



The Geometry of Chance:  
On the theory of non-equilibrium  
statistical mechanics

Schuyler Nicholson

Submitted for the degree of  
Doctor of Philosophy  
School of Mathematics and Statistics

March 2015

Supervisors: Dr. Eun-jin Kim

University of Sheffield

## ACKNOWLEDGEMENTS:

WHEN STARTING A THESIS ONE IS OFTEN LEAD TO BELIEVE THAT THEY ARE TAKING ON THE ENDEAVOUR. THE TRUTH IS I PERSONALLY HAVE NOT UNDERTAKEN THIS THESIS. I AM MERELY THE INSTRUMENT WHICH HAS RESULTED FROM SO MUCH SUPPORT FROM MY FAMILY (ON BOTH SIDES OF THE OCEAN) AND FRIENDS THROUGHOUT MY LIFE.

The most recent influence would have to be my advisor Eun-jin. The chance she took on me and the guidance she has provided was essential to my present direction and progress. Eun-jin gave me the nudge I needed to look at the thermodynamic length and many other interesting areas of current research that I might have missed otherwise, all the while giving me the independence to meander down often obscure tangents which have, if nothing else, enriched my understanding of physics and its interconnectedness.

My previous advisers Dr. David Wick and Dr. Larry Schulman have also had a large impact on my academic career. Dr. Wick who persisted with me, even when I was perhaps slower than others in the beginning to comprehend some concepts. Dr. Schulman who directed me into statistical mechanics in the first place and who taught me through his quote, "A wise Polish physicist once said, everything is interesting once you get into it" (I always suspected that he was the physicist). Which is true, there is no reason to pigeon hole oneself into a specialization, all of physics is interesting and people are doing great work in every field.

My fiancée Beth who has stuck with me and believed in me without fail. I am sorry for the long nights and missed weekends. Thank you for your patience, support and not to mention proof reading of this entire thesis, along with every publication.

I would be amiss if I didn't thank Nabil Freij, Stuart Mumford, Stephen Chaffin, Chris Nelson and Freddie Mather for enriching my time here in England and for the many hours of fruitful and helpful discussions.

Finally I want to thank my father for inspiring my love of science and my curiosity and my mother for giving me the ability to see the beauty all around me and the belief that I could take on this challenge. My only regret is that I did not finish this work in time for her to see it. She would have thoroughly enjoyed it.

**For Penelope Alice Hoblin.**

## ABSTRACT

The study of Statistical Mechanics goes back to the 1800s and the work of Boltzmann. Since that time the field has been divided into equilibrium theory and non-equilibrium theory, with the former's progression far outpacing the latter. That is until relatively recently. New insights such as the thermodynamic length [1], fluctuation theorems [2, 3] and spectral methods such as the Observable Representation [4, 5] have given us new tools to deal with large and complex non-equilibrium systems. In this work we will look at two specific tools in depth. The Observable Representation (OR) and its irreversible extension the Non-Detailed balance Observable Representation (NOR) and the information length. The NOR allows one to take the complex and often messy calculations of a system's evolution operator and represent it with a much simpler geometric version. In this version distances correspond to relationships in the original system. We will show how these distances can be used to elucidate the underlying structure of a given system and even to control chaotic systems by forming periodic orbits from said distances.

The second method to be analysed in detail is the thermodynamic length and its non-equilibrium extension the information length. This gives us a measure of distance between probability distributions the system takes in its evolution. Each distribution is represented as a point in statistical space and as the system evolves each point generates a path we can measure the distances of. This abstract space then allows us to often calculate fundamental quantities of systems under study such as the maximum available work or the dissipation as the system evolves.

Both methods may seem abstract and un-necessarily far removed from the actual systems they represent. What we gain from this abstractness far outweighs its added mathematical machinery, for from abstraction we gain generality. These methods allow us to analyse huge classes of system under one umbrella, such as irreversible or chaotic systems which before were out of reach of equilibrium statistical mechanics.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.2	Equilibrium Statistical Mechanics . . . . .	3
1.2.1	Equiprobability Assumption . . . . .	4
1.2.2	Additional constraints . . . . .	7
1.3	Dynamical systems . . . . .	10
1.3.1	Chaos . . . . .	12
1.3.2	Logistic map . . . . .	15
1.4	Information and disorder . . . . .	21
1.4.1	Relative Entropy . . . . .	24
1.4.2	Fisher Information . . . . .	25
<b>2</b>	<b>Geometry of Perron-Frobenius Operators</b>	<b>27</b>
2.1	Introduction . . . . .	27
2.2	Perron-Frobenius operators . . . . .	27
2.3	Observable Representation . . . . .	33
2.4	B matrix and Irreversible systems . . . . .	41
2.4.1	Chaos control in the logistic map . . . . .	46
<b>3</b>	<b>Statistical Mechanics Through Distances</b>	<b>73</b>
3.1	Introduction . . . . .	73
3.2	Thermodynamic length . . . . .	74
3.3	Information length . . . . .	77
3.3.1	Fisher information and intensive variables . . . . .	82
3.3.2	$\mathcal{L}$ for discrete systems . . . . .	83
3.3.3	Work for discrete non-equilibrium systems . . . . .	85
3.4	The Logistic map and $\mathcal{L}$ . . . . .	86
3.4.1	Unstable fixed points and distance to reach $p_0$ . . . . .	90

3.4.2	Conditions for $\Delta\mathcal{L} = 0$ . . . . .	91
3.5	Music through $\mathcal{L}$ . . . . .	94
3.6	Information variation ( $\mathcal{L}$ and $\mathcal{J}$ ) . . . . .	96
3.6.1	Music as a non-equilibrium system . . . . .	98
3.6.2	Power-law scalings . . . . .	101
3.6.3	Oscillations . . . . .	107
3.6.4	Continuous signals . . . . .	109
<b>4</b>	<b>Conclusion</b>	<b>117</b>

# Chapter 1

## Introduction

### 1.1 Introduction

The world is filled with large complicated systems that have many degrees of freedom. For example take one centimetre cubed of gas at  $0^\circ$  c on the surface of the earth. This system has on average  $10^{19}$  molecules of gas [6], or particles as we shall refer to them. The simple rules of classical mechanics that govern these particles has been known since the time of Newton <sup>1</sup>. To evolve this system forward we simply evolve the initial conditions forwards using a simple set of rules (Newton's laws), yet using this steady repetition of rules one is already lost and unable to calculate how the system will evolve. The system is simply too large, too complex and the precision one needs is not attainable. Still all around us we see a staggering amount of repeatable order and structure that comes from this or any system. Meaning there should be an underlying rule which governs the formation of this structure. Examples range all length scales, at the enormous scales of galaxies there is the formation of spiral galaxies such as ours [7]. The orbit of our sun and the eleven year sunspot cycle [8] are also examples of the formation of structure and order from a set of simple rules.

---

<sup>1</sup>Here we are assuming that for the length scales being considered there are no Quantum mechanical effects. This is of course in general an incorrect but necessary assumption.

As we go down in scale to those of the earth we see the repetition of weather cycles such as El Nino [9] or trade winds as repeating large scale evolutions of large complex systems. Human beings themselves are a prime example steady and repeatable formation of structure which forms from an apparent random interaction of constituents. Rosenfeld et.al in [10] showed that the creation of proteins in individual cells is not steady as was previously thought, and instead undergoes large fluctuations. This means that the basic operation of our bodies, making proteins, has a strong random element to it. The question is then, how does this set of random processes turn into the order and structure that is the human body?

Equilibrium Statistical Mechanics aims to analyse the small scale behaviour of say, atoms or molecules of a system in thermal equilibrium and extract the large scale or macroscopic consequences. The statistical flavour of this approach would make one naively assume that it is a good starting point to understand non-equilibrium systems. Despite efforts going back to the 1800's there has yet to be a unified theory of non-equilibrium statistical mechanics. To understand how we might start to construct such a theory we shall start by looking at systems in equilibrium. This will begin the first of three main sections in this thesis. Major concepts that will be needed later will be reviewed. These include concepts from equilibrium statistical mechanics we will need such extensive and intensive variables, relative entropy and Fisher information, along with a brief introduction to dynamical system and chaos theory. Next we will introduce two approaches to dealing with non-equilibrium systems. Chapter two will present the Observable Representation (OR) which is a geometric interpretation of the operator that describes the evolution of certain systems. The OR can be thought of as creating a picture which allows one to interpret the information about a system encapsulated in its evolution operator. Chapter three will deal with the thermodynamic length and its non-equilibrium extension the Information length. This is a generalization of equilibrium statistical mechanics in terms of distances over a surface of states. As a system evolves it traces out a path over

this surface, the distance of this path tells us valuable information about the available work and dissipation of the system as it evolves. The key application of this work will be into analysing classical music using the information length.

Before we go any farther though, we need to define several concepts that will be used throughout this work,

- **system:** Any large grouping of smaller components or constituents one wishes to study. This could be the stock market, a musical score, or as we will see later on, more abstract objects such as dynamical systems which will be used as mathematical experiments to test our ideas.
- **state:** A state is a set of variables that together give a complete description of the system. These are traditionally the position  $q$  and momentum  $p$  of all particles that make up the system.
- **phase space/state space:** The phase space is an  $N$  dimensional surface in which every point is a state of the system. As the system evolves in time it will trace out a path over the phase space.

