular choice of histogram, as long as it is fine enough. In the limit of large $\tau$, $s(t)$ and $s(t + \tau)$ have nothing to do with each other and $p_{ij}$ thus factorizes to $p_i p_j$ and the mutual information becomes zero. The mutual information $I(\tau)$ between measurements $s(n)$ and time lagged measurements $s(n + T)$ is both easy to evaluate directly from the time series $s(n)$ and easy to interpret. A very nice property of average mutual information is that it is invariant under smooth changes of coordinate system. Fraser and Swinney suggest that one use the function $I(\tau)$ as a kind of nonlinear autocorrelation function to determine the optimal $\tau$ [19]. The first minimum of $I(\tau)$ marks the time lag value to use in time delay reconstruction of phase space where $s(t+\tau)$ adds maximal information to the knowledge we have from $s(t)$, or, in other words, the redundancy is least. Let us take the Lorenz system $\dot{x} = \sigma(y - x)$, $\dot{y} = \gamma x - y - xz$, $\dot{z} = -bz + xy$ as an example. Figure 5(a) shows how the mutual information $I(\tau)$ changes with $\tau$ when the parameters are taken as $\sigma = 16$, $\gamma = 45.92$, $b = 4$. It is easy to see that the first minimum of $I(\tau)$ occurs at $\tau = 10$. Figures 5(b) to 5(d) represent the reconstructed attractors by $\tau = 1, 10, 20$, respectively.
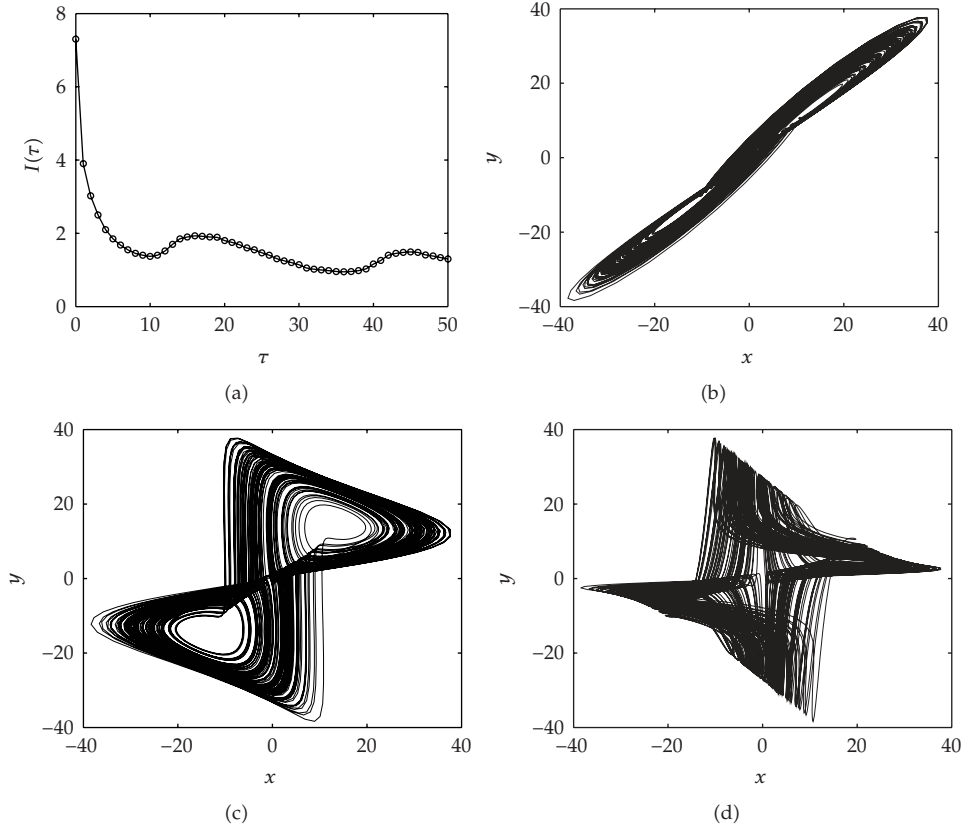
(a)

(b)

(c)

(d)

**Figure 5:** (a) The mutual information $I(\tau)$ versus the time delay $\tau$ for Lorenz system; (b) to (d) represent the cases where the attractors are reconstructed by $\tau = 1, 10, 20$, respectively.

Obviously, Figure 5(c) reconstructed by the first minimum of $\tau = 10$ reflects the structure of double-scroll best.

The time delay can be also chosen by other methods such as the autocorrelation function [7]. It is pointed out the delay spanned window $(m - 1)\tau$, rather than $\tau$ and $m$ separately, is a crucial issue for the attractor reconstruction. The reason is that the window $(m - 1)\tau$ determines the characteristics of the correlation integral whose linear part gives the correlation dimension $D_2$. However, the first minimum of mutual information is not consistently successful in identifying the optimal window [20]. The optimal window length for successful embedding $\tau_w$ should satisfy $\tau_w \geq \tau_p$ with $\tau_p$ being the mean orbital period of the underlying chaotic system and is approximated from the oscillations of the time series [21].

## 3. Modeling and Forecasting: Applications of Chaotic Time Series Analysis

The time series analysis has been widely applied in all the fields where the dynamical equations are unknown and only one or a few data are available. The most popular application is to calculate the dynamical parameters which partially reflect the properties

of the attractor. For example, from a time series one can first use the delay-coordinate embedding technique to reconstruct the phase space and then calculate its dimensions and Lyapunov exponents and so forth. by the methods introduced in the previous section. As this part is relatively easy and straightforward, we will not focus on it in this section. We would like to introduce a few typical cases where the time series analysis is necessary and useful tool to tell more information on the system, not only limited to the attractor.

### 3.1. Constructing Dynamical Equations

The most important question of time series analysis may be that: Is it possible to figure out the ordinary differential equations (ODEs) from a scalar time series, which govern the behavior of the system? This problem has been well studied and the answer is yes. A lot of methods have been developed to solve this problem. We here introduce two of them, that is, the standard approach and synchronization approach

### 3.1.1. The Standard Approach

This approach assumes every ODE can be written in a standard form [22, 23]

$$
\begin{aligned}
\dot{y}_1 &= y_2, \\
\dot{y}_2 &= y_3, \\
&\vdots \\
\dot{y}_D &= f(y_1, y_2, \ldots, y_D),
\end{aligned}
\tag{3.1}
$$

where $y_1$ is an observable, $f$ is a polynomial of an order $K$:

$$
f(y_1, y_2, \ldots, y_D) = \sum_{l_1, l_2, \ldots, l_D}^{K} c_{l_1, l_2, \ldots, l_D} \prod_{j=1}^{D} y_j^{l_j}, \quad \sum_{j=1}^{D} l_j \leq K.
\tag{3.2}
$$

For example, the Rössler system $\dot{x} = -y - z$, $\dot{y} = x + ay$, $\dot{z} = b + z(x - c)$ can be rewritten as $\dot{x} = y$, $\dot{y} = z$, $\dot{z} = ab - cx + x^2 - axy + xz + (ac - 1)y + (a - c)z - (y/(a + c - x))(x + b - ay + z)$. The Lorenz system $\dot{x} = \sigma(y - x)$, $\dot{y} = \gamma x - y - xz$, $\dot{z} = -bz + xy$ can be rewritten as $\dot{x} = y$, $\dot{y} = z$, $\dot{z} = b\sigma(\gamma - 1)x - b(\sigma + 1)y - (b + \sigma + 1)z - x^2 y - x^3 \sigma + (y/x)[(\sigma + 1)y + z]$. Then the task is to determine the coefficient from the data.

