Introduction to Classical Chaos

WeiHan Hsiao[a]

[a] *Department of Physics, The University of Chicago*

E-mail: weihanhsiao@uchicago.edu

ABSTRACT: This note is a contribution to Kadanoff Center for Theoretical Physics journal club meeting in 2017 spring quarter. In this talk, we review/introduce the ideas/notions emerging in classical chaos problems. In particular, going beyond history telling, we emphasize quantitative measures that can be generalized to quantum mechanical realm.

Contents

1 Introduction

Tracing back to the flow of history, as the author knows it the first chaotic phenomenon emerged when Poincaré studied 3-body problem around 1890s. In his works, he realized and pointed out that the problem is no longer integrable, and moreover, the numerical solution depends extremely sensitively on initial conditions. One may reference Ref. [1] for more elaboration. Later on in mid-20th century, when Edward Lorenz tried to model the atmospheric problem using 3 coupled non-linear ordinary differential equations [2], he discovered similar sensitivity to initial conditions, implying practical impossibility to predict weather in a large enough amount of time. Therefore the quote by him goes as

> "When the present determines the future, but the approximate present does not approximately determine the future."

Besides, he recognized that in phase space, the seemingly unpredictable trajectories do not scatter all over the place. Instead, they actually approach a subspace, the attractor, of the whole possible state space. With some effort, it can be shown that the dimension of such attractor lies between 2 and 3. Moreover, the aforementioned sensitivity to initial condition manifests itself in terms of instability on the attractor, and thus the attractor is dubbed the name *strange.*[1]

[1] The definition of strangeness here follows one in Ref. [4]. One can easily find different opinions in the literature, say Ref. [5].

Nowadays, we have realized more. In addition to the Hamiltonian dynamical systems raised above, similar phenomena appear in many other models in the realm of quantitative science. The simplest example may be the logistic map as a kindergarten model describing population growth, that $x_{n+1} = ax_n(1 - x_n)$, where a determines the *growth rate*. It can be shown easily, at least at numerical level, for most values of $a > 3.6$, the system is chaotic (and of course we have to specify the meaning of *chaotic* more precisely in the following sections). More examples include, but are not limited to, the double pendulum, Hénon map, that $x_{n+1} = 1 - ax_n^2 + y_n, y_{n+1} = bx_n$, and Chirikov map, that $p_{n+1} = p_n - K \sin\theta_n, \theta_{n+1} = \theta_n + p_{n+1}$. See Ref. [3].

The talk and this note is organized as follows. In the next section, we review some qualitative features that are commonly related or adopted to be criteria of chaos. To be illustrative, we try to visualize some of them in terms of (numerical solutions) for some concrete models. Next we introduce several quantities that measure the extent of chaos including *Lyapunov Exponent* and *Kolmogorov-Sinai (KS) Entropy*. In the final section, we give some comment to conclude the talk.

There are no doubt literatures on this subject, yet the author had a difficult time to find a single reference that fits into the structure and the size of the agenda here. Refs. [3–5] are comprehensive but maybe a bit time-consuming.

A particularly good introduction (in my opinion) is the online lecture notes by Michael Cross: http://www.cmp.caltech.edu/~mcc/Chaos_Course/, from which the author have borrowed many arguments.

2 Qualitative Features of Chaos

As the authors knows, despite lacking universal formal definitions, most chaotic systems have the features that we will introduce shortly and their mathematical definitions are borrowed from Ref. [6].

2.1 Topological Transitivity

A map f mapping the space $\mathcal{M} \to \mathcal{M}$ is topologically transitive if for any pair of open sets U and V in $\mathcal{M}$, there exists a natural number N such that $f^N(U) \cap V$ is non-empty.

An immediate outcome of this property is that for a chaotic system there is no *innocent* region in the state space. For any point in the space, there is an actual trajectory that eventually comes arbitrarily close to it.

