united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

H4.SMR/1330-9

"Sixth Workshop on Non-Linear Dynamics and Earthquake Prediction"

15 - 27 October 2001

Some Topics on Time Series Analysis

A. M. Correig

Departament d'Astronomia i Meteorologia
Universitat de Barcelona
Barcelona, Spain
e-mail: anton@mizar.am.ub.es

SOME TOPICS ON TIME SERIES ANALYSIS

Antoni M. Correig

(ICTP, Trieste, October 2001)

1 INTRODUCTION

The present notes, mainly conceptual, pretend to give an account on the principles in which time series analysis is based. They are motivated by the fact that very often, methods valid for linear stationarity phenomena are blindly applied to nonlinear ones getting nonsense and useless results. Time series analysis, specially when applied to nonlinear processes, is mainly an art. Quoting Hegger *et. al.* (2000), "it remains a difficult business (...) When treating the programs as a black box, you are certain to get nonsensical results. Even if you know what you are doing, output produced by the programs has to be interpreted very carefully and can often not be taken as it is". We could add that, previous to its application of any method, an analysis should be made about its applicability.

There is actually a vast literature on the subject and it is not my purpose to review it. tentatively I refer to a few representative references, mainly chosen for the clarity of its exposure and because they address the fundamental problems.

- Theoretical foundations of time series analysis: Bendat and Piersol (1986).
- Numerical techniques: Press *et al.* (1992).
- Nonlinear time series analysis: Abarbanel *et al.* (1993) and Kantz *et al.* (1997).
- Numerical techniques for nonlinear analysis: Hegger *et al.* (2000).

2 LINEAR ANALYSIS

Linear analysis has played a key role in the analysis of time series, either deterministic or stochastic, until the 70's of last century when chaotic dynamics was applied to the study of some experiments (such as turbulence) and natural phenomena, for which the well known methods failed to provide any answer.

2.1 Fourier series

The study of natural phenomena is based on observations of the irregular evolution that we consider are typically measurements of a single scalar observable at a fixed spatial point: call, for example the measured quantity $s(t_0+n\tau_s) = s(n)$, where t_0 is some initial time and τ_s is the sampling time used in the observation or experiment. Our actual interest is in the modeling of the dynamics from the point of view of the physicist, interested in the understanding of the system that produces the measured signals. The measured quantities $s(n)$ define an array usually referred to as a **time series**, containing all the information provided by the underlying dynamical system. The analysis of the time series will provide us with some of this information.

Time Series analysis is based on the discovery of the French mathematician Joseph Fourier that any function $f(t)$ satisfying certain restrictions, can be expressed as a sum of an infinite number of sinusoidal terms (J. Fourier, *Theorie analytique de la chaleur*, 1882). The main assumption is that $f(t)$ should be periodic.

Thus, for a periodic function (or, in the present context, a periodic time series),

$$x(t) = x(t \pm nT), \quad n = 1, 2, 3, ... \tag{1}$$

where T is the period, data may be expanded into a Fourier series

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n cosn\omega t + b_n sinn\omega t), \quad \omega = 2\pi f \tag{2}$$

where

$$\begin{aligned} f &= \frac{1}{T} \\ a_n &= \frac{2}{T}\int_0^T x(t)cosn\omega t dt, \quad n = 1, 2, ... \\ b_n &= \frac{2}{T}\int_0^T x(t)sinn\omega t dt, \quad n = 1, 2, ... \end{aligned} \tag{3}$$

In complex form, equation (2) reads

$$x(t) = c_n \sum_{-\infty}^{+\infty} e^{in\omega t} \tag{4}$$

where

$$c_n = \frac{1}{2T} \int_{-\infty}^{+\infty} x(t) e^{-in\omega t} dt. \tag{5}$$

Equations (2) or (4) constitute the **Fourier series theorem.**

Note that a function $f(t)$ defined as a Fourier series is a linear superposition of an infinity of sinusoidal terms defined in the interval $(0, \infty)$, or $(-\infty, +\infty)$ if the series is represented in complex form. From the definition, it follows that $x(t)$ is stationary.

2.2 Stationarity

A time series is said to be stationary if its essential statistical properties do not depend on time. From a mathematical point of view, two types of stationarity are distinguished, strong and weak stationarity. **Strong stationarity** requires the constancy of all the moments of the time series, whereas the **weak stationarity** requires the constancy of the first and second moments only (that is, the mean and the variance). In time series analysis usually it is sufficient to consider the weak stationarity, in the sense that the mean and the variance must be independent of the length of the time series considered.