## 1.2 Equilibrium Statistical Mechanics

The aim of statistical mechanics is to derive all the equilibrium properties of a macroscopic molecular system from the laws of molecular dynamics” [11]. In other words we derive the macroscopic values we see in our everyday world from the microscopic dynamics that make up the system. For a classical system in equilibrium the probability of finding the system in a certain state  $x$  is simply given by the Boltzmann distribution,

$$p(x) = \frac{e^{-\lambda^r X_r(x)}}{\mathcal{Z}}. \quad (1.1)$$

The sum extends over all observables of the system ( $X_r(x)$ , such as pressure, energy etc), where  $\lambda^r$  are a set of Lagrange multipliers and we will use

throughout this work that repeated upper and lower indices are assumed to be summed over.  $\mathcal{Z}$  is the *zustandsumme* or sum over states as named by Max Planck [12]. Today it is more commonly referred to as the partition function,  $\mathcal{Z} = \sum_x \exp[-\lambda^r X_r(x)]$ . As we shall show later if we simply have a system whose only observable is energy such as a heat bath,  $E(x) = X_1(x)$  and Eq. (1.1) becomes,

$$p(x) = \frac{e^{-\beta E(x)}}{\mathcal{Z}}. \quad (1.2)$$

$\beta$  in the above equations is the inverse temperature of the system in contact with a reservoir,  $\beta = 1/T$ . Having found Eq. (1.1), as Richard Feynman puts it [13], “That is the end of classical statistical mechanics”, for now we can calculate all the quantities we desire about the system. As we shall see though, everything in this elegant theory depends on the system being in equilibrium, meaning “that a system has settled down to the point where its macroscopic properties are constant in time” [12].

### 1.2.1 Equiprobability Assumption

Imagine a set of  $N$  particles, each of which can assume specific values of energy pressure etc,  $E_i, P_i, i = 1, 2, \dots, \sqrt{M}$ . Together the collective value of ever particle gives a state or configuration of the system. Each state as mentioned above is a complete description of the system and is represented as one point in phase space. A macrostates of the system, is an average over the probability of being in each state of the system. For example if we only consider energy the macrostate for a discrete set of states is given by,

$$\langle E \rangle = \sum_i p(i) E_i. \quad (1.3)$$

There are an infinite number of points on any given phase space. To help deal with this, above we have coarse grained each variable of interest into a finite number of values,  $\sqrt{M}$ . Every value a particle can take,  $E_i, P_j$  will be

given by a coarse grained microstate  $G_k$ ,  $k = 1, 2, \dots, M$ . The union of all  $G_k$  covers the phase space  $\Omega$ . The question then is how do we decide which set of microstates will be occupied?

Boltzmann's genius is that he proposed every configuration has an equal probability of happening. Given a large number of identical systems, the most probable set of states will simply be the set of states which has the most equivalent configurations. Or to give it a more precise definition, to calculate the configuration which has the largest multiplicity. As an example, take the set of microstates  $G_k$ ,  $k = 1, 2, \dots, M$ , each microstate corresponds to a different set of values,  $E_i$  and  $P_i$ ,  $i = 1, 2, \dots, \sqrt{M}$ . We should note, the number of coarse grains is typically very large, meaning we can to a good approximation (as is often done in the literature) take integrals over quantities instead of summations. The number of particles that occupy  $G_k$  are called the occupation number,  $n_k$ ,  $k = 1, 2, 3, \dots, M$ . As an example consider a system with just two particles as shown in Fig. (1.1). First consider the case where both particles have the same value of  $E_i$ ,  $P_j$  as shown in Fig. (1.1) (a). Since there is only one way to have this configuration the multiplicity is one. If instead we have one particle with values  $E_i$ ,  $P_j$  and the second particle with a different set of values, then there are two configurations for this situation, since switching the particles with each other gives the same configuration. Thus this situation has a multiplicity of two as seen in (1.1) (b). In general given  $N$  particles and the variables of interest having  $M$  possible microstates, the total number of arrangements is given by [14, 15],

$$C = \frac{N!}{\prod_{k=1}^M n_k!} \quad (1.4)$$

According to the equal probability hypothesis, we then want to find the set of  $n_k$  that maximizes  $C$  subject to the constraint that the probability that a particle is in  $G_k$  is,  $p(k) = n_k/N$  where  $p(k)$  is conserved,  $\sum_k p(k) = 1$ . To calculate the distribution we are most likely to observe along with the corresponding occupation numbers, we will first turn Eq. (1.4) into a more manageable form.

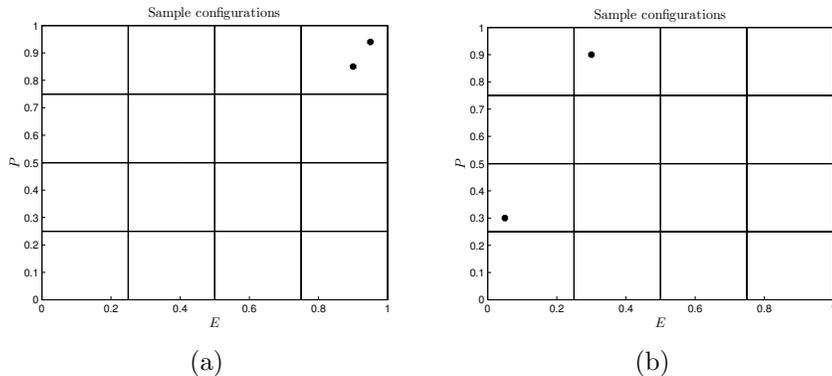


Figure 1.1: Each figure is a two dimensional coarse grained configuration space. In (a) both particles have the same values of  $E$  and  $P$  giving a multiplicity of one. In (b) both particles have two different sets of values, giving a multiplicity of two.

Note, see [16] for a nice introduction to this material. First apply Stirling's approximation since we are assuming  $N$  is large,

$$C = \frac{N^N e^{1-N}}{\prod_{k=1}^M n_k^{n_k} e^{1-n_k}}, \quad (1.5)$$

$$C = \frac{N^N}{\prod_{k=1}^M n_k^{n_k}}. \quad (1.6)$$

The second line comes from conservation of probability. Since the logarithm is a monotonic function, it shares a maximum with its argument, meaning we can take the logarithm without affecting the maximization of  $n_k$ 's,

$$\log(C) = N \log(N) - \sum_k n_k \log(n_k). \quad (1.7)$$

Finally using our definition of  $p(k)$  we have,

$$\log(C) = -N \sum_{k=1}^M p(k) \log(p(k)). \quad (1.8)$$

Thus the set of occupation numbers that are most probable is the set that corresponds to maximizing the Shannon entropy,

$$S = - \sum_{k=1}^M p(k) \log(p(k)). \quad (1.9)$$

### 1.2.2 Additional constraints

In practice a given system will be more complicated in that we will have more than one constraint on our system. Equilibrium can be defined by having constant macroscopic values, these could be energy, temperature, pressure, particle density etc. When we say “constant”, as mentioned above we mean that the system has a constant average with respect to the variables over its phase space. For a continuous system this is the integral over all phase space,

$$\int_{\Omega} p(x) X_i(x) dx = \langle X_i \rangle. \quad (1.10)$$

The variables are divided up into two groups, intensive variables and extensive variables [12],

**Intensive variables:** Quantities that do not change (in equilibrium) as the system size is scaled. These include temperature or particle density.

**Extensive variables:** Quantities that change (in equilibrium) as the system size is scaled. Examples include, internal energy, volume, or entropy.

So given a set of extensive variables  $X_i$ ,  $i = 1, 2, \dots, n$  which are on average constant  $\int_{\Omega} p(x) X_i(x) dx = \langle X_i \rangle$  we have  $n$  additional constraints. Jaynes showed [17] that by maximizing the Lagrangian we can recover the Boltzmann distribution. Using the Lagrange multipliers  $\lambda^i$ ,

$$L = - \int_{\Omega} p(x) \log(p(x)) dx - \lambda^0 \left( \int_{\Omega} p(x) dx - 1 \right) - \lambda^i \left( \int_{\Omega} p(x) X_i dx - \langle X_i \rangle \right) \quad (1.11)$$

In the above equation we have dropped the  $x$  dependence on our extensive variables  $X(x)_i = X_i$  to ease notational clutter, though it is still assumed. Also as previously noted summation over upper and lower indices is assumed. As we shall see the Lagrange multipliers which were introduced so that this problem was not under determined, are nothing more than the system's intensive variables. As an example of how to find the most likely distribution, consider the case where the system is closed and the only additional constraint is the average energy  $\langle E \rangle$ . The Lagrangian we seek to maximize writing  $p(E) = p$  is,

$$L = - \int_{\Omega} p \log(p) dE - \lambda^0 \left( \int_{\Omega} p dE - 1 \right) - \lambda^1 \left( \int p E dE - \langle E \rangle \right) \quad (1.12)$$

Taking the variational derivative [18],

$$\frac{\delta L}{\delta p} = -1 - \log(p) - \lambda^0 - \lambda^1 E = 0. \quad (1.13)$$

We see that we can immediately solve this for  $p$  giving,

$$p(E) = e^{-(1+\lambda^0+\lambda^1 E)}. \quad (1.14)$$

Though Eq. (1.14) looks similar to the Boltzmann distribution, we still need to solve for our Lagrange multipliers. We first use the conservation of probability constraint and integrate over all energies,

$$\int_0^{\infty} p dE = 1 = \int_0^{\infty} e^{-(1+\lambda^0+\lambda^1 E)} dE, \quad (1.15)$$

$$= e^{-(1+\lambda^0)} \int_0^{\infty} e^{-\lambda^1 E} dE. \quad (1.16)$$

The second integral has the solution of the form,

$$1 = e^{-(1+\lambda^0)} \frac{1}{\lambda^1}$$

$$\lambda^1 = e^{-(1+\lambda^0)} \quad (1.17)$$

Next we use the second constraint,

$$\langle E \rangle = \int_0^\infty E e^{-(\lambda^0+1+\lambda^1 E)} dE = e^{-(1+\lambda^0)} \int_0^\infty E e^{-\lambda^1 E} dE. \quad (1.18)$$

Taking the derivative under the integral leads to,

$$\langle E \rangle = e^{-(1+\lambda^0)} \left( -\frac{\partial}{\partial \lambda^1} \right) \int_0^\infty e^{-\lambda^1 E} dE = \frac{1}{\lambda^1}. \quad (1.19)$$

Using Eq. (1.16) and Eq. (1.19) in Eq. (1.14) we finally arrive at,

$$p = \frac{1}{\langle E \rangle} e^{-E/\langle E \rangle} = \frac{e^{-\beta E}}{\mathcal{Z}}. \quad (1.20)$$

The final term on the right is simply using the more traditional notation for the inverse temperature and partition function. One should keep in mind, that out of equilibrium the allowed configurations are not in general known, meaning we cannot use the equiprobability assumption as we did in this section. The proof of this difference in allowed configurations is the plethora of non-Boltzmann-Gibbs distributions found in nature out of equilibrium. Under mild requirements such as irreducibility, there is a unique equilibrium distribution, therefore, if we do not occupy this distribution we are not in equilibrium and we must have a different set of occupation numbers. Furthermore it has been argued that simply maximizing Eq. (1.9) out of equilibrium results in a backwards arrow of time, which is of course completely at odds with our everyday experience and the second law [19]. The second law states that for a closed system, entropy must on average increase until it reaches  $p_0$ . Deciding what the correct entropy to use and what the correct constraints are so as to find the non-equilibrium distribution of a given system is the goal of non-equilibrium statistical mechanics.