Reference [24] points that the above method is inefficient in many situations. In particular, universal models often contain a large number of coefficients and demonstrate divergent solutions. For solving this problem, [24] gives a modified approach as follows. Its main idea is to let the function $f$ depend explicitly on time and takes the form

$$
f(y_1, y_2, \ldots, y_D, t) = \sum_{l_1, l_2, \ldots, l_D}^{K} (c_{l_1, l_2, \ldots, l_D} + a_{l_1, l_2, \ldots, l_D} \cos \omega t + b_{l_1, l_2, \ldots, l_D} \sin \omega t) \prod_{j=1}^{D} y_j^{l_j}, \quad \sum_{j=1}^{D} l_j \leq K,
\tag{3.3}
$$

where the linear terms $\cos \omega t$ and $\sin \omega t$ is chosen for simple and effective approximation. In general, using higher powers of $\cos \omega t$ and $\sin \omega t$ are also possible. For (3.3), we still have one parameter to be determined, that is, the driving frequency $\omega$ or driving period $T$. The idea to find $T$ is as follows [25]. First, a sufficiently big value of a polynomial order $K$ is selected. Second, an initial estimate $T = T^*$ of the period value is found (this estimate can be derived as a location of a peak in the power spectrum of the observed series). At this value of $T$, one obtains the values of $c_{l_1,l_2,\ldots,l_D}$, $a_{l_1,l_2,\ldots,l_D}$, $b_{l_1,l_2,\ldots,l_D}$, and an approximation error $\epsilon$ by the linear least squares method. Then, the trial value of $T$ is varied through a certain range near the initial estimate $T = T^*$ and approximation is performed for each of the trial values. The graph of $\epsilon(T)$ has a sharp and deep minimum which corresponds quite precisely to the "true" value of the driving period.

### 3.1.2. Synchronization Approach

Recently, Sorrentino and Ott proposed an adaptive strategy that, by using synchronization, is able to obtain a set of ODE that describes the unknown real system [26]. They assume that the ODEs governing the system dynamics are expressible or approximately expressible in terms of polynomials of an assigned degree. Then they extract the whole set of parameters of the unknown system from knowledge about the dynamical evolution of its state vector and its first derivative. The detailed steps are as follows [26].

Consider a system $\dot{\mathbf{x}} = F(\mathbf{x})$ with $\mathbf{x} = (x_1, x_2, \ldots, x_m)^T$ and $F(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})]^T$, where $f_i(\mathbf{x})$ is a degree two polynomial

$$f_i(\mathbf{x}) = \sum_{j=1}^{m} \sum_{k=1}^{m} a_{ijk} x_j x_k + \sum_{j=1}^{m} b_{ij} x_j + c_i. \tag{3.4}$$

For example, in the case of Rössler system, $m = 3$ and $f_1 = -x_2 - x_3$, $f_2 = x_1 + 0.165x_2$, $f_3 = 0.2 + (x_1 - 10)x_3$, all the coefficients ($a_{ijk}, b_{ij}$, and $c_i$) are zero, except $a_{313} = 1, b_{12} = -1, b_{13} = -1, b_{21} = 1, b_{22} = 0.165, b_{33=-10}$, and $c_3 = 0.2$. Although the system function $F$ is unknown, it is appropriate to try to model the dynamics of the true system by $\dot{\mathbf{x}}' = F'(\mathbf{x}')$ with

$$f_i'(\mathbf{x}') = \sum_{j=1}^{m} \sum_{k=1}^{m} a_{ijk}' x_j' x_k' + \sum_{j=1}^{m} b_{ij}' x_j' + c_i'. \tag{3.5}$$

Then the task is how to make $a_{ijk}', b_{ij}', c_i'$ evolve to the true coefficient $a_{ijk}, b_{ij}, c_i$. To accomplish this goal, we envision coupling the true system to the model system (3.4). It is then hoped that when the synchrony is achieved, the model coefficients will be a good approximation to the corresponding coefficients of the real system.

The coupling from the true system to the model is designed as follows:

$$\dot{\mathbf{x}}' = F'(\mathbf{x}') + \Gamma(H(\mathbf{x}) - H(\mathbf{x}')). \tag{3.6}$$

The quantity $H$ is in general an $n \leq m$ vector of $n$ observable scalar quantities that are assumed to be known functions of the system state $\mathbf{x}(t)$. $\Gamma$ is an $n \times m$ constant coupling matrix. Here we assume $H(\mathbf{x}) = \mathbf{x}$ and $\Gamma = \gamma I_m$, where $\gamma$ is a scalar quantity and $I_m$ is the

identity matrix of dimension $m$. To obtain the synchronized solution $\mathbf{x}(t) = \mathbf{x}'(t)$, we introduce a potential

$$\Psi_i = \left\langle \left[ \dot{x}_i - f_i'(x_1', x_2', \ldots, x_m') \right]^2 \right\rangle_v, \quad i = 1, 2, \ldots, m, \tag{3.7}$$

where $\langle G(t) \rangle_v$ denotes the sliding exponential average $v \int^t e^{-v(t-t')} G(t') dt'$. Note that $\Psi_i$ is a function of time and also of the coefficients $a_{ijk}', b_{ij}',$ and $c_i'$. Also note that if $x_1(t), x_2(t), \ldots, x_m(t)$ are chaotic, then the quantities $f_i(x_1(t), x_2(t), \ldots, x_m(t)), i = 1, 2, \ldots, m,$ vary chaotically as well. It is easy to see that the potential satisfies $\Psi_i \geq 0$. Once $\Psi_i = 0$, we have synchronization and the coefficients are obtained.

To make $\Psi_i = 0$, we let the parameters $a_{ijk}', b_{ij}',$ and $c_i'$ adaptively evolve according to the following gradient descent relations:

$$\frac{da_{ijk}'(t)}{dt} = -\beta_a \frac{\partial \Psi_i}{\partial a_{ijk}'},$$

$$\frac{db_{ij}'(t)}{dt} = -\beta_b \frac{\partial \Psi_i}{\partial b_{ij}'}, \tag{3.8}$$

$$\frac{dc_i'(t)}{dt} = -\beta_c \frac{\partial \Psi_i}{\partial c_i'}$$

$\beta_a, \beta_b, \beta_c > 0$. Our hope is that $a_{ijk}', b_{ij}',$ and $c_i'$ will converge under this evolution to the true parameter values, $a_{ijk}, b_{ij}, c_i$.

We consider the first equation of (3.8). Let $(f_i')_{jk}$ denote $f_i'(x_1', x_2', \ldots, x_m')$ evaluated at $a_{ijk}' = 0$, we have $f_i'(x_1', x_2', \ldots, x_m') = a_{ijk}' x_j' x_k' + (f_i')_{jk}$. Substituting this into the right-hand side of the first equation of (3.8), we obtain

$$\frac{da_{ijk}'(t)}{dt} = -2\beta_a \left\langle a_{ijk}' x_j'^2 x_k'^2 + (f_i')_{jk} x_j' x_k' - \dot{x}_i x_j' x_k' \right\rangle_v. \tag{3.9}$$

Similarly letting $(f_i')_j$ denote $f_i'(x_1', x_2', \ldots, x_m')$ evaluated at $b_{ij}' = 0$, the second equation of (3.8) gives

$$\frac{db_{ij}'(t)}{dt} = -2\beta_b \left\langle b_{ij}' x_j'^2 + (f_i')_j x_j' - \dot{x}_i x_j' \right\rangle_v. \tag{3.10}$$

Finally, we consider the third equation of (3.8) with $(f_i')$ denote $f_i'(x_1', x_2', \ldots, x_m')$ evaluated at $c_i' = 0$. Then

$$\frac{dc_i'(t)}{dt} = -2\beta_c \left\langle c_i' + (f_i') - \dot{x}_i \right\rangle_v. \tag{3.11}$$