The **mean value** of continuous data is defined as

$$\overline{x} = \frac{1}{T} \int_0^T x(t) dt \tag{6}$$

and the corresponding value for a discrete set of N measurements, x_i, is

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{7}$$

Mean square value. Is a single figure describing the intensity of random data. It is the average of the squared values of the time-history record, which for continuous data is

$$\overline{x^2} = \frac{1}{T} \int_0^T x^2(t) dt, \tag{8}$$

and for discrete data

$$\overline{x^2} = \frac{1}{N} \sum_{i=1}^{N} (x_i)^2 \tag{9}$$

Variance. Is a measure of the scatter of data about its mean value. For continuous data,

$$V(x) = \frac{1}{T}\int_0^T [x(t) - \overline{x}]^2 dt \tag{10}$$

and for discrete data

$$V(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2. \tag{11}$$

2.3 Superposition

We have already said that the representation of a function $x(t)$ by means of a Fourier series constitutes a superposition of different harmonic components, each one with a given amplitude coefficient. This superposition can be visualized as the contribution of a very large number of harmonic oscillators, each one defined as

$$m\ddot{x} + kx = 0. \tag{12}$$

However, such simple oscillators do not exist in nature. It is more realistic to consider the motion of an oscillator in the presence of damping and subject to an applied external force $F(t)$. The applied force $F(t)$ can be regarded as the sum of two or more other forces $F_1(t), F_2(t), ...$, and the resulting motion $x(t)$ is the sum of motions $x_1(t), x_2(t), ...$ which each of the oscillators would have if each of the forces were acting separately. This is the contents of the principle of superposition, that can be expressed in the following way:

Let the set of functions $x_n(t), \quad n = 1, 2, ...$ be solution of the equation

$$m\ddot{x_n} + b\dot{x_n} + kx_n = F_n(t) \tag{13}$$

where the mass m and the spring constant k may differ for each oscillator, and let

$$F(t) = \sum_n F_n(t). \tag{14}$$

Then the function

$$x(t) = \sum_n x_n(t). \tag{15}$$

satisfies the equation

$$m\ddot{x} + b\dot{x} + kx = F(t). \tag{16}$$

In particular, whenever $F(t)$ can be written as a sum of sinusoidally oscillating terms

$$F(t) = \sum_n C_n \cos(\omega_n t + \theta_n) \tag{17}$$

and if $F(t)$ is periodic

$$F(t+T) = F(t) \tag{18}$$

where T is the period of the force, then $F(t)$ can be written as a Fourier series as defined in (2). It is worth to note that, at the contrary of the solutions of eq. (12), if any of the frequencies $f = 2\pi n/T$ coincides with the natural frequency ω_0 of the oscillator (phenomenon known as *resonance*) then the corresponding term of the solution will be relatively much larger than the rest. For more details see Symon (1963).

2.4 Fourier analysis

A useful approach to the understanding of the dynamics of the system consists on making models of the observation or experiment in the usual variables of velocity, pressure, temperature, voltage, etc. The job of the physicist is then to explore the existence of **invariant quantities** of the model under the dynamics. In the case of linear systems Fourier analysis has shown to be a very powerfool tool. The location of the sharp lines in the Fourier spectrum are characteristic of the specific physical system being measured. If we drive the linear system harder, this would result in more poser under each of these lines, and if we start the system at a different time, the phase of the signal will be altered, but the location of the Fourier peaks will remain unchanged. Thus the characteristic frequencies, usually called resonant frequencies for the linear system, are invariants of the dynamics and can be used to classify the physics. If we see a particular set of frequencies emerging from a driven linear system, we can recognize that set of frequencies at a later data and distinguish the source with confidence.

Let $\vec{x}(t) = (x_1(t), x_2(t), ..., x_n(t)$ be a recorded time series of an observed natural phenomenon. Under the assumptions that conditions (17) and (18) apply, Fourier analysis, also known as harmonic analysis, consists in the retrieval of the frequencies $\omega_n = 2\pi n f$ and the coefficients a_0, a_n and b_n of the Fourier series we assume represents the observed time series. Very often this retrieval is performed by applying a Fast Fourier Transform (FFT) algorithm (see, for example, Brigham (1974)).