### 1.3 Dynamical systems

A dynamical system may be defined as a deterministic mathematical prescription for evolving the state of a system forward in time [20]. Dynamical systems can be either continuous or discrete depending on the variables of interest. For a discrete system whose phase space is given by  $\Omega$ , initial points  $x_n$  are evolved by a map over discrete time steps,

$$x_{n+1} = f(x_n). \quad (1.21)$$

A continuous dynamical system evolves the initial condition  $x \in \Omega$  for a time  $t$ ,

$$x(t) = f^t(x). \quad (1.22)$$

Together,  $(\Omega, f)$  form a dynamical system. In this work subscripts will denote component directions, i.e.  $x_i(t)$  is the  $i$ th component of  $x(t)$ . We usually assume  $f$  to be a smooth function meaning we can take as many derivatives of it as we need. The equations of motion that we are usually given are  $\dot{x}(t) = df^t/dt$  where  $x$  is a vector of dimension  $d$  giving a set of ordinary differential equations (ODEs). To understand how the neighbourhood of a point in  $\Omega$  behaves, we take the evolution of a second test orbit initially some small displacement  $\delta x$  away giving,

$$f^t(x + \delta x) = x(t) + \delta x(t) = f^t(x) + \frac{\partial x_i(t)}{\partial x_j} \delta x + \dots \quad (1.23)$$

To linear order we define the variational equation,

$$\delta x(t) = J^t(x) \delta x_0, \quad (1.24)$$

where we have defined the Jacobian at time  $t$  as,

$$J_{ij}^t(x) = \left. \frac{\partial x_i(t)}{\partial x_j} \right|_{x=x_0}. \quad (1.25)$$

The Jacobian measures the deformation of an infinitesimal neighbourhood around  $x$  [21]. Similarly define the variational equation of motion to linear order as,

$$\delta \dot{x}(t) = A(x)\delta x, \quad (1.26)$$

where,

$$A_{ij}(x) = \frac{\partial \dot{x}_i}{\partial x_j} \quad (1.27)$$

Then for a linear flow, taking the time derivative of Eq. (1.24) and using, Eq. (1.26) we have,

$$\begin{aligned} \delta \dot{x}(t) &= \frac{d}{dt} J(x)\delta x = A(x)\delta x(t) \\ &= A(x)J^t(x)\delta x \\ \frac{d}{dt} J^t(x) &= A(x)J^t(x), \end{aligned} \quad (1.28)$$

at time zero the Jacobian,  $J^0(x)$  is given by the identity matrix,  $J^0(x) = \mathbb{I}$ . Eq. (1.28) then gives us a way to numerically integrate the variational equation of motion. This link means that we do not need to construct  $J$  which can be very difficult in practice even with the equations of motion. This will be seen later on when we construct the Lyapunov exponents for the Lorenz system.

Of note are some special kinds of points in phase space. A fixed point of period  $p$  is a special solution such that,  $x^* = f^p(x^*)$ , where  $p$  is the minimum time such that, the orbit first returns to  $x^*$ . The set of points along the evolution,  $f^{t_0}(x^*), \dots, f^t(x^*), \dots, f^p(x^*)$ ,  $t_0 < t < p$  form what is called a periodic orbit. Fixed points can be classified as either stable or unstable. A stable fixed point is one such that any point a small distance away will evolve to the fixed point. An

unstable fixed point is the opposite, any nearby points are evolved away from the fixed point. The stability is determined by the stability coefficient. Again take an initial point  $x$  together with a second points a small distance away  $x' = x + \delta x$ . In one dimension the nearby point is evolved by,  $f(x + \delta x)$ . Writing as a Taylor expansion to first order we see that,  $|f(x + \delta x) - f(x)| = |\delta x| \left| \frac{df(x)}{dx} \right|$ .  $|df(x)/dx|$  is called the stability coefficient with good reason. It is clear that if  $|df(x)/dx| > 1$  the points move apart from each other and we have an unstable point  $x$ . If  $|df(x)/dx| = 0$  We have a marginal point and if  $|df(x)/dx| < 1$ ,  $x$  is a stable point. In higher dimensions our derivative turns into the Jacobian Eq. (1.25), whose eigenvalues give the stability of the neighbourhood around  $x$  [22]. Given the eigenvalue equation for  $J$ ,

$$J e_j = \eta_j e_j, \tag{1.29}$$

when  $|\eta_j| < 1 \forall j$  the neighbourhood around  $x$  is attracting. On the other hand if  $|\eta_j| > 1$  for any  $j$  then the neighbourhood is expanding and unstable. If  $\eta_j$  has a non-zero imaginary component,  $Im \eta_j \neq 0$  then the neighbourhood has an oscillatory behaviour.

In this Thesis we will not seek to discover anything new about the dynamical systems we study (though we will). Instead a dynamical system will be a substitute for running an experiment in the real world. They will give us a way to test and elucidate theoretical concepts that otherwise would remain abstractions. The two systems we will use are the logistic map and the Lorenz system.

### 1.3.1 Chaos

There are many different although similar definitions for chaos. Here we shall define it as follows:

1. The system demonstrates sensitive dependence on initial conditions: given two initially close points, the points will be iterated far apart.

2. The system is Topologically Transitive. Given any two intervals  $\mathbf{P}$ ,  $\mathbf{Q}$ , there is a positive integer  $n$ , such that,  $f^n(\mathbf{P}) \cap \mathbf{Q} \neq \emptyset$ .

The rate that two initial conditions are either iterated apart, or together is called the Lyapunov exponent ( $L$ ),  $dx \approx e^{Ln} dx_0$ . The finite, but long time Lyapunov exponent is defined as,

$$L = \frac{1}{N} \sum_i^N l_i. \quad (1.30)$$

$l_i$  is defined as the instantaneous Lyapunov exponent,

$$l_i = \log |f'(x_i)|. \quad (1.31)$$

The traditional definition of the largest Lyapunov exponent is,

$$\lambda_{max} = \lim_{t \rightarrow \infty} \frac{1}{t} \log \frac{\|\delta x(t)\|_2}{\|\delta x_0\|_2}. \quad (1.32)$$

$\|\delta x(t)\|_2$  is the  $L_2$  norm of the vector  $\delta x$ . For higher dimensional systems we generalize Eq. (1.30) from the average divergence in between two points, to the divergence of a volume of space in  $n$ -dimensions. This means that the rate of growth of a volume element is given by the determinant of the Jacobian, Eq. (1.25),

$$|\det J_{ij}^t(x)| = \exp \left( t \sum_{i=1}^n \lambda_i \right). \quad (1.33)$$

Here  $\lambda_i$  are the lyapunov exponents, [23].

There is a generalization of the Lyapunov exponents [24], which defines the Lyapunov exponent of order  $p$ ,  $p \leq n$ ,  $n$  is the dimension of the system. This method is often of use when running long simulations of dynamical systems which have large stretching in one or more directions. Given a parallelepiped whose edges are given by the vectors  $u_p$ ,  $p = 1, 2, \dots, n$ , the  $p$  Lyapunov exponent

is then defined as,

$$\lambda_p = \lim_{t \rightarrow \infty} \log (Vol (J^t(x)\delta x)), \quad (1.34)$$

“where  $Vol (J^t(x)\delta x)$  is the  $p$  dimensional volume in the tangent space” [24]. From [25] we know that  $p$  linearly independent vectors,  $u_p$  are guaranteed to exist meaning, there are a set of scalars  $a_1, a_2, \dots, a_p$  not all zero, such that [26],

$$a_1 u_1 + a_2 u_2 + \dots + a_p u_p \neq 0. \quad (1.35)$$

The volume of our parallelepiped is then,  $Vol^p = \|v_1\| \|v_2\| \dots \|v_p\|$ , where  $v_i$  are the orthogonal vectors generated using Gram-Schmidt orthogonalization [27]. The orthonormal vectors  $w_i$  which also come out of Gram-Schmidt are then evolved forward in time. This is due to the matrix whose eigenvalues generate the Lyapunov exponents can often become ill conditioned which leads to numerical errors [24]. To recap to find the Lyapunov exponents, we must simply integrate Eq. (1.28) along the flow and find the vectors  $v_p$  through Gram-Schmidt Orthogonalization [27], thus leading to our final definition for  $\lambda_i$ ,

$$\lambda_i \approx \frac{1}{T} \int_0^T dt \log \|v_i(t)\|_2, \quad (1.36)$$

with,  $\lambda_p$  being given by,

$$\lambda_p = \sum_{i=1}^p \lambda_i, \quad (1.37)$$

In conclusion, if a system is Topologically Transitive, and satisfies either  $L > 0$  or  $\lambda_i > 0$ , it is defined as being chaotic,  $L \leq 0$ , the system is not chaotic. To see why this mixing is needed, imagine a one dimensional system bound to the real numbers with two initial conditions separated by some small amount  $\epsilon$ ,  $x_0$  and  $x_0 + \epsilon$ . The rule for evolving this system is  $x_{n+1} = x_n^p$ , where  $p > 1$ ,  $p \in \mathbb{Z}$ . This system has sensitive dependence of initial conditions but it is obviously not chaotic for we can easily predict all solutions go to infinity.