2.2 Initial Condition Sensitivity

The map f from $\mathcal{M}$ to $\mathcal{M}$ has sensitive dependence on initial conditions if there exists $\epsilon > 0$ such that for any $x \in \mathcal{M}$ and any neighborhood $\mathcal{B}$ of x, there exists a $y \in \mathcal{B}$ and an integral number $N \neq 0$ such that $||f^N(x), f^N(y)|| > \epsilon$.

This property, sometimes dubbed as butterfly effect, is probably the most well-recognized feature of chaos. In plain English, it means that there is some distance such that no matter how little we perturb the system, there is some subset in our perturbation set differ from original trajectory at least by ϵ after long enough time.

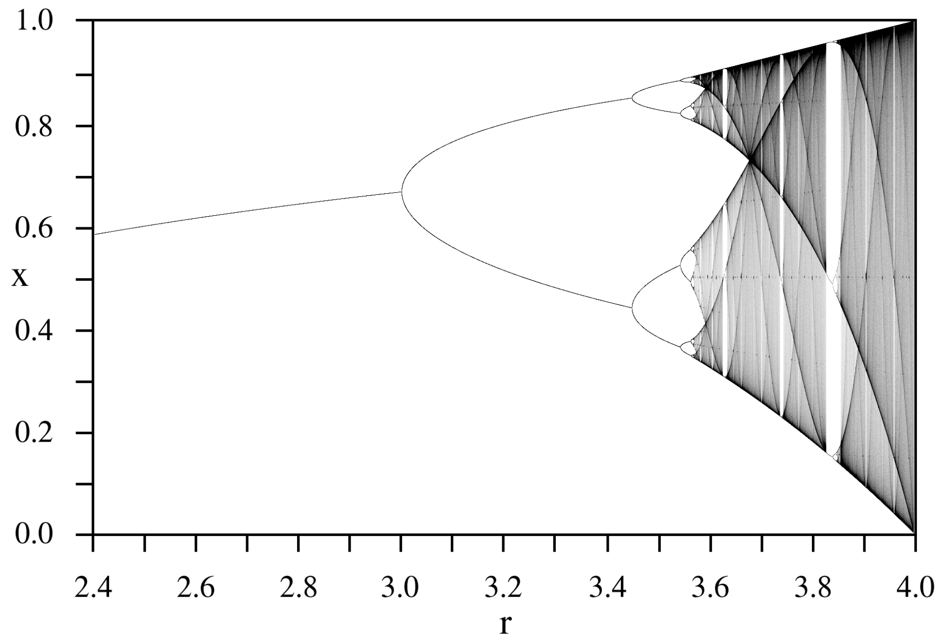

Figure 1.

2.3 Periodic Orbit Density

This property is also referred as *element of regularity*. A system possess an element of regularity if there exists a dense subset of periodic points in $\mathcal{M}$. That is, for any $x \in \mathcal{M}$, there is a point y that continually iterates back arbitrarily close to original point x.

2.4 Attractors

In some systems, they contain additionally a structure called strange attractors. In those systems the dynamics, or the map, has some points. Using the simplest $x_{n+1} = f(x_n)$ as the example, there are some points in state or phase space that map them to themselves. They form an attractor set. Taking the logistic map $x_{n+1} = rx_n(1 - x_n)$ as a toy example. When $r = 2$, the attractor contains 1 point, which is the fixed point of the map. When $r = 3.2$, the attractor contains 2 points. As r goes to 3.5, it further bifurcates to 4 points. These attractors look regular, each of them containing finite points. Nevertheless, the story becomes dramatic as r exceeds, say 3.57. The attractor becomes chaotic. It seems to cover the whole space, but actually not quite does so. To study such complex geometric structures, we try to characterize them by a non-integral dimension, and hence the name fractal. We postpone the technical details to the next section. Figure 1 borrows the celebrated bifurcation diagram from Wikipedia.

Nonetheless, in order to visualize some of these properties. Let us raise the classic Lorenz model as the example. The Lorenz model is defined by the following equations [2].

$$\frac{\mathrm{d}X}{\mathrm{d}t} = -\sigma X + \sigma Y \tag{2.1}$$

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = -XZ + rX - Y \tag{2.2}$$

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = XY - bZ, \tag{2.3}$$

with $\sigma = 10, b = 8/3$ and $r = 28$. Some visualizations are shown in figure 2.