2.5 Fourier integral and spectral analysis

Fourier series theorem, equation (2), can be generalized by taking the limit $T \to \infty$, to include nonperiodic forces, as well as to make it suitable for stochastic functions of time. This gen-

eralization, known as the Fourier integral theorem, allows us to represent any continuous (or piecewise continuous) lineal function $F(t)$, subject to certain limitations, as a superposition of harmonically oscillating forces, that is, we drop constrain (18) while maintaining assumption (17) Before proceeding further, let us remind that Fourier analysis provides us with the amplitudes of the frequencies present in the original process. The lowest frequency is known as the fundamental frequency, and the higher frequencies, which are integer multiples of the fundamental frequency, are known as the higher order harmonics. In the case of the spectral analysis (generalized Fourier analysis), the **period** of analysis (but not the signal being analyzed) is assumed to be infinitely long. This is the contents of the Fourier Integral Transformation, where all frequencies are present instead of just those which are integer products of the fundamental frequency.

Once the process of generalization of the Fourier series to the Fourier integral has been carried out, the final result (the equivalent of equations (2) or (4)) consists in the the Fourier transform pair

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \tag{19}$$

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega. \tag{20}$$

$F(\omega)$ is known as the Fourier transform of $f(t)$. Writing the two successive transformations as a repeated integral we obtain the Fourier's integral theorem

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \right] e^{i\omega t} d\omega. \tag{21}$$

For the analysis of digitized data, it is necessary to consider a finite version of the Fourier series, and to derive a discrete form of the Fourier transform. The classical Fourier series is identically computationally to the discrete Fourier transform, although its theoretical derivation is quite different. Similarly to Fourier analysis, spectral analysis is performed in practice by means of a FFT.

In a more general context, Fourier series and Fourier integrals are used to represent and/or approximate functions, and are special instances of expansions in terms of orthogonal functions, conveniently treated in the frame-word of Hilbert space. They follow directly from Weierstrass's theorem for two variables by changing to polar coordinates and restricting the domain of definition to the unit circle, and making use of Euler's formula (for a detailed presentation see. Korn and Korn (1968) and Byron and Fuller (1969)). In this context, it is straightforward to realize that Fourier transform, a kind of integral transform, is a linear process.

2.6 Power spectrum

Let x_j be a discrete time series. The *autocorrelation function* of the signal x_j is defines by

$$\psi_m = -\frac{1}{n}\sum_{j=1}^{N} x_j x_{j+m} \tag{22}$$

where $\psi_m = \psi(m.\delta t$. Physically this function represents the average of the product of the signal values at a given time and at a time $m.\delta t$ later. Therefore it can be deduced from ψ_m whether, and for how long, the instantaneous value of the signal depends on its previous values. It can also be said that the autocorrelation function is a measure of the degree of resemblance of the signal with itself as time passes.

It can be seen (Berge *et al.*, 1984) that the autocorrelation function ψ_k is the Fourier transform of $|X_k|^2$, where X_k is the Fourier transform of x_k. The graph representing $|X_k|^2$ as a function of the frequency f ($f = k.\delta f$) is called is called the *power spectrum* by analogy to the case when the ordinate represents power, *i.e.* energy per unit time. Consider, for example, the record of a wave field such as the ground motion. If the detector is linear and does not introduce distortion in the frequency band considered, then the quantity it measures is proportional to the variation of the wave field. The theory of waves shows that the power carried by a wave is proportional to the square of its amplitude averaged over time. Parseval's theorem states that the average over time can be replaced by an average over frequency, thus justifying the appellation "power spectrum".

From a practical point of view, the power spectrum is simply the squared amplitude of the Fourier transform of the signal. Note that the information about the phase of a component (X_k) is lost when we consider $|X_k|^2$.

The appearance of the power spectrum clearly depends on the way in which the signal $x(t)$ evolves over time. For the case of a pure periodic signal of period T the power spectrum will consist on a simple peak (spectral line) located at a frequency $1/T$, and if we are dealing with periodic functions containing harmonics, the power spectrum will consist on a series of harmonics of the fundamental frequency located at frequencies n/T, $n = 1, 2, ...$ Somewhat more complicated is the case of quasiperiodic signals for which the ratio of the frequency components is an irrational number. This complication arises from the fact that the Fourier spectrum of a quasiperiodic function $x(t)$ (consisting on, say, f_r frequency components), which depend non-linearly on periodic functions of the variables $\omega_i t$, contain components at all frequencies of the form $|m_1 f_1 + m_2 f_2 + ... + m_r f_r|$, where m_i are arbitrary integers.