### 1.3.2 Logistic map

The logistic map is a one-dimensional discrete map which is one of the simplest system that can demonstrate chaos. The system is governed by,

$$x_{n+1} = ax_n(1 - x_n), \quad (1.38)$$

where  $x_n$  is the position of the system at the  $n^{\text{th}}$  time step.  $a$  is the control parameter  $0 \leq a \leq 4$  for  $x_n \in [0, 1]$ . The generalized logistic map is defined as,

$$x_{n+1} = 1 - ax_n^2. \quad (1.39)$$

Here  $x \in [-1, 1]$  when  $a \in [0, 2]$ . If chaos is defined by sensitive dependence to initial conditions and the approach and departure of many stable and unstable fixed points, then we should be able to see this for the logistic map. Fig. (1.2) shows the overall evolution of Eq. (1.38) as we vary  $a$ . From  $0 \leq a \leq 1$  there is

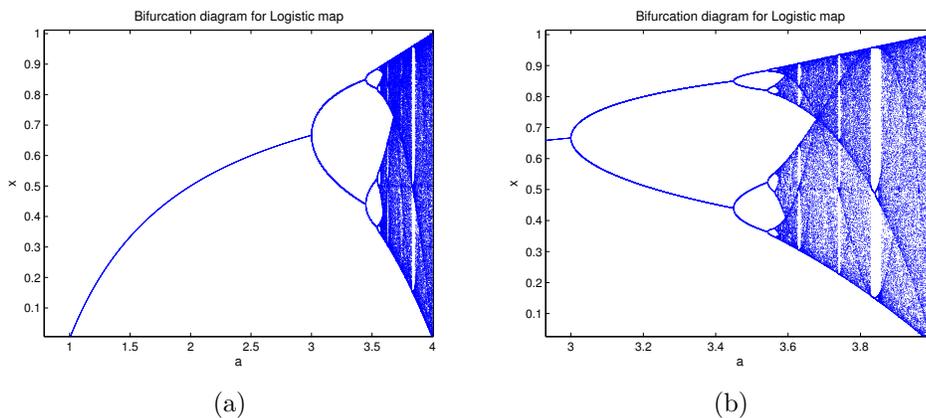


Figure 1.2: Fig. (a), Bifurcation diagram for logistic map Eq. 1.38. Fig. (b), A zoomed in view of the Bifurcation diagram. Note the sudden loss of defined bifurcations around  $a = a_c =$ . This is the initial onset of chaos. Also visible are the periodic windows, explained below.

one stable fixed point for the logistic map at  $x^* = 0$ . There is also an unstable

fixed point at  $x^* = 1 - 1/a$  for  $0 \leq a \leq 1$ . This fixed point becomes stable from  $1 < a < 3$  while  $x^* = 0$  becomes unstable for  $1 < a \leq 4$ . The transition at  $a = 3$  is the first period doubling bifurcation. Here two stable period  $n = 2$  fixed points and one unstable fixed point are created. The process that creates this is seen in fig. (1.3) where we have plotted  $f^2(x)$ ,  $f^4(x)$ ,  $f^6(x)$  and  $f^8(x)$  for  $a = 2.9$  in black over the entire domain  $[0, 1]$ . Where the line  $f(x) = x$  in red first crosses  $f^n(x)$  a fixed point of length  $n$  is created. Each stable fixed point is plotted with a red circle, while all unstable fixed points are shown with black circles. One can convince themselves from the evolution at each time step that for this value of  $a$  no other stable orbits above  $n = 2$  are created. This is how each period double bifurcation proceeds at ever shorter increments of  $a$  until at  $a_c = 3.569946 \dots$  [28]. At  $a_c$  there are an infinite number of stable and unstable fixed points and the system is chaotic.

The logistic map has another interesting feature that can be seen in Fig. (1.2), where for certain values of  $a$  the system becomes periodic. The most obvious example of this is the period-3 window between  $3.828 \lesssim a \lesssim 3.8568$  as seen in Fig. (1.4). The evolution to chaos is governed by successive period doubling bifurcations while the generation of period windows is through tangent bifurcations. Here an unstable and stable fixed point are created as the curve  $f^n(x)$  initially intersects  $f(x) = x$ . To help illustrate this, Fig. (1.5) shows a typical tangent bifurcation. Here  $f^3(x)$  is plotted for both  $a = 3.8$  and  $a = 3.9$ . As  $a$  grows the function passes through the  $f(x) = x$  line generating the unstable and stable fixed point. A view of the period-3 window as a function of  $a$  is shown in fig. (1.6) (a). To generate this figure, we calculate  $f^3(x)$  as a function of  $a$  and see how the interpolated surface intersects the  $f(x) = x$  plane. Initially where  $f^3(x)$  intersects the plane we have a tangent bifurcation. fig. (1.6) (b) shows the  $f^6(x)$  surface as a function of  $a$ . In this figure we see the initial creation of the tangent bifurcations as the  $f(x) = x$  plane first intersects  $f^6(x)$ . We also see the period doubling bifurcations around  $a = 3.825$ .

The logistic map may be the simplest dynamical system which can exhibit

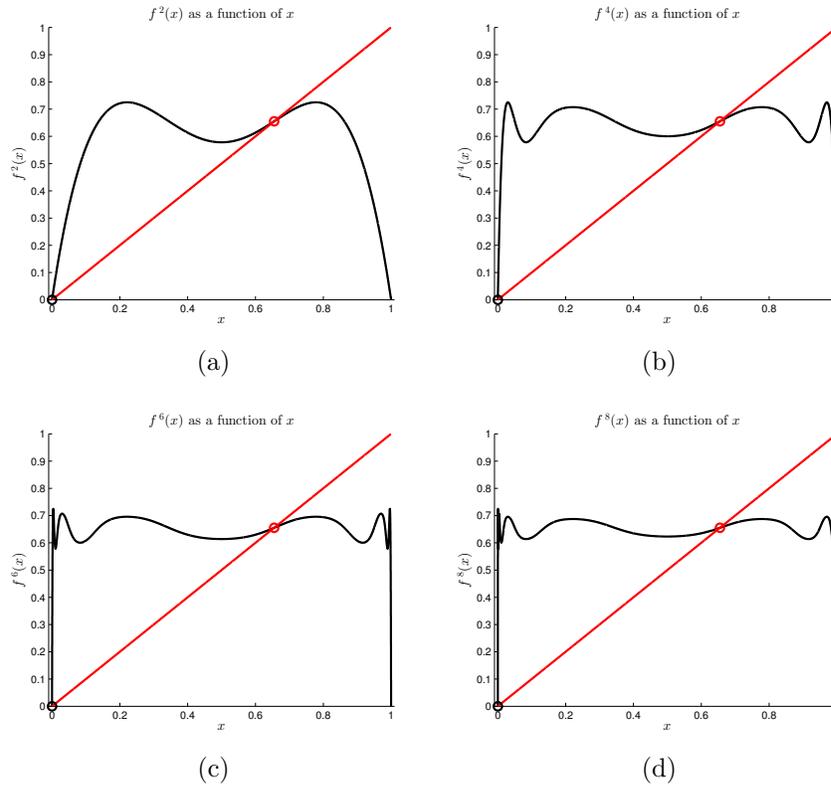


Figure 1.3:  $f^2(x)$ ,  $f^4(x)$ ,  $f^6(x)$  and  $f^8(x)$  are plotted, (a-d) in black for the logistic map when  $a = 2.9$ . We see that one stable fixed point is created for  $f^2(x)$  and for all higher powers of  $f$  no more stable fixed points are created. The one unstable fixed point is plotted at  $x(0)$  with the black circle.

chaos. But as we have just seen it has a very rich and interesting evolution to chaos. This is probably best summed up in Fig (1.7) where we have plotted the bifurcation diagram in blue with the Lyapunov exponent, Eq. (1.30) overlaid in black. We see that as the system undergoes period doubling bifurcations the Lyapunov exponent grows greater than zero. Then as the tangent bifurcations bring the system temporarily out of chaos the Lyapunov exponents fall back down below zero, signalling the system is no longer chaotic.

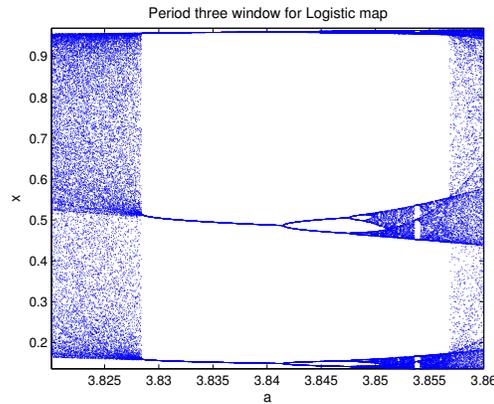


Figure 1.4: The period window for  $3.828 \lesssim a \lesssim 3.8568$ .