We here consider the case where $\beta_{a,b,c}$ are very large. For this situation the solutions $a_{ijk}', b_{ij}',$ and $c_i'$ rapidly converge to the minimum of the potentials, indicating we can set the

averages $\langle \cdots \rangle_v$ on the right-hand side of (3.9)–(3.11) to zero. We further consider that the average $\langle \cdots \rangle_v$ is performed over a time scale $v^{-1}$, which is much larger than the time scale $T_s$ for variation in $\mathbf{x}(t)$, in which case $a'_{ijk}, b'_{ij}$, and $c'_i$ vary slowly compared to $\mathbf{x}(t)$. Under these conditions, (3.8) and (3.9)–(3.11) then yield

$$\sum_{l=1}^{m}\sum_{n=l}^{m}a'_{iln}\left\langle x'_l x'_n x'_j x'_k\right\rangle_v + \sum_{l=1}^{m}b'_{il}\left\langle x'_l x'_j x'_k\right\rangle_v + c'_i\left\langle x'_j x'_k\right\rangle_v = \left\langle \dot{x}_i x'_j x'_k\right\rangle_v, \quad j=1,\ldots,m; \ k=j,\ldots,m,$$

$$\sum_{l=1}^{m}\sum_{n=l}^{m}a'_{iln}\left\langle x'_l x'_n x'_j\right\rangle_v + \sum_{l=1}^{n}b'_{il}\left\langle x'_l x'_j\right\rangle_v + c'_i\left\langle x'_j\right\rangle_v = \left\langle \dot{x}_i x'_j\right\rangle_v, \quad j=1,\ldots,m,$$

$$\sum_{l=1}^{m}\sum_{n=l}^{m}a'_{iln}\langle x'_l x'_n\rangle_v + \sum_{l=1}^{n}b'_{il}\langle x'_l\rangle_v + c'_i = \langle \dot{x}_i\rangle_v.$$

$$(3.12)$$

Equation (3.12) constitutes a system of $M = [(m^2/2) + (3m/2) + 1]m$ linear equations for the $M$ quantities $a'_{ijk}, b'_{ij}$, and $c'_i$. In practice, it is inconvenient to explicitly calculate the integrals for these quantities in terms of the form

$$I(t) = \langle G(t)\rangle_v = v\int^{t} e^{-v(t-t')}G(t')dt' \tag{3.13}$$

at every time step. Instead we will use the fact that $I(t)$ satisfies the differential equation

$$\frac{dI}{dt} + vI = vG(t) \tag{3.14}$$

and obtain $I(t)$ as a function of time by solving (3.14). Thus the adaptive system for finding estimates of the quantities $a_{ijk}, b_{ij}$, and $c_i$ is (3.6) for $\mathbf{x}'(t)$ and (3.12) for $a'_{ijk}, b'_{ij}$, and $c'_i$, where the various terms in (3.12) are of the form $I(t) = \langle G(t)\rangle_v$ obtained by integrating (3.14). Sorrentino and Ott have applied this method to both the Rössler and Lorenz systems and confirmed that it works well [26].

### 3.2. Detecting the Phase Locking

Another important application of time series analysis is in biology and medical sciences. It is found that in living systems, synchronization is often essential in normal functioning, while abnormal synchronization can lead to severe disorders. For example, the EEG data from human brain shows that synchronization under normal conditions seems to be essential for the binding problem; whereas epilepsies are related to abnormally strong synchronization [27–29]. In principal, the synchronization relationship between two time sequences may be phase synchronization (PS), generalized synchronization (GS), Lag synchronization (LS), and complete synchronization (CS), and so forth, depending on the coupling strength [30].

For two coupled identical systems, one may observe CS where there is an invariant synchronization manifold [31]. And for two coupled nonidentical systems, it may show PS and then LS when coupling is increasing [32]. At PS, their frequencies are locked whereas

their amplitudes remain chaotic and uncorrelated [33–35]. And at LS, the corresponding variables become identical after the transformation of time shift. For two coupled different systems, we may observe GS when coupling is large enough [36–45], where there is a functional relation between their states.

CS is very easy to be detected by directly checking the identity between the measured two time series. However, it is not so easy to detect the PS, LS and GS. To detect PS, we first need to calculate the phase $\phi_{1,2}$ from the two time series by the Hilbert transform. If they satisfy the condition

$$\left| n\phi_1 - \ell\phi_2 \right| < \text{const}, \tag{3.15}$$

where $n, \ell$ are integers, the two time series are phase-locking or in PS. To check LS, we calculate a similarity function $S$:

$$S^2(\tau) = \frac{\left\langle [x_2(t+\tau) - x_1(t)]^2 \right\rangle}{\left[ \langle x_1^2(t) \rangle \langle x_2^2(t) \rangle \right]^{1/2}}, \tag{3.16}$$

which is a time-averaged difference between the variables $x_1$ and $x_2$ taken with the time shift $\tau$. Then we search for its minimum $\sigma = \min_\tau S(\tau)$. If there exists a $\tau_0$ with $\sigma = 0$, we have LS, that is, $x_2(t+\tau) = x_1(t)$. Figure 6 shows the similarity function $S$ for two coupled nonidentical Rössler systems

$$\dot{x}_{1,2} = -\omega_{1,2}y_{1,2} - z_{1,2} + k(x_{2,1} - x_{1,2}),$$

$$\dot{y}_{1,2} = \omega_{1,2}x_{1,2} + ay_{1,2}, \tag{3.17}$$

$$\dot{z}_{1,2} = -b + z_{1,2}(x_{1,2} - c),$$

where $\omega_{1,2} = 0.97 \pm 0.02$, $a = 0.165$, $b = 0.2$, $c = 10$. From Figure 6 we see that the LS is possible for $k \geq 0.15$.

To check GS, there is a convenient method, that is, the auxiliary system approach [38]. The idea is as follows. Consider two coupled systems $\dot{\mathbf{x}} = F(\mathbf{x})$, $\dot{\mathbf{y}} = G(\mathbf{y}, \mathbf{x})$. Then we construct an auxiliary response system $\mathbf{y}'$ identical to $\mathbf{y}$, link it to the driving system $\mathbf{x}$ in the same way as $\mathbf{y}$ is linked to $\mathbf{x}$. That is, $\dot{\mathbf{y}}' = G(\mathbf{y}', \mathbf{x})$. Instead of checking the relationship between $\mathbf{x}$ and $\mathbf{y}$, we check the relationship between $\mathbf{y}$ and $\mathbf{y}'$. If there is an identical synchronization between $\mathbf{y}$ and $\mathbf{y}'$, then we have a GS between $\mathbf{x}$ and $\mathbf{y}$. Unfortunately, this method fails for the time series where the dynamical equations are unknown. An alternative approach to detect the GS is to estimate the maximal conditional Lyapunov exponent $\lambda_{\max}^R$. There is GS if $\lambda_{\max}^R < 0$. The conditional Lyapunov exponent is determined by the variational equation of the response system at $\delta\mathbf{x} = 0$

$$\delta\dot{\mathbf{y}} = D_y G(\mathbf{y}, \mathbf{x})\delta\mathbf{y}, \tag{3.18}$$

where $D_y G$ denotes the Jacobian matrix with respect to the $\mathbf{y}$ variable. For a dynamical system with no explicit driving and response parts, the GS can be determined by the transverse
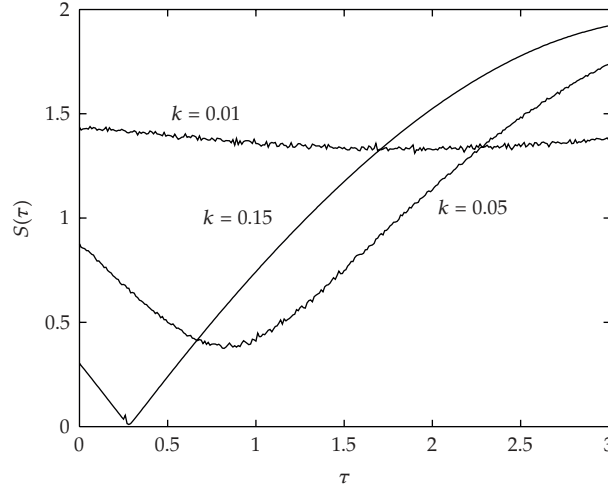
**Figure 6:** The dependence of the similarity function $S$ on the time lag for different coupling strengths.

Lyapunov exponent which is from the variational equation on the invariant manifold $\mathbf{x} = \mathbf{y}$. Its detail can be found in [41, 46].