As a short summary, we have introduced some general features that many people use to recognize chaotic phenomena. Though being stated in a rather formal manner, they are still qualitative and we will have to do more, which leads us to next section.

To see the necessity, the most common feature of chaos is the loss of predicability after long time and the result looks *random.* However, it should be emphasized that it is not really *random* or *noisy* compared to authentic randomly generated noise. Quite the contrary, it has some internal pattern if we look at the solutions from a proper view point. Having some quantitative measures also helps us to distinguish chaos from noise. A nice visual demonstration of such distinction can be found in Ref. [7].

3 Quantitative Measure of Chaos

In this section we will be studying some quantitative measures of chaos. Therefore, we should go beyond qualitative classification and seek some measures that help us to quantify chaos. I think it may be helpful first to introduce some more concepts.

Probability and invariant measure is one of them since we will talk about entropy shortly. From empirical point of view, the probability is actually the relative frequency of the occurrences of repeated trials. In the context here, we focus on the neighborhoods of the attractors in the state space. We divide the space around into small cells (boxes) and the probability p_i will be proportional to the times that a trajectory visits a cell i. An immediate question follows that which trajectory we are referring to? As we have been reminded again and again that a chaotic system is sensitive to where we start a trajectory. In many chaotic system, it is known that p_is actually do not depend on where we start the trajectory on the attractor, and nor do they depend on the unit scale of coordinates. In this sense p_i is called an invariant distribution or measure. For some models, it can be written as $p_i = \int_i d\tau\, \rho(x)$ for some probability density.

Given this distribution defined on state space, we are then able to define 2 kinds of average for a quantity $\mathcal{O}(x)$ depending on the state space variable x. One is the time average

$$\langle \mathcal{O} \rangle_t = \frac{1}{T} \int_0^T dt\, \mathcal{O}(x(t)) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(x(t_i)), \tag{3.1}$$