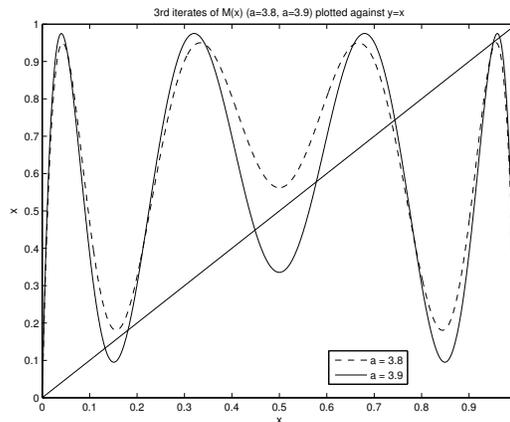


Figure 1.5:  $f^3(x)$  for  $a = 3.8$  and  $a = 3.9$  plotted with  $f(x) = x$ . The intersections that arise are the tangent bifurcations.

## Lorenz System

Edward Lorenz in 1963 [29] showed that by truncating Salzman's set of first order differential equations which described Rayleigh Benard Convection (see [30] and references within) to just three variables, he could generate non-periodic flows, [20] i.e. chaotic trajectories. The equations for this set  $x_i, i = 1, 2, 3$  have three parameters,  $\sigma, \rho$  and  $\beta$ . Together they give the approximation of Rayleigh

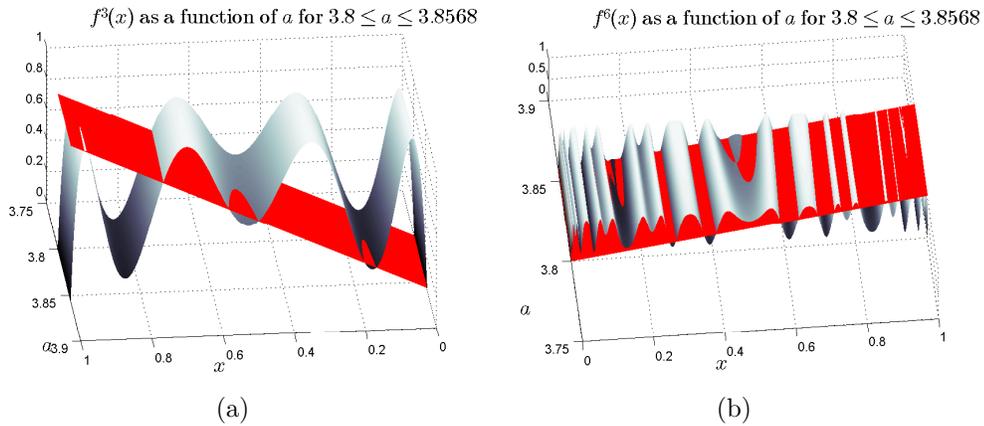


Figure 1.6: (a) shows  $f^3(x)$  as a function of  $a$ . As  $f^3(x)$  crosses the  $f(x) = x$  plane a tangent bifurcation is formed. (b) shows  $f^6(x)$  as a function of  $a$ . Here we see the initial tangent bifurcation as the plane and function intersect, but also the period doubling bifurcation as the system evolves back to chaos.

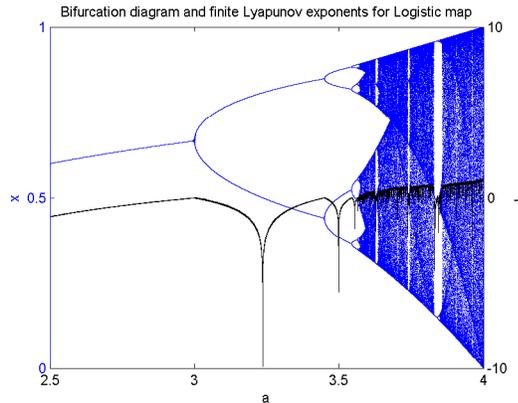


Figure 1.7: Plotting the Bifurcation diagram and  $L$  over top of each other, we see that the system changes from being chaotic to non chaotic as it undergoes both period doubling and tangent bifurcations.

Benard Convection [31],

$$\begin{aligned}
 \dot{x}_1 &= \sigma(x_2 - x_1), \\
 \dot{x}_2 &= -x_1x_3 + \rho x_1 - x_2 \\
 \dot{x}_3 &= x_1x_2 - \beta x_3.
 \end{aligned}
 \tag{1.40}$$

For  $0 \leq \rho < 1$  we have one equilibrium solution,

$$(x_1^*, x_2^*, x_3^*) = (0, 0, 0),$$

which is stable for  $\rho < 1$  and unstable for  $\rho > 1$ .  $\rho \geq 1$  we also have two more equilibrium solutions

$$(x_1^*, x_2^*, x_3^*) = \left( \pm\sqrt{\beta(\rho-1)}, \pm\sqrt{\beta(\rho-1)}, \rho-1 \right).$$

The logistic map in the previous section had a relatively simple progression to chaos since there is only one control parameter to vary. The Lorenz system on the other hand is extremely complex with the adjustment of each parameter leading to substantially different behaviour. To give a flavour of the myriad of different behaviours the fixed points can exhibit we will vary  $\rho$  from  $1 < \rho < 28$ . With  $\rho = 28$  the value which we will use later on in our analysis. The other two parameters will be held fixed at,  $\sigma = 10$  and  $\beta = 8/3$ , which with  $\rho = 28$  are the three parameter values originally used by Lorenz. For  $\rho = 1$  the fixed point at the origin goes from stable to unstable as previously mentioned. At the same time the other two fixed points come into being and are stable. This is then an example of a pitchfork bifurcation [28]. The two fixed points not at the origin have varied behaviour as  $\rho$  increases past one. For  $1 \leq \rho \lesssim 1.3457$  all the eigenvalues have zero imaginary parts as seen in Fig. (1.8), plotted with black lines and circles. This means both fixed points have linear behaviour in their respective neighbourhoods. For  $1.3457 < \rho \lesssim 24.737$  the eigenvalues  $\eta_2$  and  $\eta_3$  have non-zero imaginary parts and trajectories have oscillating behaviour around the fixed points, shown with the black lines in Fig. (1.8). Finally for  $24.737 < \rho \leq 28$ ,  $|\eta_2|$  and  $|\eta_3|$  are greater than one meaning nearby orbits are repelled. In this parameter window the system is chaotic, shown in Fig. (1.8) with red lines.

The Lorenz system exhibits a plethora of different behaviours as you vary the control parameters. We have just given a taste of these in the above discussion.

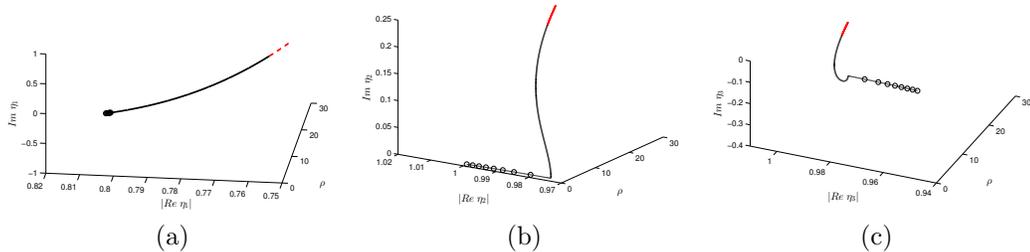


Figure 1.8: Each of the above figures depicts the real and imaginary parts of  $\eta_j$ ,  $j = 1, 2, 3$  for the Jacobian of the Lorenz system the fixed point  $\left(\sqrt{\beta(\rho-1)}, \sqrt{\beta(\rho-1)}, \rho-1\right)$  for  $1 \leq \rho \leq 28$ . We see three distinct phases for  $\eta_2$  and  $\eta_3$ , first line with circles showing the purely linear and thus real stability of the neighbourhood. Then the development of non-zero imaginary components shown with solid black lines. Finally the loss of stability as  $|\eta| > 1$  shown in red.

The values we will use in the coming work of  $\sigma = 10$ ,  $\beta = 8/3$  and  $\rho = 28$  have been selected due to their celebrated history and the fact that the largest Lyapunov exponents Eq. (1.36) is greater than zero. Using these three control parameters we generate the famous butterfly attractor Fig. (1.10).

## 1.4 Information and disorder

The Shannon entropy was shown in section 1.2 to arise from the assumption that every configuration is equally likely. There is another side to this subject, namely that the entropy is given as the missing information in the system  $S = -\mathcal{I}$ . This intuitively makes sense, if we have  $p(x, t) = 1$  while  $p(y \neq x, t) = 0$ , then we have zero missing information in that finding the system in state  $x$  gives us no new information since we knew exactly what state the system was in already.

This is the view Shannon famously proposed in 1967 [32]. One may ask are there other measures which can be used to represent Information? Khinchin [33] formulated four axioms which can be shown to uniquely produce the Shannon

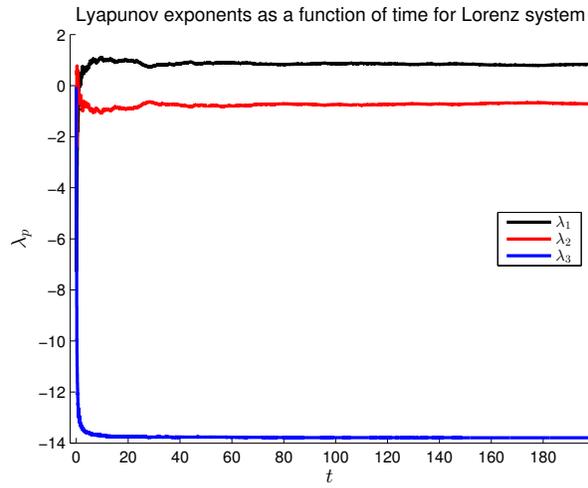


Figure 1.9: All three Lyapunov exponents for the Lorenz system using  $\sigma = 10$ ,  $\beta = 8/3$  and  $\rho = 28$ .