For the case of nonstationary time series, we do not have a fixed relationship like the CS, PS, LS, and GS for all the time, that is, the synchronized relation may change with time such as in the EEG data with epileptic seizure. An approach based on the permutation entropy may be useful for this case [47, 48]. Let us briefly introduce the concept of entropy. In probability theory, entropy quantifies the uncertainty associated to a random process. Consider an experiment with outcome $S = \{s_1, s_2, \ldots, s_n\}$. Assume that the probability of $s_i$ is $p_i$ with $\sum_i p_i = 1$. If $s_1$ has a probability very close to 1, then in most experiments the outcome would be $s_1$ thus the result is not very uncertain. One does not gain much information from performing the experiment. One can quantify the "surprise" of the outcome as information $= -\ln$ (probability). The entropy associated to the experiment is

$$H = -\sum p_i \ln p_i \qquad (3.19)$$

which is simply the expectation value of the information produced by the experiment. Entropy quantifies the information content, namely, the amount of randomness of a signal.

For two measured scalar time series data $x_i(n\Delta t), y_i(n\Delta t)$, $n = 1, 2, \ldots$, we divide the long time series into segments with fixed finite length $\tau$. Then in each segment denoted as $\tau_j$ $(j = 1, 2, \ldots)$, we use the technique of sliding window analysis to partition the time series data into short sequences of a given length $m = 3$. Each shifting in the time series corresponds to a new short sequence. For example, $x_i(n\Delta t), x_i((n + 1)\Delta t), x_i((n + 2)\Delta t)$ and $x_i((n + 1)\Delta t), x_i((n + 2)\Delta t), x_i((n + 3)\Delta t)$ are different sequences. For a time series with length $N$, the total short sequences in one segment will be $N - m + 1$. Inside a short sequence, we distinguish it as a pattern by the natural position order. Hence we have four different patterns for $m = 3$, such as $x_i(n\Delta t) < x_i((n + 1)\Delta t) < x_i((n + 2)\Delta t)$, $x_i(n\Delta t) < x_i((n + 1)\Delta t) > x_i((n+2)\Delta t), x_i(n\Delta t) > x_i((n+1)\Delta t) < x_i((n+2)\Delta t)$, and $x_i(n\Delta t) > x_i((n+1)\Delta t) > x_i((n+2)\Delta t)$.

Using $p_i$ ($i = 1, \ldots, 4$) to denote the probability that one of the patterns appears in the time series of a segment, we can define the permutation entropy $H(\tau_j)$ as

$$H(\tau_j) = -\sum_{i=1}^{4} p_i \ln p_i. \tag{3.20}$$

Because the time series is always changing with time, $H(\tau_j)$ changes with time period interval $\tau_j$ and will be determined by the local topological structure of the time series. For two time series with GS, we cannot expect them to have the same permutation entropy $H(\tau_j)$ in the corresponding time period $\tau_j$ because they are not identity, but we would expect the corresponding $H(\tau_j)$ have similar changing tendency. That is, we require the changing tendency of $H(\tau_j)$ will be in step if there is GS between the two time series $x_i(n\Delta t)$ and $y_i(n\Delta t)$. By this way we can figure out the degree of GS [47, 48].

### 3.3. Inferring Coupling Direction

Sometimes, knowing the synchronized relationship between two time series is not enough, and we need to know more information about them such as which one is the driving system and which one is the response system. For example, in the prediction and control of epileptic seizure, we need to find out the focus. This problem has been well studied in the past decade and a number of approaches have been proposed, such as the cross-correlation functions, cross-spectral analysis, Granger causality, the nearest neighbors, and phase dynamical modeling [49–53]. We here introduce typical three of them: the Granger causality, the nearest neighbors, and the phase dynamical modeling.

#### 3.3.1. Granger Causality

Granger causality is based on the notion of predictability [49]. In general, in a coupled system that involves two interacting fields, Granger causality tests whether past values of one field ($X$) statistically help to predict the current values of the other field ($Y$) better than using past values of $Y$ alone. Should past values of $X$ contain information about current values of $Y$ beyond that contained in the preceding $Y$ sequence (or any other variables contained in the information set), variability in the $X$ field is said to "Granger causal" variability in the $Y$ field. Similarly, we can test whether previous values of $Y$ cause variability in the present values of $X$.

Consider two time series $\{x_n\}$ and $\{y_n\}$. If there is a causal relation between process $y$ and process $x$, the prediction of signal $\{x_n\}$ can be improved by incorporation into the model the past of signal $\{y_n\}$. As a result, we have a "joint"/bivariate autoregressive (AR) model

$$x_n = f(x_{n-1}, \ldots, x_{n-d_1}) + g(y_{n-1}, \ldots, y_{n-d_2}), \tag{3.21}$$

where $f$ and $g$ are polynomials that can be determined from the current data. $d_1$ is the correlation length to the previous values, and $d_2$ describes "inertial" properties of the influence. If $d_2 = 1$, then the influence is instantaneous, otherwise it is nonlocal in time.

Different values of $d_1$ and $d_2$ need to be tested in order to select those values that provide the most faithful results. In the framework of linear AR, we have

$$x_n = \alpha_0 + \sum_{i=1}^{d_1} \alpha_i x_{n-i} + \sum_{i=1}^{d_2} \beta_i y_{n-i}. \tag{3.22}$$

In (3.22) the coefficients $\alpha_i$ and $\beta_i$ are selected in order to minimize the mean square error

$$\epsilon_{xy}^2 = \frac{1}{N} \sum_n \left[ x_n - \left( \alpha_0 + \sum_{i=1}^{d_1} \alpha_i x_{n-i} + \sum_{i=1}^{d_2} \beta_i y_{n-i} \right) \right]^2. \tag{3.23}$$

The null hypothesis that $y_n$ does not Granger cause $x_n$ is supported when $\beta_i = 0$, which reduces (3.23) to

$$x_n = \alpha_0 + \sum_{i=1}^{d_1} \alpha_i x_{n-i}, \tag{3.24}$$

where the coefficients $\alpha_i$ are selected in order to minimize the mean square error

$$\epsilon_x^2 = \frac{1}{N} \sum_n \left[ x_n - \left( \alpha_0 + \sum_{i=1}^{d_1} \alpha_i x_{n-i} \right) \right]^2. \tag{3.25}$$

When $\epsilon_{xy}^2$ appears to be less than the $\epsilon_x^2$, it is assumed that process $y$ influences process $x$. This model leads to the well-known alternative test statistics, the Granger-Sargent test [54]

$$S_{xy}^2 = \frac{\epsilon_x^2 - \epsilon_{xy}^2}{\epsilon_{xy}^2}. \tag{3.26}$$

Thus, the influence of $\{y_n\}$ on $\{x_n\}$ is characterized by the value of the normalized prediction improvement $S_{xy}^2$ and the reverse influence of $\{x_n\}$ on $\{y_n\}$, $S_{yx}^2$ is described by an equation similar to (3.26) in which $x$ and $y$ should be interchanged.

### 3.3.2. The Nearest Neighbor Approach

This method is based on the existence of GS and is a nonlinear prediction approach [55–57]. Since many features of real data such as EEG signals cannot be generated by linear models, it is generally argued that nonlinear measures are likely to give more information than the one obtained with conventional linear approaches. The main idea is that for a driver/response system with $y = G(x)$, the response system $y$ will follow the driver system $x$. Therefore, the nearest neighbors of $x$ will have corresponding nearest neighbors of $y$ but the inverse does not work. Based on this feature, the direction of coupling can be figured out.