2.7 Probability distribution functions

In analyzing the weak stationarity of a time series, their statistical properties have been determined through two descriptors, the mean and the variance. These descriptors refer to the statistical description of phenomena that have already occurred. From another point of view, in many cases each of the values $x_i, \quad i = 1, ...N \quad$ of a time series will have a probability of occurrence p_i. By definition

$$\sum_{i=1}^{N} p_i = 1. \tag{23}$$

The probability density function (PDF), $p(x)$, is the probability that any measurement has a value between x and $x + dx$. In terms of PDF, the mean μ and variance σ^2 of a random variable, x, are given by

$$\mu = \int_{-\infty}^{\infty} xp(x)dx \tag{24}$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 p(x)dx. \tag{25}$$

As already mentioned, a series x is called stationary if its statistical parameters do not change with time. A series is weakly stationary (or stationary up to order 2) if only the joint moments up to order 2 of the above probability distributions exist and are identical. This means that the expected values of x and x^2

$$\begin{aligned} E[x] &= \mu \\ E[x^2] &= \mu' \end{aligned} \tag{26}$$

are constants, so that from (25)

$$Var[x] = \sigma^2 \tag{27}$$

is a constant and the weak stationarity is satisfied.

3 NONLINEAR ANALYSIS

It is (implicitly) widely assumed that time series records of natural phenomena are Gaussian processes, that is, follow a normal distribution. Since a normal distribution is completely specified by its mean and variance, it follows that if x is stationary to order 2 it must be completely stationary. If this is the case, the time series is "well behaved" and all available

statistical tools can be applied for its analysis. The only problem with the Gaussian assumption is that, usually, do not hold for natural phenomena, hence new kind on analysis has to be devised. The next two sections, based on the paper by Liebovitch and Scheurle (2000), are devoted to these topics.

3.1 Fractal statistics

Very often data sets recorded from natural phenomena do not follow a Gaussian distribution, but a power law (a straight line in a plot $\log p(x)$ *vs* $\log(x)$), indicating that it has the form

$$p(x) = Ax^{-\alpha} \tag{28}$$

where x is defined in the interval $(0, \infty)$ and $\alpha > 0$. In other words, natural phenomena are self-similar or fractals. As an example, consider the magnitude distribution of earthquakes.

Taking into account equations (24) and (25) we can readily see that the mean value and variance are not defined: they diverge to ∞ as $x \to 0$ and go to 0 as $x \to \infty$. In other words, from this new statistics, there are a few large events and a large number of smaller ones. Hence, the mean value of a data set will depend on the amount of data analyzed. As a consequence the measured time series will not be stationary. This poses a serious problem because most methods of analysis are based on the stationarity of the time series. It must also be noted that power laws are nonlinear, except when the exponent of equation (28) is $\alpha = -1$. Hence, the analysis of natural time series will further require nonlinear techniques. We have already said that the averages measured depend on the amount of data analyzed. Whether the sample means increase or decrease depends on the relative number of small values compared to the large values in the data. If many small values are included as more data are analyzed, then the average will decrease. If there are a few very big values that are included as more data are analyzed, then the average will increase. Which one of these two cases will happen depends on the relative number of small values in the data compared to the large values, that is by the parameter called the 'fractal' dimension. For fractal objects the fractal dimension, d, describes the number of new pieces, N,of an object that are found when the object is viewed at a finer resolution r, namely,

$$d = \frac{log(N)}{log(\frac{1}{r})}. \tag{29}$$

For fractal data, the fractal dimension quantifies the relative number of small values compared to the large values. It is related to the power law form of the PDF. In summary, we need to use appropriate fractal measures, such as the fractal dimension, to characterize fractal data in a meaningful way.

3.2 Chaotic systems

Classical mechanics is based on Newtonian determinism: given a model (a set of rules or equations) and a set of initial conditions, past and future motions can be exactly predicted. Unfortunately this is true only for the (very reduced) class of integrable systems (in the sense of Hamiltonian mechanics: a Hamiltonian system is integrable if all equations can be reduced to quadratures: obviously, a non integrable system can be solved numerically. Note that all linear systems are integrable).