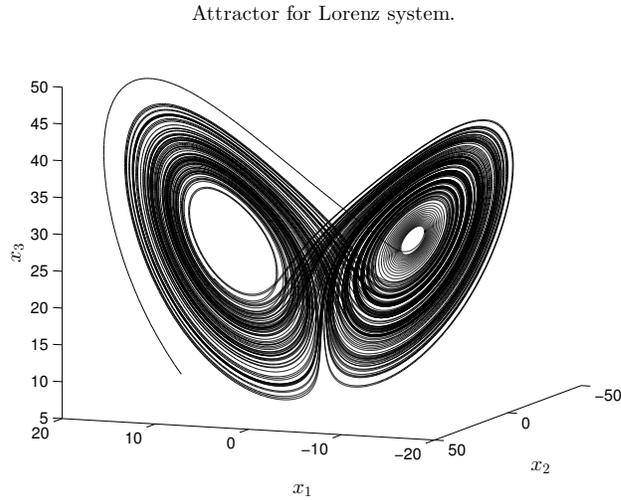


Figure 1.10: The Lorenz attractor for  $\sigma = 10$ ,  $\beta = 8/3$  and  $\rho = 28$ .

entropy [22]. The axioms are [34],

1. Information only depends on the probability of an event,

$$\mathcal{I}(p(1), p(2), \dots, p(N))$$

2. Information takes on a minimum for the uniform distribution,

$$\mathcal{I}(1/N, 1/N, \dots, 1/N) \leq \mathcal{I}(p(1), p(2), \dots, p(N)).$$

3. Including an event of probability  $p(x, t) = 0$  does not change the information.

4. Given two subsystems,  $I$  and  $II$ , the combined information is given by [35],

$$\mathcal{I}(p(x, y)^{I,II}) = \mathcal{I}(p(x)^I) + \mathcal{I}(p(y|x)^{II}). \quad (1.41)$$

$p(x, y)$  is the joint distribution of both subsystems,  $p(x)$  is the marginal distribution of subsystem  $I$  and  $p(y|x)^{II}$  is the conditional distribution of system  $II$  given system  $I$ . It follows from the joint entropy, which we can then relate to the information through  $S = -\mathcal{I}$ ,

$$\begin{aligned} S(x, y)^{I,II} &= - \sum_{x,y} p(x, y)^{I,II} \log(p(x, y)^{I,II}), \\ &= - \sum_{x,y} p(x, y)^{I,II} \log(p(x)^I) - p(x, y)^{I,II} \log(p(y|x)^{II}), \\ &= - \sum_x p(x)^I \log(p(x)^I) - \sum_{x,y} p(x, y)^{I,II} \log(p(y|x)^{II}). \end{aligned} \quad (1.42)$$

This axiom says how the order of collecting information influences the outcome. If both subsystems are independent, then the order of collection doesn't matter and we arrive at the more often quoted relation showing

the independence of information for independent systems,

$$\mathcal{I}(p(x, y)^{I,II}) = \mathcal{I}(p(x)^I) + \mathcal{I}(p(y)^I). \quad (1.43)$$

If one changes the fourth axiom then we have the generalized entropies such as the Renyi [36] and Tsallis [37] entropies. See [34, 38] and [39] for nice introductions.

### 1.4.1 Relative Entropy

A related measure of information is the Relative entropy or Kullback-Liebler distance [40],

$$DS[p(x)|q(x)] = \sum_x p(x) \log \left( \frac{p(x)}{q(x)} \right). \quad (1.44)$$

$DS$  is defined as the information in  $x$  available to discriminate between two systems  $p(x)^I$  and  $q(x)^{II}$ , per observation of system  $I$ . The quantity is non-negative and zero only when  $p(x) = q(x)$ . Though called a distance, this is technically incorrect since in general  $DS[p(x)|q(x)] \neq DS[q(x)|p(x)]$  and  $DS[p(x)|q(x)]$  does not satisfy the triangle inequality [41]. Procaccia has given another physical meaning to  $DS$  which is more applicable to our goals here, namely that for a distribution evolving to equilibrium  $p_0(x)$  the relative entropy gives the amount of available work for the system. To see this add and subtract  $S(p_0)$  to Eq. (1.44),

$$\begin{aligned} DS[p(x, t)|p_0(x)] &= \sum_x p(x) \log \left( \frac{p(x)}{p_0(x)} \right) + S(p_0(x)) - S(p_0(x)), \quad (1.45) \\ &= S(p_0(x)) - S(p(x, t)) + \sum_x \log(p_0(x)) (p_0(x) - p(x, t)). \end{aligned}$$

Using Eq. (1.1) we arrive at,

$$DS[p(x, t)|p_0(x)] = \Delta S - \lambda^i \Delta \langle X_i \rangle. \quad (1.46)$$

Procaccia identifies,  $\Delta S$  as the change in entropy and  $\Delta\langle X_i \rangle$  as the average change in the extensive variables,

$$\Delta\langle X_i \rangle = \sum_x X_i(x)(p_0(x) - p(x, t)). \quad (1.47)$$

The first law of thermodynamics says that work  $W$  is equal to the change in dissipation minus the change in energy,  $W = T\Delta S - \Delta E$ . To see how  $DS[p(x, t)|p_0]$  is a measure of work, imagine a heat bath, where the inverse temperature is given by,  $\lambda^1 = 1/T = \beta$ . Then the only extensive variable is energy, meaning Eq. (1.46) is,

$$\begin{aligned} DS[p(x, t), p_0(x)] &= \Delta S - \beta\Delta\langle E \rangle, \\ TDS[p(x, t), p_0(x)] &= T\Delta S - \Delta\langle E \rangle. \end{aligned} \quad (1.48)$$

We see that  $DS$  is the measure of available work as the system evolves from  $p(x, t)$  to  $p_0(x)$ .

### 1.4.2 Fisher Information

The Shannon entropy and the Relative entropy are both global measures of the system, in that they include averages over the entire phase space. If you then change the order that you sum over the states you will not change the value of  $S$  or  $DS$ . The next information measure is different. The Fisher information is a local measure over a distribution  $p(x, t)$ , in that changing the order of summation matters. The Fisher information [42] is a measure of information about the variance of a set of measurements. This is shown in the celebrated Cramer Rao inequality,

$$\langle (x - \langle x \rangle)^2 \rangle I_F \geq 1, \quad (1.49)$$

Here  $x = \langle x \rangle + \epsilon$ , where  $\langle x \rangle$  is the true value of a parameter of the system and  $x$  is the value of  $\langle x \rangle$  shifted by some noise term. It is assumed that the average of  $x$  is

better than any individual measurement, meaning,  $|\langle \langle x \rangle + \epsilon \rangle - \langle x \rangle| < |x - \langle x \rangle|$ . A variable whose average converges to a true value is called a *smart measurement*. Here we have used the average notation to follow from our definition of To derive the Fisher information ( $I_F$ ) and Eq. (1.49) we first take the derivative with respect to  $\langle x \rangle$  of,

$$\int dx (x - \langle x \rangle) p(x|\langle x \rangle) = 0, \quad (1.50)$$

giving,

$$\int dx (x - \langle x \rangle) \frac{\partial p}{\partial \langle x \rangle} - \int p = 0. \quad (1.51)$$

The probability distribution is the conditional distribution of  $x$  given the true value of  $\langle x \rangle$ ,  $p(x|\langle x \rangle) = p$ . Using the identity  $\frac{\partial p}{\partial \langle x \rangle} = p \frac{\partial \log(p)}{\partial \langle x \rangle}$  and squaring each side gives,

$$\int dx ((x - \langle x \rangle)^2 p) p \left( \frac{\partial \log(p)}{\partial \langle x \rangle} \right)^2 = 1. \quad (1.52)$$

Applying the Cauchy-Schwarz inequality leads to,

$$\int dx (x - \langle x \rangle)^2 p \int p \left( \frac{\partial \log(p)}{\partial \langle x \rangle} \right)^2 dx = \langle (x - \langle x \rangle)^2 \rangle I_F \geq 1. \quad (1.53)$$

The Fisher information ( $I_F$ ) is then a lower bound on the variance from the repeated measurement of  $x$ . This also represents a kind of uncertainty relationship reminiscent of the Heisenberg uncertainty relation [42], in that as  $I_F$  increases, the variance which is the average squared error in the measurement decreases. In the previous subsection we saw how the Shannon entropy is related to the relative entropy. The Fisher information is also related to both these measures as we will see later on.

# Chapter 2

## Geometry of Perron-Frobenius Operators

### 2.1 Introduction

The Observable Representation (OR) is a way of visualizing and understanding the information inherent in a matrix of transition probabilities. The matrix of transition probabilities is also called a Perron-Frobenius (PF) operator due to the original work by Perron in 1907 [26] regarding positive matrices and the extension of these results to irreducible matrices by Frobenius. All of the information inherent to the dynamics of a reversible system is encapsulated in the PF operator. Or as we will see for irreversible systems the complete information is given by the combination of the forward time evolution operator and the time reversed evolution operator. Before we look at this geometric representation in detail, we will introduce some theory which will prove useful.

### 2.2 Perron-Frobenius operators

Consider a set of random variables,  $\{X_n; n = 1, 2, \dots\}$ . These define the possible states of a system all of which belong to the state space,  $x \in \Omega$ .