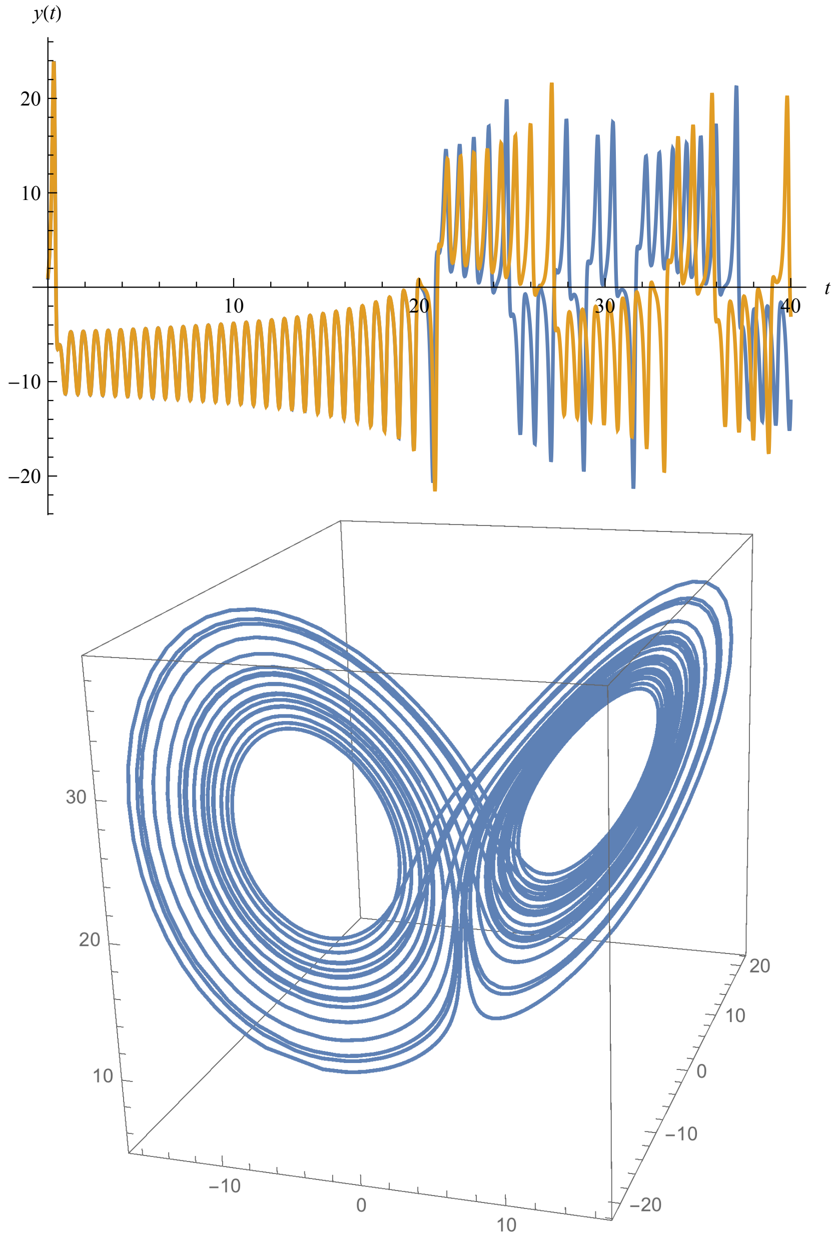

Figure 2. The top panel shows the sensitivity to initial conditions. As we can see, the blue and brown trajectory start with almost the name initial value, yet around $t = 25$ they become dramatically different. The bottom panel depicts the attractor in Lorenz system.

while the other is the state space average

$$\langle \mathcal{O} \rangle_p = \int d\tau \, \rho(x) \mathcal{O}(x) = \frac{1}{N'} \sum_{i=1}^{N'} \mathcal{O}(x_i) p_i. \tag{3.2}$$

If these 2 calculations happen to yield the same result, the system is called ergodic.

3.1 Lyapunov Exponent

One of the features that all chaotic systems embrace is, as we have mentioned repeatedly, sensitivity to initial conditions. The Lyapunov exponent, or the characteristic exponent, is a quantity that rephrases the statement.

To start with, let us first consider a discrete time one dimensional map $x_{n+1} = f(x_n)$. If we can formulate such a quantity for this toy model, with some modification we should be able to generalize it to more complicated systems.

Given this map, we may consider 2 initial values x_0 and x'_0, which become x_N and x'_N after N iterations. The difference between these 2 trajectories is identified as

$$\delta x_N = x_N - x'_N = f^N(x_0) - f^N(x'_0) \simeq \frac{\mathrm{d}f^N(x_0)}{\mathrm{d}x}(x_0 - x'_0) = \frac{\mathrm{d}f^N(x_0)}{\mathrm{d}x}\delta x_0. \tag{3.3}$$

The derivative of composition map $f^N(x) = f(f(...f(x)...))$ can be evaluated using chain rule.

$$\frac{\mathrm{d}f^N(x_0)}{\mathrm{d}x} = \frac{\mathrm{d}f(x_{N-1})}{\mathrm{d}x}\frac{\mathrm{d}f(x_{N-2})}{\mathrm{d}x}...\frac{\mathrm{d}f(x_0)}{\mathrm{d}x}. \tag{3.4}$$

From such a product structure and an anstaz $\delta x_N = e^{\lambda N}\delta x_0$, we can define, for this simple model,

$$\lambda = \lim_{N\to\infty}\frac{1}{N}\log\Big|\frac{\mathrm{d}f^N(x_0)}{\mathrm{d}x}\Big| = \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}\log\Big|\frac{\mathrm{d}f(x_k)}{\mathrm{d}x}\Big|. \tag{3.5}$$

This starting point has an obvious and immediate extension. f can be regarded as a map $f : \mathbb{R}^n \to \mathbb{R}^n$, while $x \in \mathbb{R}^n$. Denote the differential map $\partial f_i/\partial x_j = T_{ij}$. Then $\partial(f^N)_i/\partial x_j$ becomes a matrix product

$$T^N(x) = \frac{\partial f^N}{\partial x} = T(f^{N-1}(x))...T(f(x))T(x). \tag{3.6}$$