Consider two time series $\{x_n\}$ and $\{y_n\}$. Let us reconstruct delay vectors $\mathbf{x}_n = (x_n, \ldots, x_{n-(m-1)\tau})$ and $\mathbf{y}_n = (y_n, \ldots, y_{n-(m-1)\tau})$, where $n = 1, \ldots, N$, $m$ is the embedding
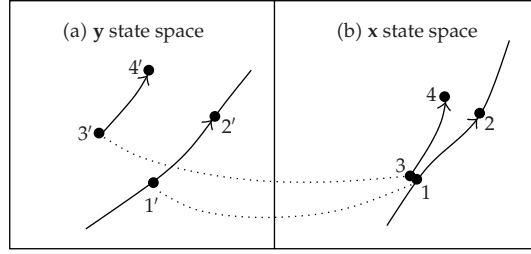
**Figure 7:** Schematic illustration for the indices of the nearest neighbors in **x** and **y** state space, respectively, where the number $1, 3$ denote the $r_{n,j}$ and $1', 3'$ the $s_{n,j}$.

dimension, and $\tau$ denotes the time lag. Let $r_{n,j}$ and $s_{n,j}$, $j = 1, \ldots, k$, denote the time indices of the $k$ nearest neighbors of $\mathbf{x}_n$ and $\mathbf{y}_n$, respectively. Figure 7 shows the schematic illustration with $k = 1$. For each $\mathbf{x}_n$, the square mean Euclidean distance to its $k$ neighbors is defined as

$$R_n^{(k)}(\mathbf{X}) = \frac{1}{k} \sum_{j=1}^{k} \left( \mathbf{x}_n - \mathbf{x}_{r_{n,j}} \right)^2, \tag{3.27}$$

and the **Y**-conditional squared mean Euclidean distance is defined by replacing the nearest neighbors by the equal time partners of the closest neighbors of $\mathbf{y}_n$,

$$R_n^{(k)}(\mathbf{X} \mid \mathbf{Y}) = \frac{1}{k} \sum_{j=1}^{k} \left( \mathbf{x}_n - \mathbf{x}_{s_{n,j}} \right)^2. \tag{3.28}$$

If the point cloud $\{\mathbf{x}_n\}$ has an average squared radius $R(\mathbf{X}) = (1/N) \sum_{n=1}^{N} R_n^{(N-1)}(\mathbf{X})$, then $R_n^{(k)}(\mathbf{X} \mid \mathbf{Y}) \approx R_n^{(k)}(\mathbf{X}) \ll R(\mathbf{X})$ if the system is strongly correlated, while $R_n^{(k)}(\mathbf{X} \mid \mathbf{Y}) \approx R(\mathbf{X}) \gg R_n^{(k)}(\mathbf{X})$ if they are independent. Accordingly, we can define an interdependent measure $S^{(k)}(\mathbf{X} \mid \mathbf{Y})$ as [28, 56, 57]

$$S^{(k)}(\mathbf{X} \mid \mathbf{Y}) = \frac{1}{N} \sum_{n=1}^{N} \frac{R_n^{(k)}(\mathbf{X})}{R_n^{(k)}(\mathbf{X} \mid \mathbf{Y})}. \tag{3.29}$$

Since $R_n^{(k)}(\mathbf{X} \mid \mathbf{Y}) \geq R_n^{(k)}(\mathbf{X})$ by construction, we have

$$0 < S^{(k)}(\mathbf{X} \mid \mathbf{Y}) \leq 1. \tag{3.30}$$

Low values of $S^{(k)}(\mathbf{X} \mid \mathbf{Y})$ indicate independence between **X** and **Y**, while high values indicate synchronization. This dependence becomes maximal when $S^{(k)}(\mathbf{X} \mid \mathbf{Y}) \to 1$.

The opposite dependence $S^{(k)}(\mathbf{X} \mid \mathbf{Y})$ is defined in complete analogy. It is in general not equal to $S^{(k)}(\mathbf{Y} \mid \mathbf{X})$. Both $S^{(k)}(\mathbf{X} \mid \mathbf{Y})$ and $S^{(k)}(\mathbf{Y} \mid \mathbf{X})$ may be of order 1. Thus, **X** can depend on **Y**, and at the same time **Y** can depend on **X**. If $S^{(k)}(\mathbf{X} \mid \mathbf{Y}) > S^{(k)}(\mathbf{Y} \mid \mathbf{X})$, that is, if **X** depends more on **Y** than vice versa, we say that **Y** is more "active" than **X**.

This approach is good at the case of only two coupled systems. For the case of multiple coupled systems, we need to pay attention to not only the values of $S^{(k)}(\mathbf{X} \mid \mathbf{Y})$ or $S^{(k)}(\mathbf{Y} \mid \mathbf{X})$ but also their difference. Take a ring of three unidirectionally coupled oscillators as an example. We will obtain that both the values of $S^{(k)}(\mathbf{X} \mid \mathbf{Y})$ and $S^{(k)}(\mathbf{Y} \mid \mathbf{X})$ between the oscillators 1 and 2 are greater than zero, but it does not means that their coupling is bidirectional. That is, the indirect coupling should be also considered.

### 3.3.3. Phase Dynamical Modeling

This approach is presented by Rosenblum and Pikovsky [58] and widely applied in real situations [59, 60]. The main idea is to use a general property that a weak coupling affects the phases of interacting oscillators, whereas the amplitudes remain practically unchanged. In detail, consider a case of weak coupling where the dynamics can be reduced to those of two phases $\phi_{1,2}$:

$$
\begin{aligned}
\dot{\phi}_1 &= \omega_1 + q_1(\phi_1) + \epsilon_1 f_1(\phi_1, \phi_2) + \xi_1(t), \\
\dot{\phi}_2 &= \omega_2 + q_2(\phi_2) + \epsilon_2 f_2(\phi_2, \phi_1) + \xi_2(t),
\end{aligned}
\tag{3.31}
$$

where the small parameters $\epsilon_{1,2} \ll \omega_{1,2}$ characterize the strength of coupling and the random terms $\xi_{1,2}$ describe noisy perturbations that are always present in real-world systems. Functions $q_{1,2}, f_{1,2}$ are $2\pi$-periodic in all arguments. If the coupling is bidirectional, both $\epsilon_1$ and $\epsilon_2$ will be greater than zero. In the case of an unidirectional driving, say from system 1 to system 2, we have $\epsilon_1 = 0$ and $\epsilon_2 > 0$. System (3.31) describes the phase dynamics of weakly coupled noisy limit cycle oscillators, Josephson junctions, and phase locked loops, as well as phase dynamics of weakly coupled continuous-time chaotic systems [58].

The coupling direction is represented by the ratio between $\epsilon_1$ and $\epsilon_2$, thus the goal is to estimate the ratio from the measured time series. To do it, we first extract the phase $\phi_{1,2}(t_k)$ from data by Hilbert transform approach, where $t_k = k\delta t$, $\delta t$ is the sampling interval, $k = 1, \ldots, N$. Then we compute for each time point the increments $\Delta_{1,2}(k) = \phi_{1,2}(t_k + \tau) - \phi_{1,2}(t_k)$, $\tau$ is a time delay. These increments can be considered as generated by some unknown two-dimensional noisy map $\Delta_{1,2}(k) = F_{1,2}[\phi_{1,2}(k), \phi_{2,1}(k)] + \eta_{1,2}(k)$. Next, we fit (in the least mean square sense) the dependencies of $\Delta$ on $\phi_1$ and $\phi_2$ using a finite Fourier series as the probe function:

$$
F_{1,2} = \sum_{m,l} A_{m,l} e^{im\phi_1 + il\phi_2}.
\tag{3.32}
$$

The sum in (3.32) is taken to the terms with $|l| \le 3$ for $m = 0$, $|m| \le 3$ for $l = 0$, and $|m| = |l| = 1$. The results of fitting are used to quantify the cross-dependencies of phase dynamics of two systems by means of the coefficients $c_{1,2}$ defined as

$$
c_{1,2}^2 = \int \int_0^{2\pi} \left( \frac{\partial F_{1,2}}{\partial \phi_{2,1}} \right)^2 d\phi_1 d\phi_2.
\tag{3.33}
$$

Finally, the *directionality index* is introduced as

$$r = \frac{c_2 - c_1}{c_2 + c_1}.$$

(3.34)

In the case of unidirectional coupling $r = \pm 1$, while for nearly symmetrical coupling $r$ is close to zero.