In a classical paper, Henon and Heiles (1964), in a study of a conservative non-integrable Hamiltonian system by solving it by numerical integration, discovered that depending on the values of the energy (a constant of motion, the only one), that can be viewed as a tuning parameter, the solution could be periodic or stochastic. Note that a Hamiltonian system is fully deterministic! This discovery was previously made by Poincare at the beginning of the XX century, but was unnoticed to physicists. The main difference, however, between the works of Poincare and Henon-Heiles is that the last were able to follow the motion of the system in phase space by means of a computer, and thus able to quantify the results.

At the same time, Lorenz (1963) made another seminal discovering: the *sensitivity to initial conditions*. He developed a very simplified model of the motion of air in the atmosphere in a study of predictability of atmospheric phenomena. Starting from Navier-Stokes equation and after truncation, Lorenz gave an approximate description of a horizontal fluid layer heated from below. The warmer fluid formed at the bottom tends to rise creating convection currents. For sufficiently intense heating the time evolution has sensitive dependence on initial conditions, thus representing a very irregular and chaotic convection (butterfly effect). A spectral analysis of a time series generated from this chaotic motion shows that it is neither periodic nor quasi-periodic: simply it is aperiodic. No predominant frequency is present (except in the case of the presence of an external harmonic force, in which case the fundamental, higher harmonics and sub-harmonics could be present in the power spectrum).

The Fourier spectrum of an aperiodic signal is continuous (in the sense of its mean value, computed, for example, by averaging $|X_k|^2$ values locally over a certain number of consecutive values of k). The real difficulty is that a Fourier spectrum which looks continuous cannot be automatically attributed to an aperiodic signal, because this is also the appearance of the spectrum of a quasiperiodic signal with a large number of frequencies.

If we suppose that the signal is truly aperiodic, we must still decide whether the number of degrees of freedom is small (say, less than ten) or, on the contrary, very large. In the first case we have the capability of developing a completely deterministic description of the system. By contrast, in the second case only a probabilistic approach can be achieved with the current state of knowledge.

By similarity to Hamiltonian mechanics, instead of working in *configuration space*, representation of time evolution of the system in position coordinates, an alternative approach consists in working in *phase space*, defined by position and momentum coordinates. In dealing with dynamical systems, phase space is defined in a more general way, in terms of position and velocity, where by position it is understood the measured (observed) parameter, and by velocity its temporal variation. In dealing with observed time series, representation of motion in phase space is usually carried out by the method of *delays* (Packard *et al.*, 1980).

The representation of the evolution of the system (time series) in phase space reveals that the points of the trajectory define what is known as a *strange attractor*, defined by the asymptotic behavior of the chaotic motion. The strange attractor is confined in a subspace which dimension may be fractal.

A very simple example that may present both regular and chaotic motion is provided by the Duffing oscillator, defined as

$$\ddot{x} + r\dot{x} + \omega_0^2 x + \beta x^3 = f\cos(\omega t) \tag{30}$$

where r is the damping coefficient, ω_0 is a parameter (that would correspond to the proper frequency of the system for $\beta = 0$), β is the coefficient of nonlinearity and f is the strength of the harmonic external force. Next figures display some of the key features of chaotic systems.

Roughly speaking, a nonlinear system is that for which the output is not proportional to the input. Mathematically, the signature of a nonlinear system is the breakdown of the superposition principle (defined by equations (13) - (16)): as it can be seen from eq. (30), due to the nonlinear term βx^3, if $x_1(t)$ and $x_2(t)$ are both solutions of the equation, their sum $x_1(t) + x_2(t)$ is not a solution. On the other hand, the breakdown of the superposition may be caused, for a linear equation, by unknown or moving boundary conditions.

The analysis of a nonlinear system is difficult and non-trivial. For example, a small change in the initial conditions may result in large differences in the behavior of the system at later times, and the breakdown of the superposition principle renders the Fourier transform technique useless: resonances are no longer invariants of the system. We are thus on the need of some other kind of invariants with the following properties: these invariants, quantities that are unchanged under various operations on the dynamics of the orbits, are all unchanged under small variations in initial conditions. We have already said that the asymptotic motion in phase space is 'trapped' by the attractor, and that the attractor can be found by transforming the time series into the phase space by the method of delays. Let us emphasize that it is the attractor, nor the time series, that is reproducible from experiment to experiment. The time series is different in each experiment, but the attractor is the same in each experiment. Thus the attractor (not the time series, not the individual orbits in phase space) is the true invariant

of the system, and the only meaningful measures are those found from the properties of the attractor that can be generated from a transformation of the time series.