The next step is to extract some numbers out of this matrix. If the system is ergodic, Oseledec?s multiplicative ergodic theorem then ensures the existence of the following limit.

$$\lim_{N\to\infty}((T^N)^T T^N)^{1/2N} = \Lambda. \tag{3.7}$$

We erase the x dependence owing to the ergodic assumption. The logarithms of the eigenvalues of Λ are Lyapunov exponents. Another intuitive approach is, for given initial point or time, we try to find the normalized eigenvectors v_i to T^N. Then

$$\lambda^{(i)} = \lim_{N\to\infty}\frac{1}{N}\log||T^N v_i||. \tag{3.8}$$

For general complex systems, there are technical details that we have to deal with. To have some taste, let us take Arnold's cat map as an example. Let x_1 (mod 1) and x_2 (mod 2) be the coordinates in state space (2 torus). The cat map is defined as $f : T^2 \to T^2$.

$$f\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{pmatrix}x_1 + x_2\\x_1 + 2x_2\end{pmatrix} \text{ (mod 1)}. \tag{3.9}$$

This map is known to be chaotic. If we write $\delta x_N = \delta x_0 \, e^{\lambda N}$. The Lyapunov exponent is the logarithm of the larger eigenvalue of the matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \tag{3.10}$$

which is

$$\lambda = \log \frac{1}{2}(3 + \sqrt{5}). \tag{3.11}$$

3.2 Kolmogorov-Sinai Entropy

Given the definition of p_i, we are mimic what people have done in thermodynamics and information theory to define an *entropy*

$$S = -\sum_{i=1}^{N} p_i \log p_i. \tag{3.12}$$

We, however, have to be careful about the meaning of this quantity. A choice of partition is made when we define the probability p_i.

There are 2 related quantities giving more intrinsic properties. One is the information dimension, which tells how this entropy scales with the size of the boxes defining its partition. We will be talking about it in next section. The other quantifies how information evolves in time or under each iteration and is the main character of this section, the Kolmogorov-Sinai, or K-entropy.

Roughly speaking, the K-entropy tells us how our (precision of) knowledge of the state x decreases under time evolution, which reflects the sensitivity to initial conditions. Another way to put it is how the knowledge is improved if we pull things back in time.

To be more concrete, suppose we start with a primitive partition $\mathfrak{z}_0 = \{B_i\}$, where B_i can be regarded as some small volumes covering the attractor. The time evolution or iteration is done by the map f. For each volume B_i we define the preimage $f^{-1}(B_i)$, which is weaker than the *inverse map.* The partition of $\mathfrak{z}_0$, to some extent, reflects the precision we have about the evolution. Let's say we construct the partition in a way that all we can say about a measurement x is which bin it lies in. With these assumptions, if $x_0 \in B_{i_0}$ and $f(x_0) = x_1 \in B_{i_1}$, our knowledge of x_0 is refined. We know it is actually lying in $B_{i_0} \cap f^{-1}(B_{i_1})$. This way, we are capable of defining a *better* partition $\mathfrak{z}_1 = \{B_i \cap f^{-1}(B_j)\}$ and computing $S(\mathfrak{z}_1)$. That the precision gets better as we pull the iteration back entails the precision becomes worse in time, and in this sense the K-entropy quantifies how information evolves. The refined partition after n-iteration is

$$\mathfrak{z}_n = \{B_{i_0} \cap f^{-1}(B_{i_1}) \cap f^{-2}(B_{i_2}) \cap ... \cap f^{-n}(B_{i_n})\}. \tag{3.13}$$

The K-entropy is

$$K = \lim_{n\to\infty} \frac{1}{n} S(\mathfrak{z}_n). \tag{3.14}$$

Again to reflect the property of sensitivity to initial values, this quantity is related to Lyapunov Exponents by a bound

$$K \leq \sum_{\lambda^{(i)}>0} \lambda^{(i)} \tag{3.15}$$

and the bound is saturated for a certain class of attractors. This can be motivated as follows. Imagine that at $t = 0$ we start with small boxes of side ϵ which define the resolution. After n-iteration, the size of each box expands by a factor $e^{n(\lambda^{(1)}+\lambda^{(2)}\ldots)}$ which redefines the resolution later in time.