## 3.4. Other Applications

### 3.4.1. Coherence Resonance Induced by Noise

When two chaotic systems are coupled bidirectionally, there will be a synchronization when the coupling strength is strong enough. However, when there is noise in the system, the synchronization manifold will be destroyed from time to time, resulting in some kind of bursting. It is revealed that the time intervals of bursting will show some regularity, which can be enhanced by an optimal external noise and results in a phenomenon called coherence resonance [61, 62]. Coherence resonance is referred to the fact that noise can actually be utilized to improve the temporal regularity of the bursting time series, in the absence of an external periodic signal.

Consider two coupled Lorenz systems

$$\dot{x}_{1,2} = 10(y_{1,2} - x_{1,2}) + k(x_{2,1} - x_{1,2}) + D\xi_x(t),$$

$$\dot{y}_{1,2} = 28x_{1,2} - y_{1,2} - x_{1,2}z_{1,2} + D\xi_y(t),$$

$$\dot{z}_{1,2} = -\frac{8}{3}z_{1,2} + x_{1,2}y_{1,2} + D\xi_z(t),$$

(3.35)

where $k$ is the coupling strength, $\xi_{x,y,z}(t)$ are independent Gaussian noise, and $D$ quantifies the noise strength. When $D = 0$, there is a synchronized manifold for $k > k_c \approx 3.92$ [61]. But for $D > 0$, the difference between the two systems exhibits on-off intermittence. Take two time series from the variable $y$ of the two coupled systems and let $|\Delta y|$ denote their difference. Figure 8 shows the time series of $|\Delta y|$ for three typical $D$. It is easy to see that the bursting interval, denoted by $T$ in Figure 8(b), shows different behaviors in the three panels of Figure 8.

To show the regularity of time series $|\Delta y|$, Figure 9 shows the corresponding power spectra. From this figure we see that there are no peak in the spectra for both the small and large $D$, indicating a lack of temporal regularity in the bursting time series. A pronounced peak does exist at the intermediate noise level (see Figure 9(b)), indicating the existence of a strong time-periodic component in the time series. This apparent temporal regularity can be quantified by the characteristics of the peak at a nonzero frequency $\omega_p$ in the spectrum. In particular, we utilize the quantity $\beta_s = H\omega_p/\Delta\omega$, where $H$ is the height of the spectral peak, $\Delta\omega$ is the half width of the peak [63]. By its definition, a high value of $\beta_s$ indicates a strong temporal regularity in the bursting time series. This phenomenon can be also described by another quantity $\beta_T = \langle T \rangle / \sqrt{\text{Var}(T)}$, where $\langle T \rangle$ and $\text{Var}(T)$ are the average and variance of the interval $T$ [64]. Both quantities $\beta_s$ and $\beta_T$ show a bell shape on the noise strength $D$, that is, a resonance on $D$, which has been confirmed by experiment [62].
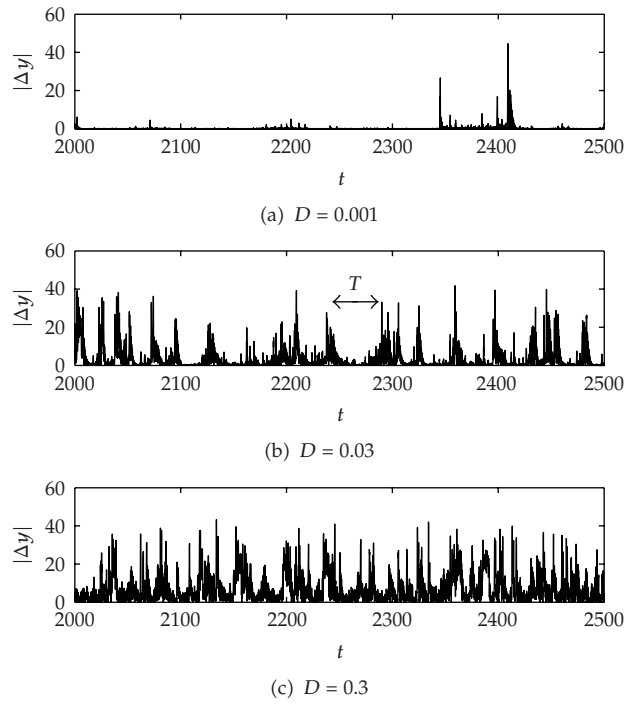
(a) $D = 0.001$



(b) $D = 0.03$



(c) $D = 0.3$

**Figure 8:** Intermittency of time series $|\Delta y|$ from synchronization manifold.



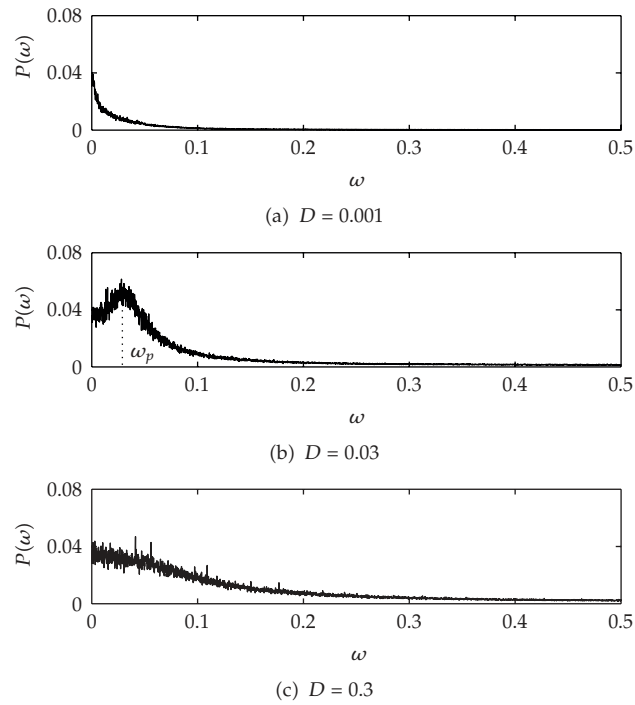(a) $D = 0.001$



(b) $D = 0.03$



(c) $D = 0.3$

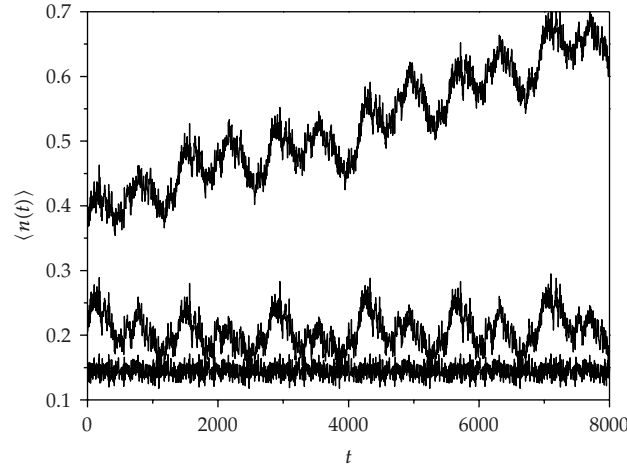**Figure 9:** Power spectra of time series $|\Delta y|$ corresponding to Figure 8.

**Figure 10:** Average packets per node for different status in an Internet model.

### 3.4.2. Correlation of Traffic Data in Internet

Undoubtedly, the internet has become a very important tool in our daily life [65–68]. The operations on the internet, such as browsing World Wide Web (WWW) pages, sending messages by email, transferring files by ftp, searching for information on a range of topics, and shopping, have benefited us a lot. Therefore, sustaining its normal and efficient functioning is a basic requirement. However, the communication in the internet does not always march/go freely. Similar to the traffic jam on the highway, the intermittent congestion in the internet has been observed [69]. Once it happens, the accumulated packets in Internet will increase linearly with time. Figure 10 shows three status from an Internet model [70], where the top, middle and bottom curves denote the phases of congestion, busy/buffer and free, respectively. From Figure 10 it is easy to see that there exist erratic fluctuation, heterogeneity, and nonstationarity in the data. These features make the correlation difficult to be quantified.