3.3 Characterization of the attractor

Before proceeding to its characterization let us first define what a strange attractor is. Following Schuster (1995), consider a chaotic dissipative system, described either by a flow or a map. For simplicity, consider the case of a flow. A flow is described by a set of autonomous first-order differential equations

$$\dot{\vec{x}} = \vec{F}(\vec{x}), \quad \vec{x} = (x_1, x_2, ..., x_d) \tag{31}$$

and the term dissipative means that an arbitrary volume element V enclosed by some surface S in phase space $|\vec{x}|$ contracts. The surface S evolves by having each point on it an orbit generated by (31). This yields, by divergence theorem,

$$\frac{dV}{dt} = \int_V d^d x \left(\sum_{i=1}^{d} \frac{\partial F_i}{\partial x_i} \right) \tag{32}$$

and dissipative systems are defined by $dV/dt < 0$. It is found that the trajectories generated by flow (31) (consider, for example, the Lorenz model) are *i)* attracted to a bounded region in phase space, *ii)* the motion is erratic, and *iii)* there is a sensitive dependence of trajectories on the initial conditions. Let us see that the trajectories are attracted to a bounded region in phase space, and under the assumption that the system is dissipative, the volume of this region contracts to zero. This means that the d-dimensional system generates a set of points whose dimension is less than d and necessarily has to be greater than 2; *i.e.*, its volume in d-dimensional space is zero. Very often, this dimension is not an integer, *i.e.* it is a fractal. This leads, in a natural way, to the concept of *strange attractor.*

A strange attractor has the following properties:

a) It is an attractor, *i.e.*, a bounded region of phase space $|\vec{x}|$ to which all sufficiently close trajectories from the so-called basin of attraction are attracted asymptotically for long enough times. Furthermore, the attractor itself should be indecomposable, that is, the trajectory should visit every point on the attractor in the course of time.

b) The property which makes the attractor strange is the sensitive dependence on the initial conditions; *i.e.*, despite the contraction of the volume, lengths need not shrink in all directions, and points, which are arbitrarily close initially, become exponentially separated at the attractor for suficienly long times.

The above two properties tells us that the attractor can be characterized by mainly two invariants: a geometrical one, the *dimension of the attractor*, and a dynamical one, the *Lyapunov exponent*; both can be retrieved from observed time series. As said these invariants are independent of the initial conditions and provide a direct analogy to the Fourier frequencies of a linear dynamical system. The following two sections follow Abarbanel *et.al.* (1993).

3.3.1 Dimension

Beyond the simplest concept of dimension as the number of coordinates needed to specify a state is the geometrically related concept of how (hyper) volumes scales as a function of a characteristic length parameter:

$$V \propto L^D \tag{33}$$

Planar areas scale quadratically with the length of a side, and volumes in rteal world space go as the cube of the side length. Contingent upon some workable generalization of the idea of volume, eq. (33) can be inverted to 'define' a dimension:

$$D = \frac{\log V}{\log L} \tag{34}$$

Early techniques for estimating the dimension used a covering of the attractor to calculate V. Consider a partitioning of the d_E-dimensional phase space with an ϵ-sized grid. Count how many of the partition elements contain at least on point from the sample data, and use this value as the measure of V at resolution ϵ. Then refine the phase space partition by decreasing ϵ and count again how many of these smaller partition elements are not empty, and repeat the process until ϵ has spanned several orders of magnitude. In theory, in the limit as $\epsilon \to \infty$ the ratio in eq. (34) describes the space-filling properties of the point set being analyzed. this procedure constitutes the *box-counting* method. Intuitively eq. (33) takes the amount of an object to be equivalent to its volume, but this is not the only measure for how much of something there is. Eq. (33) has the more general form

$$bulk \propto size^{dimension}. \tag{35}$$

Consider now the case in which the bulk is measured in terms of some moment of the invariant distribution. The invariant distribution or or invariant measure on the attractor is the natural density of points within that volume. The definition of this density is

$$\rho(\vec{x}) = \frac{1}{N} \sum_{k=1}^{N} \delta^d(\vec{x} - \vec{y}(k)) \tag{36}$$

in phase space of dimension d. It can be shown that any function of phase space $g(\vec{x}) \to \vec{F}(\vec{x})$.