3.3 Fractal Dimensions

As it is pointed out in last section, to study the complicated geometric structure of strange attractors, non-integral dimension is a characterization.

To start with, we first define the so-called *capacity* or *box counting* (in $\mathbb{R}^n$). Suppose the set of interest belongs to a subset in $\mathbb{R}^n$. We cover it with $N(\epsilon)$ n-cube of side ϵ. The *Capacity* is defined as

$$d_C = -\lim_{\epsilon\to 0} \frac{\log N(\epsilon)}{\log \epsilon}. \tag{3.16}$$

Let us first apply this definition to a standard example to make reason of it. We consider a straight line of length ℓ in $\mathbb{R}$. The cubic here is merely a segment of length ϵ. The number needed to cover ℓ is $N(\epsilon) = \ell/\epsilon$ and therefore

$$d_C = -\lim_{\epsilon\to 0} \frac{\log \ell - \log \epsilon}{\log \epsilon} = 1. \tag{3.17}$$

Another example can be a square of side ℓ embedded in $\mathbb{R}^2$. $N(\epsilon) = \ell^2/\epsilon^2$. Then clearly $d_C = 2$.

Next we look at 2 fractal examples. The first one is the famous 1/3 Cantor set, which can be produced from $[0, 1]$ by removing iteratively the middle 1/3 of each line element at each step. The length, or the Lebesgue measure or the remaining set is 0, yet the set still contains infinitely many points. At nth iteration, there are 2^n remaining segments, each of which is of length $\epsilon = (1/3)^n$. Consequently,

$$d_C = \lim_{n\to\infty} \frac{n \log 2}{n \log 3} = \frac{\log 2}{\log 3} = 0.63... \tag{3.18}$$

Another exactly computable example is Koch curve. At each stage, we need 4^n boxes, each of which is of size $(1/3)^n$. Therefore,

$$d_C = \lim_{n\to\infty} \frac{n \log 4}{n \log 3} = \frac{\log 4}{\log 3} = 1.26. \tag{3.19}$$

Contrast to the 0-length of 1/3 Cantor set, the Koch curve has ∞ length.

In addition to Capacity, *Hausdorff* dimension is another similar measure that gives more sensible results in some cases. The problem is that either of them and the examples we look at are defined from a pure geometrical perspective, while in chaotic problems the attractors are determined by dynamics. The measures of attractors may make them not applicable to real studies. However, d_C defined above

still provides us a foundation to think about general dimensions.
Following the same idea of box counting, again we cover the attractor of interest with n-boxes of side ϵ and assign a box label i to each box. Suppose there are N_i out of total N points in the box. We define the probability $p_i = N_i/N$.[2] The qth *Generalized Dimension* d_n is defined as

$$d_n = \lim_{\epsilon\to 0} \frac{1}{n-1} \frac{\log \sum_i {p_i}^n}{\log \epsilon}. \tag{3.20}$$

As $n = 0$, $d_0 = d_C$.
As n approaches 1, writing $n = 1 + \delta$ and bearing $\delta \to 0$ in mind,

$$\begin{aligned}\frac{1}{n-1} \log \sum_i p_i^n =& \frac{1}{\delta} \log \sum_i p_i e^{\delta \log p_i} = \frac{1}{\delta} \log \left(1 + \delta \sum_i p_i \log p_i\right) + \mathcal{O}(\delta^2) \\ =& \sum_i p_i \log p_i + \mathcal{O}(\delta).\end{aligned}$$

Consequently,

$$d_1 = \lim_{\epsilon\to 0} \frac{\sum_i p_i \log p_i}{\log \epsilon}. \tag{3.21}$$