A conventional approach to measure the correlation in this situation is by the detrended fluctuation analysis (DFA), which can reliably quantify scaling features in the fluctuations by filtering out polynomial trends. The DFA method is based on the idea that a correlated time series can be mapped to a self-similar process by integration [71–74]. Therefore, measuring the self-similar feature can indirectly tell us information about the correlation properties. The DFA method has been successfully applied to detect long-range correlations in highly complex heart beat time series [71], stock index [72], physiological signals [73], and particle condensation [74].

The DFA method is a modified root-mean-square (rms) analysis of a random walk and its algorithm can be worked out as the following steps.

(1) Start with a time series $s(j)$, where $j = 1, \ldots, N$, and $N$ is the length of the data, and integrate $s(j)$ to obtain

$$y(i) = \sum_{j=1}^{i} \left[ s(j) - \langle s \rangle \right], \tag{3.36}$$

where $\langle s \rangle = (1/N) \sum_{j=1}^{N} s(j)$.

(2) Divide the integrated profile $y(i)$ into boxes of equal length $m$. In each box, we fit $y(i)$ to get its local trend $y_{fit}$ by using a least-square fit.

(3) The integrated profile $y(i)$ is detrended by subtracting the local trend $y_{fit}$ in each box:

$$Y_m(i) \equiv y(i) - y_{fit}(i).$$

(3.37)

(4) For a given box size $m$, the rms fluctuation for the integrated and detrended signal is calculated

$$F(m) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [Y_m(i)]^2}.$$

(3.38)

(5) Repeat this procedure for different box size $m$.

For scale-invariant time series with power-law correlations, there is a power-law relationship between the rms fluctuation function $F(m)$ and the box size $m$:

$$F(m) \sim m^\alpha.$$

(3.39)

The scaling $\alpha$ represents the degree of the correlation in the signal; the signal is uncorrelated for $\alpha = 0.5$ and correlated for $\alpha > 0.5$. Using the DFA method to the data in Figure 10, we find that the values of $\alpha$ for the three curves from top to bottom are $1.3, 0.7$, and $0.5$, respectively, indicating that the data of congestion phase are correlated while data of free phase are uncorrelated.

Except the above applications, there are many other fields of time series analysis such as chaos controlling, testing for nonlinearity with surrogate data, EEG data analysis and multifractals, and stock market analysis, which have gotten widely interesting from both physics and engineers [7–9].

## 4. Conclusions

The chaos theory and its time series analysis have been well studied in the past decades. A lot of important results have been achieved. This paper is contributed to the brief summary of chaotic time series analysis. The concepts, such as dimension, Lyapunov exponents, Hilbert transform, and attractor reconstruction, have been discussed. Several applications of time series analysis have been explained in detail, such as constructing dynamical equations, detecting the phase locking, inferring coupling direction, and coherence resonance induced by noise and correlation of traffic data in Internet. These are only a few of the applications of time series analysis, and a lot of other applications are not included here, such as the analysis of transient chaotic time series. Even in these mentioned fields, new techniques and methods are continuously showing up.

## Acknowledgments

## References

[1] E. Ott, *Chaos in Dynamical Systems*, Cambridge University, Cambridge, UK, 1993.

[2] K. T. Alligood, T. D. Sauer, and J. A. Yorke, *Chaos: An Introduction to Dynamical Systems*, Textbooks in Mathematical Sciences, Springer, New York, NY, USA, 1997.

[3] J. Guckenheimer and P. Holmes, *Nonlinear Oscillators, Dynamical Systems, and Bifurcations of Vector Fields*, vol. 42 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 1983.

[4] A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics*, vol. 38 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 1983.

[5] L. E. Reichl, *The Transition to Chaos in Conservative Classical Systems: Quantum Manifestations*, Institute for Nonlinear Science, Springer, New York, NY, USA, 1992.

[6] S. N. Elaydi, *Discrete Chaos*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2000.

[7] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University, Cambridge, UK, 2000.

[8] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer, New York, NY, USA, 1996.

[9] T. S. Parker and L. O. Chua, *Practical Numerical Algorithms for Chaotic Systems*, Springer, New York, NY, USA, 1689.

[10] Y.-C. Lai, Z. Liu, N. Ye, and T. Yalcinkaya, "Nonlinear time series analysis," in *The Handbook of Data Mining*, N. Ye, Ed., Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2003.

[11] L. A. Aguirre and C. Letellier, "Modeling nonlinear dynamics and chaos: a review," *Mathematical Problems in Engineering*, vol. 2009, Article ID 238960, 35 pages, 2009.

[12] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence , Warwick 1980 (Coventry, 1979/1980)*, D. A. Rand and L.-S. Young, Eds., vol. 898 of *Lecture Notes in Mathematics*, pp. 366–381, Springer, London, UK, 1981.

[13] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D*, vol. 9, no. 1-2, pp. 189–208, 1983.

[14] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Physical Review Letters*, vol. 50, no. 5, pp. 346–349, 1983.

[15] J.-P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Reviews of Modern Physics*, vol. 57, no. 3, pp. 617–656, 1985.

[16] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, "Liapunov exponents from time series," *Physical Review A*, vol. 34, no. 6, pp. 4971–4979, 1986.

[17] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, Academic Press, San Diego, Calif, USA, 4th edition, 1995.

[18] C. G. Schroer, T. Sauer, E. Ott, and J. A. Yorke, "Predicting chaos most of the time from embeddings with self-intersections," *Physical Review Letters*, vol. 80, no. 7, pp. 1410–1413, 1998.

[19] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.

[20] J. M. Martinerie, A. M. Albano, A. I. Mees, and P. E. Rapp, "Mutual information, strange attractors, and the optimal estimation of dimension," *Physical Review A*, vol. 45, no. 10, pp. 7058–7064, 1992.

[21] D. Kugiumtzis, "State space reconstruction parameters in the analysis of chaotic time series—the role of the time window length," *Physica D*, vol. 95, no. 1, pp. 13–28, 1996.

[22] G. Gouesbet and J. Maquet, "Construction of phenomenological models from numerical scalar time series," *Physica D*, vol. 58, no. 2, pp. 202–215, 1992.

[23] G. Gouesbet and C. Letellier, "Global vector-field reconstruction by using a multivariate polynomial $L_2$ approximation on nets," *Physical Review E*, vol. 49, no. 6, pp. 4955–4972, 1994.

[24] B. P. Bezruchko and D. A. Smirnov, "Constructing nonautonomous differential equations from experimental time series," *Physical Review E*, vol. 63, no. 1, Article ID 016207, 7 pages, 2001.

[25] R. Hegger, H. Kantz, F. Schmuser, M. Diestelhorst, R.-P. Kapsch, and H. Beige, "Dynamical properties of a ferroelectric capacitor observed through nonlinear time series analysis," *Chaos*, vol. 8, no. 3, pp. 727–736, 1998.

[26] F. Sorrentino and E. Ott, "Using synchronization of chaos to identify the dynamics of unknown systems," *Chaos*, vol. 19, no. 3, Article ID 033108, 2009.

[27] S. F. Farmer, "Rhythmicity, synchronization and binding in human and primate motor systems," *Journal of Physiology*, vol. 509, no. 1, pp. 3–14, 1998.

[28] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, "A robust method for detecting interdependences: application to intracranially recorded EEG," *Physica D*, vol. 134, no. 4, pp. 419–430, 1999.

[29] F. Mormann, K. Lehnertz, P. David, and C. E. Elger, "Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients," *Physica D*, vol. 144, no. 3, pp. 358–369, 2000.

[30] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Science*, vol. 12 of *Cambridge Nonlinear Science Series*, Cambridge University, Cambridge, UK, 2001.

[31] L. M. Pecora and T. L. Carroll, "Synchronization in chaotic systems," *Physical Review Letters*, vol. 64, no. 8, pp. 821–824, 1990.

[32] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "From phase to lag synchronization in coupled chaotic oscillators," *Physical Review Letters*, vol. 78, no. 22, pp. 4193–4196, 1997.