The probability of  $X_n$  being in state  $x$  is given by,  $p(x) = Pr\{X_n = x\}$  with conservation of probability,  $\int_{\Omega} p(x)dx = 1$ . In general the probability that the system finds itself at  $X_{n+1}$  under the condition that it was previously in states  $x_0, x_1, \dots, x_n$  is given by,

$$Pr\{X_{n+1} = x_{n+1}|X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0\}. \quad (2.1)$$

If, on the other hand, the probability of being in state  $x_{n+1}$  only depends on being at state  $x_n$ , on the previous time step we have,

$$Pr\{X_{n+1} = x_{n+1}|X_n = x_n, \dots, X_0 = x_0\} = Pr\{X_{n+1} = x_{n+1}|X_n = x_n\}, \quad (2.2)$$

and the system is Markovian [43]. Though this definition seems limiting one should remember that every deterministic ODE can be thought of as a Markovian system given a large enough time step. While nature follows continuous equations of motion, we will be forced by computational limitations to use discrete approximations. These discrete approximations are commonly called coarse graining of the system in the literature. How one goes about coarse graining the system is an open topic which will not be covered here. Instead we will most often take the naive approach of uniformly dividing  $\Omega$  into disjoint sets  $I_i$  such that  $\bigcup_i I_i = \Omega$ . The continuous state of the system will be given by  $x \in \Omega$  though we will use  $x = x(t)$  interchangeably. The discrete approximations of the states will be given by indices  $i$  or  $j$ .

A Perron-Frobenius operator simply describes how a density,  $p(x, t)$  evolves to some new density  $p(x, t')$ ,  $t' = t + \Delta t$ . Here we will always assume our densities are probabilities. Again for some function  $x(t) = f^t(y, t_0)$  which evolves an orbit from state  $y$  at time  $t_0$  to  $x(t)$ ,  $t > t_0$ , the Perron-Frobenius operator is formally defined as [21, 22],

$$p(x, t) = \int_{\Omega} \delta(x - f^t(y, t_0))p(y, t_0)dy, \quad (2.3)$$

where  $\delta(x)$  is the normal dirac-delta function. Eq. (2.3) can be written in operator notation as,

$$p(x, t) = Lp(y, t_0). \quad (2.4)$$

A more useful form is found by using the dirac-delta function's composition identity (here in 1-D),

$$\delta(f(x)) = \sum_k \frac{\delta(x - y_k^*)}{\left| \frac{df(y_k^*)}{dy} \right|}. \quad (2.5)$$

$y_k^*$  are the roots of the equation,  $f(x)$  and the summation is over each pre-image of  $x(t)$ . Relating this to Eq. (2.3) means  $y_k^*$  solves,  $x - f^t(y, t_0) = 0$ . As an example that can be solved analytically take the generalized logistic map (1.39).

$$f(y_n) = 1 - ay_n^2,$$

this gives,  $y_n^* = \pm \sqrt{\frac{1-x_{n+1}}{a}}$  and  $df(y)/dy = 2ay_n$ . Using this in Eq. (2.3) and Eq. (2.5) leads to,

$$p(x_{n+1}) = \frac{\delta\left(x_{n+1} \mp \sqrt{\frac{1-x_{n+1}}{a}}\right)}{\left|2a\sqrt{\frac{1-x_{n+1}}{a}}\right|} p\left(\sqrt{\frac{1-x_{n+1}}{a}}\right). \quad (2.6)$$

So now the density at any point  $x_{n+1}$  can be known, given knowledge of  $f(y_n)$  and  $y_n^*$ . Of course this is very nice in the idealized world of mathematics but in practice there are in infinite set of points in  $\Omega$  so we are forced to consider subsets of these in  $\Omega$ . This fundamental change from single points along orbits to densities of points along orbits will allow us to define an approximation of the operator  $L$  which we can use experimentally.

Given our space  $\Omega$ , as mentioned above we partition it into  $N$  sub-intervals  $I_i$  where  $\bigcup_i I_i = \Omega$  and  $I_i \cap I_j = \emptyset$ . This change means that we now follow sets of points that evolve from one sub interval to another, as is shown in Fig. (2.1).

Following from, [44] we start at Eq. (2.3),

$$p(x, t) = \int_{\Omega} \delta(x - f^t(y, t_0))p(y, t_0)dy.$$

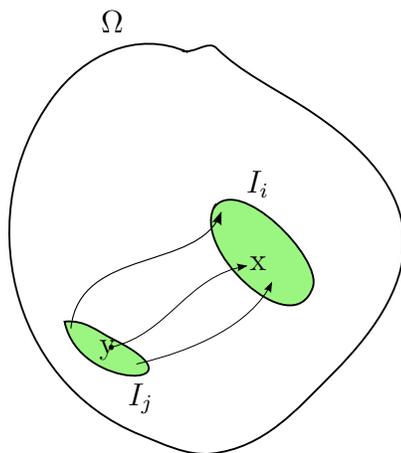


Figure 2.1: Evolutions of states in  $I_j$  to  $I_i$

Only now, we define the initial distribution to be constant over each interval  $I_j$ ,  $p = \sum_{i=1}^N p_i \chi_j$ . Here,  $\chi_j$  is the characteristic function of  $I_j$ , and allows us to keep track of which sets we are tracking the orbits to and from. Its property is that,

$$\chi_j = \begin{cases} 1 & \text{if } x \in I_j \\ 0, & \text{else.} \end{cases}$$

We also now use the pulse function in place of the delta function,

$$\delta_n(x) = \begin{cases} n & \text{if } x \in I_i, \\ 0 & \text{else,} \end{cases} \quad \int_{\Omega} \delta_n(x) dx = n \int_{\Omega} dx = 1,$$

where  $n$  is the normalization constant. Combining these definitions into, Eq.

(2.3) we have,

$$p(x, t) = \sum_{j=1}^N p_j \int_{\Omega} \delta_n(x - f^t(y, t_0)) \chi_j dy. \quad (2.7)$$

Now, since we are only considering orbits that start in  $I_j$  and are mapped to  $I_i$ , the integral only receives a contribution from said trajectory giving,

$$p(x, t) = \sum_j n \int_{f(y) \in I_i | y \in I_j} p_j \chi_j dy = \sum_{j=1}^N p_j \frac{m(f^t(y, t_0) \in I_i | y \in I_j)}{m(y \in I_j)}. \quad (2.8)$$

$m(f^t(y, t_0) \in I_i | y \in I_j)$  is read as the measure of all orbits that originate in  $I_j$  and are mapped to  $I_i$ . From Eq. (2.8) we can identify the matrix of transition probabilities that approximates the true PF operator  $L$ ,

$$R_{ij} = \frac{m(f^t(y, t_0) \in I_i | y \in I_j)}{m(y \in I_j)}. \quad (2.9)$$

This is the PF operator that we will use in our experimental results to follow.  $R \rightarrow L$  in the limit of  $I_i \rightarrow 0$ , meaning for a fine enough partition of states  $R$  will be a good approximation to  $L$ . Notice that we are again making the assumption that the system is Markovian. This is often a good approximation if the system has separation of time scales, meaning our time step is long enough that the shorter time scales of the system have averaged themselves to zero.

## Pulse functions and divergence of orbits

Now we apply the composition rule to our pulse functions. This will show how the divergence of orbits is related to the PF operator. Again let  $y^* \in I_j$  be all points in  $I_j$  such that,  $y^* = \{f(y) \in I_i \mid y \in I_j\}$  and the inverse of  $f$ ,  $f^{-1}(y^*) = y$ . Here we have used the short hand notation,  $f^t(y, t_0) = f(y)$ . We will also assume that  $f(y)$  is invertible, so that there is only one preimage of  $y^*$ .  $y^*$  from the more general case shown in the logistic map when  $f(y)$  is not invertible, meaning  $y^*$  could have more than one solution.

Using the composition rule for pulse functions where again  $x$  are all points  $x \in I_i$ ,

$$\int_{\Omega} \delta_n(x - f(y)) dy = n \int_{u(\Omega)} \frac{\delta_n(u(y^*))}{|f'(y^*)|} du = 1. \quad (2.10)$$

$u$  is defined as,  $u(y) = x - f(y)$  and  $y^*$  are again the set of all points such that  $\{u(y^*) = 0 | z \in I_j\}$ . Again taking the piecewise constant function gives,

$$\sum_j np_j \int_{\Omega} \delta_n(x - f(y)) \chi_j dy = \sum_j np_j \int_{\Omega} \frac{1}{|f'(y^*)|} \chi_j du. \quad (2.11)$$

The normalization constant  $n$  is given by,

$$\frac{1}{n} = \int_{I_j} \frac{\delta_n(u(y))}{|f'(y)|} du. \quad (2.12)$$

For higher dimensions the denominator,  $|f'(y, t_0)| \rightarrow |\det J^t(y, t_0)|$ , where  $|\det J^t(y, t_0)|$  is the determinant of the Jacobian at time  $t$  having originated from  $y$  at  $t_0 < t$ . The physical interpretation of  $|\det J^t(y, t_0)|$  is of the change in volume, occupied by the flow in evolving from  $y \rightarrow x$  which by re-writing  $R_{ij}$  using Eq. (2.12) gives,

$$R_{ij} = n \int_{I_j} \frac{\delta(u(y^*) \in I_i)}{|\det J^t(y^*)|} du. \quad (2.13)$$

Thus we can view  $R_{ij}$  as the total change in volume of  $I_j$  as it evolves to  $I_i$ . If we assume  $J$  is diagonalizable then  $\exists$  a similarity transform  $P$  such that,  $J^t(y, t_0) = PDP^{-1}$ , where  $D$  is the  $N \times N$  matrix,

$$D_{ij} = \begin{cases} \eta_i & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

$\eta_i(t) = \eta_i$  is the  $i$ th eigenvalue of  $J^t(y, t_0)$  at time  $t$ , which gives the average expansion of orbits in the direction of the  $i$ th eigenvector [22]. The determinant

of the Jacobian in Eq. (2.13) is then,

$$\det J^t(y^*) = \det(P)\det(D)\det(P^{-1}) = \prod_{i=1}^N \eta_i \quad (2.14)$$

This then gives a connection between the PF operator and the Lyapunov numbers, using the definition from (1.33) we now define our new Lyapunov exponent  $\lambda_\alpha$  as,

$$\lambda_\alpha = \lim_{t \rightarrow \infty} \frac{1}{t} \log(\eta_\alpha(z, t)). \quad (2.15)$$

Then when  $\det J^t(z)$  is well approximated by a single value such as in the limit of the volume,  $I_j \rightarrow 0$  while keeping  $\Omega$  fixed, then combining Eq. (2.14) and Eq. (2.15) and dropping the limit gives,

$$\exp\left(t \sum_{\alpha=1}^N \lambda_\alpha\right) = \det J^t(z). \quad (2.16)$$

This shows the connection between the divergence of orbits and the change of volume of space as the orbit evolves from  $z$  at  $t_0$ .