Define the moment $< g >$ as

$$< g >= \int d^d x g(\vec{x})\rho(\vec{x}) = \frac{1}{N}\sum_{k=1}^{N} g(\vec{y}(k)), \tag{37}$$

take $g(\vec{x}) = \rho^p(\vec{x})$ in eq. (36) and use $< g >^{1/p}$ to measurew the bulk of eq. (35). The following result is obtained:

$$D = \lim_{r\to 0} \frac{1}{q-1}\frac{\log \sum \rho^q}{\log r} = D_q. \tag{38}$$

This definition of the *generalized dimension* D_q provides a whole spectrum of invariant quantities for $-\infty < q < \infty$.

The case $q = 0$,

$$D_0 = -\frac{\log N_0}{\log r} \tag{39}$$

is just the capacity dimension estimated by the box-counting method.

Next consider D_1. Taking the limit $q \to 1$ results in

$$D_1 = \lim_{r\to 0} \frac{\sum_i p_i \log p_i}{\log r}. \tag{40}$$

Note that, when $p_i = 1/N$ for all i, then $D_1 = D_0$. But when the p_i are not all equall $D_i \leq D_0$.

The difference between the two measures is explained by how we use our phase-space paratition. For D_0 the bulk is how many partitions elements are nonempty. D_1 is calculated from a bulk in which each partition element's contribution is proportional to how many samples it contains. So bulk in D_0 is simply the volume, whereas for D_1 bulk is equivalent to the mass. Because of the functional form of the numerator in eq. (40), D_1 is commonly called the *information dimension.* That numerator express ther amount of information needed to specify a state to precision r, and the ratio defining the dimension tells us how fast this required information grows as r is decreased.

Finally, let us consider D_2, the so-called *correlation dimension*:

$$D_2 = \lim_{r\to 0} \frac{-\log \sum_i p_i^2}{\log r}. \tag{41}$$

We can see that the numerator constitutes a two-point correlation function, measuring the probability of finding a pair of random points within a given paratition element, just as the numerator in the definition of D_1 measures the probability of finding one point in a given element. The numerical value of that numerator can be estimated by counting how many pairs of points have a separation distance less than some value r. (Grassberger and Procaccia (1983), see Abarbanel et al (1993)) suggested using

$$C_2(r) = \int d^d x \rho(\vec{x}) n(r, \vec{x}) = \frac{2}{N(N-1)}\sum_{i\neq j}^{N} \theta(r - |\vec{y}(j) - \vec{y}(i)|) \tag{42}$$

as a simple and computationally efficient means of estimating $\sum_i p_i^2$.

3.3.2 Lyapunov spectrum

Whereas dimensions characterize the distribution of points in phase space, Lyapunov exponents describe the action of the dynamics defining the evolution of trajectories. Dimensions only deal with the way measure is distributed through the space, whereas Lyapunov exponents examine the structure of the time-ordered points making up a trajectory.

To understand the significance of the spectrum of Lyapunov exponents, consider the effect of the dynamics on a small spherical fiducial hipervolume in the (possibly reconstructed from observed time series) subspace. Arabitrarily complicated dynamics, like those associated with chaotic systems, can cause the fiducial element to evolve into extremely complex shapes. However, for small enough length scales and short enough time scales the initial effect of the dynamics will be to distort the evolving spheroid into an ellipsoidal shape, with some directions being stretched and others contracted. The primary, longest, axis of this ellipsoid will correspond to the most unstable direction of the flow, and the asymptotic rate of expansion of this axis is what is measurwed by the largest Lyapunov exponent. More precisely, if the infinitesimal radius of the initial fiducial volume is called $r(0)$, and the length of the ith principal axis at time t is called $l_i(t)$, then the ith Lyapunov exponent can be defined as

$$\lambda_i = \lim_{t\to\infty} \frac{1}{t} \log \frac{l_i(t)}{r(0)}. \tag{43}$$

By convention, the Lyapunov exponents are always ordered so that $\lambda_1 > lambda_2 > \lambda_3$.... From a measured time series, usually only the greatest Lyapunov exponent can be computed with precision.