[33] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "Phase synchronization of chaotic oscillators," *Physical Review Letters*, vol. 76, no. 11, pp. 1804–1807, 1996.

[34] A. S. Pikovsky, M. G. Rosenblum, G. V. Osipov, and J. Kurths, "Phase synchronization of chaotic oscillators by external driving," *Physica D*, vol. 104, no. 3-4, pp. 219–238, 1997.

[35] W. Wang, Z. Liu, and B. Hu, "Phase order in chaotic maps and in coupled map lattices," *Physical Review Letters*, vol. 84, no. 12, pp. 2610–2613, 2000.

[36] N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, and H. D. I. Abarbanel, "Generalized synchronization of chaos in directionally coupled chaotic systems," *Physical Review E*, vol. 51, no. 2, pp. 980–994, 1995.

[37] L. M. Pecora, T. L. Carroll, and J. F. Heagy, "Statistics for mathematical properties of maps between time series embeddings," *Physical Review E*, vol. 52, no. 4, pp. 3420–3439, 1995.

[38] H. D. I. Abarbanel, N. F. Rulkov, and M. M. Sushchik, "Generalized synchronization of chaos: the auxiliary system approach," *Physical Review E*, vol. 53, no. 5, pp. 4528–4535, 1996.

[39] L. Kocarev and U. Parlitz, "Generalized synchronization, predictability, and equivalence of unidirectionally coupled dynamical systems," *Physical Review Letters*, vol. 76, no. 11, pp. 1816–1819, 1996.

[40] B. R. Hunt, E. Ott, and J. A. Yorke, "Differentiable generalized synchronization of chaos," *Physical Review E*, vol. 55, no. 4, pp. 4029–4034, 1997.

[41] K. Pyragas, "Conditional lyapunov exponents from time series," *Physical Review E*, vol. 56, no. 5, pp. 5183–5188, 1997.

[42] Z. Liu and S. Chen, "General method of synchronization," *Physical Review E*, vol. 55, no. 6, pp. 6651–6655, 1997.

[43] R. Brown, "Approximating the mapping between systems exhibiting generalized synchronization," *Physical Review Letters*, vol. 81, no. 22, pp. 4835–4838, 1998.

[44] C. L. Goodridge, L. M. Pecora, T. L. Carroll, and F. J. Rachford, "Detecting functional relationships between simultaneous time series," *Physical Review E*, vol. 64, no. 2, Article ID 026221, 2001.

[45] S. Boccaletti, L. M. Pecora, and A. Pelaez, "Unifying framework for synchronization of coupled dynamical systems," *Physical Review E*, vol. 63, no. 6, Article ID 066219, 2001.

[46] K. Pyragas, "Weak and strong synchronization of chaos," *Physical Review E*, vol. 54, no. 5, pp. R4508–R4511, 1996.

[47] Z. Liu, "Measuring the degree of synchronization from time series data," *Europhysics Letters*, vol. 68, no. 1, pp. 19–25, 2004.

[48] Z. Liu, H. U. Bambi, and L. D. Iasemidis, "Detection of phase locking from non-stationary time series," *Europhysics Letters*, vol. 71, no. 2, pp. 200–206, 2005.

[49] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, p. 424, 1969.

[50] W. Wang, B. T. Anderson, R. K. Kaufmann, and R. B. Myneni, "The relation between the north Atlantic oscillation and SSTs in the north Atlantic basin," *Journal of Climate*, vol. 17, no. 24, pp. 4752–4759, 2004.

[51] B. Schelter, M. Winterhalder, M. Eichler, et al., "Testing for directed influences among neural signals using partial directed coherence," *Journal of Neuroscience Methods*, vol. 152, no. 1-2, pp. 210–219, 2006.

[52] E. Sitnikova, T. Dikanev, D. Smirnov, B. Bezruchko, and G. Luijtelaar, "Granger causality: cortico-thalamic interdependencies during absence seizures in WAG/Rij rats," *Journal of Neuroscience Methods*, vol. 170, no. 2, pp. 245–254, 2008.

[53] D. A. Smirnov and I. I. Mokhov, "From Granger causality yo long-term causality: application to climatic data," *Physical Review E*, vol. 80, no. 1, Article ID 016208, 2009.

[54] K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Physics Reports*, vol. 441, no. 1, pp. 1–46, 2007.

[55] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, "Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble," *Physical Review E*, vol. 54, no. 6, pp. 6708–6724, 1996.

[56] R. Q. Quiroga, J. Arnhold, and P. Grassberger, "Learning driver-response relationships from synchronization patterns," *Physical Review E*, vol. 61, no. 5, pp. 5142–5148, 2000.

[57] R. Q. Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, "Performance of different synchronization measures in real data: a case study on electroencephalographic signals," *Physical Review E*, vol. 65, no. 4, Article ID 041903, 2002.

[58] M. G. Rosenblum and A. S. Pikovsky, "Detecting direction of coupling in interacting oscillators," *Physical Review E*, vol. 64, no. 4, Article ID 045202, 2001.

[59] B. Bezruchko, V. Ponomarenko, M. G. Rosenblum, and A. S. Pikovsky, "Characterizing direction of coupling from experimental observations," *Chaos*, vol. 13, no. 1, pp. 179–184, 2003.

[60] D. A. Smirnov, M. B. Bodrov, J. L. Perez Velazquez, R. A. Wennberg, and B. P. Bezruchko, "Estimation of coupling between oscillators from short time series via phase dynamics modeling: limitations and application to EEG data," *Chaos*, vol. 15, no. 2, Article ID 024102, 10 pages, 2005.

[61] Z. Liu and Y.-C. Lai, "Coherence resonance in coupled chaotic oscillators," *Physical Review Letters*, vol. 86, no. 21, pp. 4737–4740, 2001.

[62] L. Zhu, Y.-C. Lai, Z. Liu, and A. Raghu, "Can noise make nonbursting chaotic systems more regular?" *Physical Review E*, vol. 66, no. 1, Article ID 015204, 2002.

[63] H. Gang, T. Ditzinger, C. Z. Ning, and H. Haken, "Stochastic resonance without external periodic force," *Physical Review Letters*, vol. 71, no. 6, pp. 807–810, 1993.

[64] A. S. Pikovsky and J. Kurths, "Coherence resonance in a noise-driven excitable system," *Physical Review Letters*, vol. 78, no. 5, pp. 775–778, 1997.

[65] M. Li and W. Zhao, "Representation of a stochastic traffic bound," to appear in *IEEE Transactions on Parallel and Distributed Systems*.

[66] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.

[67] M. Li and P. Borgnat, "Forward for the special issue on traffic modeling, its computations and applications," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 181–195, 2009.

[68] M. Li and S. C. Lim, "Modeling network traffic using generalized Cauchy process," *Physica A*, vol. 387, no. 11, pp. 2584–2594, 2008.

[69] B. A. Huberman and R. M. Lukose, "Social dilemmas and Internet congestion," *Science*, vol. 277, no. 5325, pp. 535–537, 1997.

[70] X.-Y. Zhu, Z.-H. Liu, and M. Tang, "Detrended fluctuation analysis of traffic data," *Chinese Physics Letters*, vol. 24, no. 7, pp. 2142–2145, 2007.

[71] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series," *Chaos*, vol. 5, no. 1, pp. 82–87, 1995.

[72] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, and H. E. Stanley, "Statistical properties of the volatility of price fluctuations," *Physical Review E*, vol. 60, no. 2, pp. 1390–1400, 1999.

[73] Z. Chen, K. Hu, P. Carpena, P. Bernaola-Galvan, H. E. Stanley, and P. Ch. Ivanov, "Effect of nonlinear filters on detrended fluctuation analysis," *Physical Review E*, vol. 71, no. 1, Article ID 011104, 2005.

[74] M. Tang and Z. Liu, "Detrended fluctuation analysis of particle condensation on complex networks," *Physica A*, vol. 387, no. 5-6, pp. 1361–1368, 2008.