The reduction of d_n to d_1 is formally exactly the same as one of Rényi entropy to von Neumann entropy. Usually the dimension of a geometry object determines how it scales with the box size ϵ. d_1, similarly, gives how the information content $\sum_i p_i \log p_i$ scales with ϵ.
The last dimension I would like to introduce here is the *Lyapunov Dimension*. It was proposed by Kaplan and Yorke to express dimensions in terms of dynamical quantities, Lyapunov exponents. For a dissipative system, we know the volume of states in the phase space contracts, implying the sum of all Lyapunov exponents should be negative. The fact that some exponents must be negative supports the existence of attractors, while the chaotic phenomena appearing on the attractor are attributed to those positives λs. We define $\mu(n) = \sum_{i=1}^n \lambda_i$ as the sum of the n largest λ_i in the Lyapunov spectrum. Between 2 integers n and $n+1$, it is defined by linear interpolation

$$\mu(t) = \sum_{i=1}^n \lambda_i + (t-n)\lambda_{n+1}. \tag{3.22}$$

The Lyapunov dimension is then defined as

$$d_L = \max\{t, \mu(t) \geq 0\}. \tag{3.23}$$

Suppose $\mu(n) > 0$ and $\mu(n+1) < 0$. The extremum can be evaluated as

$$d_L = n + \frac{1}{|\lambda_{n+1}|} \sum_{i=1}^n \lambda_i. \tag{3.24}$$

[2] We have avoid diving into defining the invariant (physical) measure ρ. p_i can be viewed as an estimate of $\int_{V_i} \mathrm{d}\tau\, \rho(x)$.

As we have pointed out, the dimension of a space, roughly speaking, tells us how it scales with the size of boxes used to cover it. The sum of Lyapunov exponents tells how a phase space volume expands or contracts along time evolution. d_L can be interpreted as the dimension at which the volume stays invariant. The conjecture by Kaplan and Yorke is $d_L = d_1$.

In addition to these 3 quantities, there are some other quantitative criteria for the onset of chaos. Examples include *decay of autocorrelation function* and *broad distribution of frequencies in power spectrum*.

In most literatures, strong criteria for chaos are the existence of a positive Lyapunov exponent and positive Kolmogorov-Sinai entropy[3].

4 Summary

In this talk and note we have review/introduce both qualitative feature and quantitative measures for classical chaotic systems. Some qualitative features are visualized with figures. As for quantitative measures we do our best to raise some simple analytical soluble toy examples, though much care is required when serious numerical calculation concerns us.

With the quantities defined in last section, I think, at philosophical level, they opened another route to think about physical problems. In most physics curriculums we started with some exactly soluble models, thoroughly discussing their solutions and properties. Next, we often turned on perturbations and utilized perturbation machinery to refine the existing solutions. Such approach strongly depends on the property that the systems there are separable and the linear nature of the unperturbed models, that is to say, we are capable of talking about a harmonic mode, or some of them. It is not the case for nonlinear dynamics. It is seen from the spectrum analysis that one needs infinitely many modes to deal with non-linear problems.

Instead, for nonlinear problems, we have to utilize quantities such as Lyapunov exponents and K-entropy to help us characterize systems.

References

[1] Alain Chenciner, *Poincaré and the Three-Body Problem*, Séminaire Poincaré XVI, 45-133 (2012).

[2] E. Lorenz, Journal of the Atmospheric Science, **20**, (2): 130-141 (1963).

[3] H. G. Schuster and W. Just, *Deterministic Chaos: An Introduction*, 4th Ed. (2005).

[4] J. P. Eckmann and D. Ruelle, *Ergodic Theory of Chaos and Strange Attractors*, Rev. Mod. Phys. **57**, 617 (1985).

[5] R. C. Hilborn, *Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers* (1994).

[6] J. T. Townsend, *Chaos Theory: A Brief Tutorial and Discussion*

[7] Geoff Boeing, *Visual Analysis of Nonlinear Dynamical Systems: Chaos, Fractals, Self-Similarity and the Limits of Prediction*,System **2016**, 4, 37.

[3] Actually a positive K-entropy implies more properties such as mixing and ergodic, but the conditions and the statements need to be made more precise.