3.4 Kolmogorov entropy

In computing the Lyapunov exponents, it is not taken into account the possible dependence on the initial point. Thie dependence is included in the definition of the Kolmogorov-entropy, or K-entropy for revity:

$$K = \int_M \rho(x) \sum{}' \lambda_i(x) dx, \tag{44}$$

where $\rho(x)$ represents an invariant distribution function, normalized to unity, on the attractor M, and the prime on the sum denotes that the sum is only over the positive Lyaponov exponents. Mosr frequently systems are considered that have Lyapunov exponents independent of the point on the attractor and thus because the distribution function is normalized, the K-entropy is given by

$$K = \sum{}' \lambda_i. \tag{45}$$

The K-entropy represents the rate of information production and the growth of uncertainty. In a chaotic system with $K > 0$ the number of possible states of the system that evolve from some initial distribution increases with time. Hence information is increasing. Another way of viewing this information grow focuses on making a measurement of the system point in phase space. In a system with $K > 0$, a determination of the system point to within a minimum uncertainty at time $t > 0$ determines the initial state of the system to within an uncertainty at $t = 0$ much less than the minimum. Consequently, athere has been a growth in information. As a chaotic system evolves the subsequent knowledge about the satate of the system, as determined by the initiasl state, deteriorates. The system is generating information and unless a subsequent measurement is performed, one knows less and less about the system.

Following Schuster (1995), the K-entropy can thus be used to determine the average time over which thwe state of a system, displaying detrerministic chaos, can be predicted. Consider, for example, a one-dimensional map, located on a unit squae, for which, after n steps an interval l increases to $L = le^{\lambda n}$, where λ is the Lyapunov exponent. If L becomes larger than 1 (that is, with a positive Lyapunov exponent), we can no longer locate the trajectory in $[0, 1]$, and all we can say is that the system has a

$$p(x) = \rho_0 dx \tag{46}$$

of being in an interval $[x, x + dx] \in [0, 1]$, where $\rho_0(x)$ is the invariant density odf the system. In other words, precise prediction about the state of this system are only possible for times n that are smaller T_m:

$$le^{\lambda T_m} = 1 \rightarrow T_m = \frac{1}{\lambda} \log\left(\frac{1}{l}\right). \tag{47}$$

Above T_m one can only make statistical predictions. Eq. (46) casn be generalyzed to higher dimensional dynamicasl sysems by replacing λ by the K-entropy

$$T_m \propto \frac{1}{K} \log\left(\frac{1}{l}\right). \tag{48}$$

Note that the precision l, with which the initial state is located, only influences T_m logarithmically.

REFERENCES

Abarbanel, H.D.I., Brown, R., Sidirowich, J.J. and Tsimiring L.Sh. (1993). The analysis of observed chaotic data in physical systems, *Rev. Mod. Phys.*, 65, 1331-1392.

Bendat, J.S. and Piersol, A.G. (1986). *Random Data. Analysis and measurement procedures*, Wiley.

Berge, P., Pomeau, Y. and Vidal, C. (1984). *Order within chaos*, Wiley.

Brigham, E.O. (1974). *The fast Fourier transform.* Prentice-Hall.

Byron, F.W. and Fuller, R.W. (1969). *Mathematics of classical and quantum physics.* Addison-Wesley.

Hegger, R., Kantz, H. and Schreiber, T. (2000). TISEAN-2.1, Nonlinear Time Series Analysis. www.mpipks-dresden.mpg.de/ tisean

Henon, M. and Heiles, C. (1964). The applicability of the third integral of the motion: some numerical results, *Astron. J.*, 69, 73.

Kantz, H. and Schreiber (1997). *Nonlinear time series analysis*, Cambridge.

Korn, G.A. and Korn, T.M. (1968). *Mathematical handbook for scientists and engineers.* McGraw-Hill.

Liebovitch, L.S. and Scheurle, D. (2000). Two lessons from fractals and chaos, *Complexity*, 5, 34-43.

Lorenz, E.N. (1963). Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20, 130.

Packard, N.H., Crutchfield, J.F., Farmer, J.D. and Shaw, R.S. (1980). Geometry from a time weries, *Phys. TRev. Lett.*, 45, 712.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992). Numerical Recipes in Fortran. The art of scientific computing, Cambridge.

Schuster, H.G. (1995). Deterministic chaos, VCH.

Symon, K.R. (1963). *Mechanics*, Addison-Wesley.