## 2.3 Observable Representation

### Phase space reconstruction

The easiest way to get an understanding of the Observable Representation (OR) [4, 5] is to introduce it through an example. This example will be the approximate reconstruction of a system's phase space, represented through a space made up of a PF operator's left eigenvectors.

The system under study will be represented by an  $N \times N$  matrix of transition probabilities  $R_{ij}$  which is defined in Eq. 2.9. The states (coarse grained states) of the system belong to the state space  $\Omega$ ,  $i, j \in \Omega$ , where  $\Omega$  is of cardinality  $\Omega < \infty$ , the system transitions through states according to Eq. (2.2). We

assume the system is irreducible which means if the system starts in any initial state  $j$ , then in  $N-1$  steps the system can get to any other state, mathematically this is given by,

$$(\mathbb{I} + R)_{ij}^{N-1} > 0 \quad \forall i, j, \quad (2.17)$$

where  $\mathbb{I}$  is the identity matrix. This requirement allows most of the results for positive matrices, i.e.  $R_{ij} > 0 \quad \forall i, j$  to carry over for non-negative matrices,  $R_{ij} \geq 0 \quad \forall i, j$  such as our PF operators. We also require the matrix be diagonalizable to avoid the added complication of requiring a Jordan form, though most results do carry over for non-diagonalizable matrices, see [45].

Through the Perron-Frobenius theorem there exists a unique positive vector  $p_0(i)$  such that  $\sum_i p_0(i) = 1$  and  $\sum_j R_{ij} p_0(j) = \lambda_0 p_0(i)$ , where  $\lambda_0$  corresponds to the spectral radius of  $R$ . Here  $R$  is normalized to unity  $\sum_i R_{ij} = 1$  which results in  $\lambda_0 = 1$ .  $p_0$  can be seen as a unique stationary distribution for  $R$ .

The next requirement which we will relax later is that the system satisfies detailed balance (DB). This is a strict condition that states the probability of transitioning from state  $j \rightarrow i$  times the probability of staying at  $j$  is the same as the back transition  $\forall i, j$ . More formally it is defined as,

$$J_{ij}^{st} = 0 = R_{ij} p_0(j) - R_{ji} p_0(i). \quad (2.18)$$

The superscript  $st$  is to differentiate the stationary current from the more general case to be defined later for any probability distribution  $p(i, t) \neq p_0(i)$ . In this work a system is in equilibrium if it is time independent and reversible. For a reversible system, “The microscopic equations of motion describe reversible processes, i.e. don’t change their form if time is reversed and if all quantities are appropriately transformed” [46]. Therefore any detailed balance system is in equilibrium. The reverse is not true though, a stationary (time independent) system is not necessarily reversible, meaning  $J_{ij} \neq 0$  implies the system is irreversible [47].

If we relax the irreducibility requirement then we are no longer guaranteed

$p_0$  and thus  $\lambda_0$  are unique. In this case given our normalization requirement each  $\lambda = 1$  will correspond to a separate stationary distribution. Going back to the irreducible case, the remaining eigenvalues are re-arranged in decreasing value,  $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ . The left and right eigenvectors of  $R$  are defined as,

$$\sum_i A_\alpha(i)^\dagger R_{ij} = \lambda_\alpha A_\alpha(j), \quad \sum_j R_{ij} P_\alpha(j) = \lambda_\alpha P_\alpha(i). \quad (2.19)$$

The subscript of vectors, (i.e.  $A$  or  $P$ ) in Eq. (2.19) denotes the columns, while the argument denotes the row.  $p_0$  is given by the zeroth right eigenvector,  $A_0(i) = 1, \forall i$  is the corresponding left eigenvector and represents conservation of probability. The eigenvectors are normalized to form an orthonormal basis,  $\langle A_\alpha | P_\beta \rangle = \delta_{\alpha\beta}$ . The OR is then defined as the set of m-tuples made up from the left eigenvectors of  $R$ ,

$$\mathcal{A} \equiv \begin{array}{cccc} \{A_1(1), & A_2(1), & \dots & A_m(1)\} \\ \vdots & \vdots & \vdots & \vdots \\ \{A_1(N), & A_2(N), & \dots & A_m(N)\} \end{array} \quad (2.20)$$

Each m-tuple will thus represent a coarse grained state of the system. Using the notation from above, the state  $I_i$  will be the set of all states,  $y$ , such that  $m(y \in I_i)$ . This is no trivial assumption, since how one defines the size of  $I_i$  can affect the outcome of any subsequent analysis. The problem is aptly stated in [48] “There is no law of Nature that defines the coarse grains”. In this work we will typically take  $I_i$  as small as possible and leave defining a more physically correct set of states to future work. Nevertheless as we will see shortly, even the most naive coarse graining can often correctly capture the underlying structure of a system. The first and simplest example comes from [48]. Imagine a Brownian particle  $\xi$  confined to a circle as illustrated in Fig. (2.2). In one unit time step, the particle starting in  $I_j$  will either transition to a new state  $I_{j+1}$  with probability 1/2 or transition to state  $I_{j-1}$  with equal

probability.

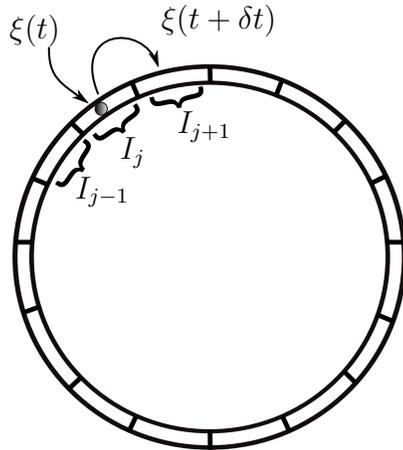


Figure 2.2: Illustration of Brownian motion on a ring. The particle, represented by the grey circle will either jump to state  $I_{j+1}$  or  $I_{j-1}$  on each time step.

To generate the OR for this example we use  $R$  of the form,

$$R = \begin{pmatrix} 0, & 1/2, & 0, & 0, & \dots, & 1/2 \\ 1/2, & 0, & 1/2, & 0, & \dots, & 0 \\ 0, & 1/2, & 0, & 1/2, & \dots, & 0 \\ \vdots & & & & \vdots & \vdots \\ 1/2, & 0, & 0, & 0, & \dots & 1/2 \end{pmatrix}. \quad (2.21)$$

Plotting the first and second left eigenvectors of Eq. (2.21) we see in Fig. (2.3) that we indeed correctly recreate the state space of the system.

The reconstruction of state space can be extended to many other examples as is shown in [49]. For the particle on a ring it is fairly obvious that two dimensions gives a correct representation of the system. But what if we have a more complicated example, or we do not know the dimensions of the underlying space? Usually one looks for a spectral gap, meaning there is a eigenvalue  $\lambda_m$ , where  $\lambda_m \gg \lambda_{m+1}$ . As an example of such a system, imagine a sets of  $n$  states that the system can readily transitions between. These states are represented

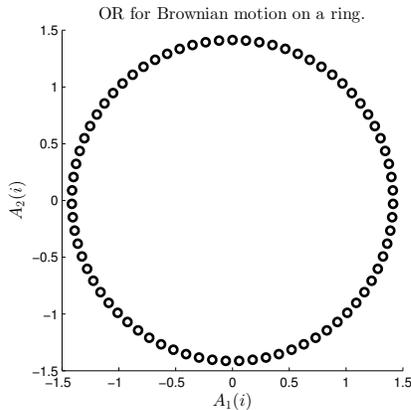


Figure 2.3: OR for Brownian motion on a ring using  $N = 75$  states.

by the random matrix  $r_1 \in \mathbb{R}^{n \times n}$ . If for example, we have three such sets of states,  $r_1$ ,  $r_2$  and  $r_3$  we can generate a system with three “clusters” as we shall refer to them. Define the matrix  $W$ ,

$$W = \begin{pmatrix} r_1 + r_1^\dagger, & \epsilon, & \epsilon \\ \epsilon, & r_2 + r_2^\dagger, & \epsilon \\ \epsilon, & \epsilon, & r_3 + r_3^\dagger \end{pmatrix} \quad (2.22)$$

The  $\dagger$  denotes transpose and  $\epsilon$  is an  $n \times n$  matrix of some small positive constant  $\epsilon \ll 1$ , which ensures  $R$  is irreducible. Adding each sub-matrix with its transpose guarantees detailed balance.  $R$  is then given by  $R_{ij} = W_{ij} / \sum_i W_{ij}$ . Using,  $n = 30$  and plotting the first ten eigenvalues of this system in Fig. (2.4) (a) we indeed see a defined spectral gap for  $m = 2$ . Plotting the first two left eigenvectors in Fig. (2.4) (b) also shows the correct grouping of states into three clusters, while Fig. (2.4) (c) shows the first three left eigenvectors. Using the extra dimension has distorted one of the clusters to the size of the space, which may lead one to incorrectly believe these states do not belong to the same cluster. Later, when dealing with chaotic systems such as the logistic map, or the Lorenz system, we will not have a clearly defined spectral gap and our job will be much harder.

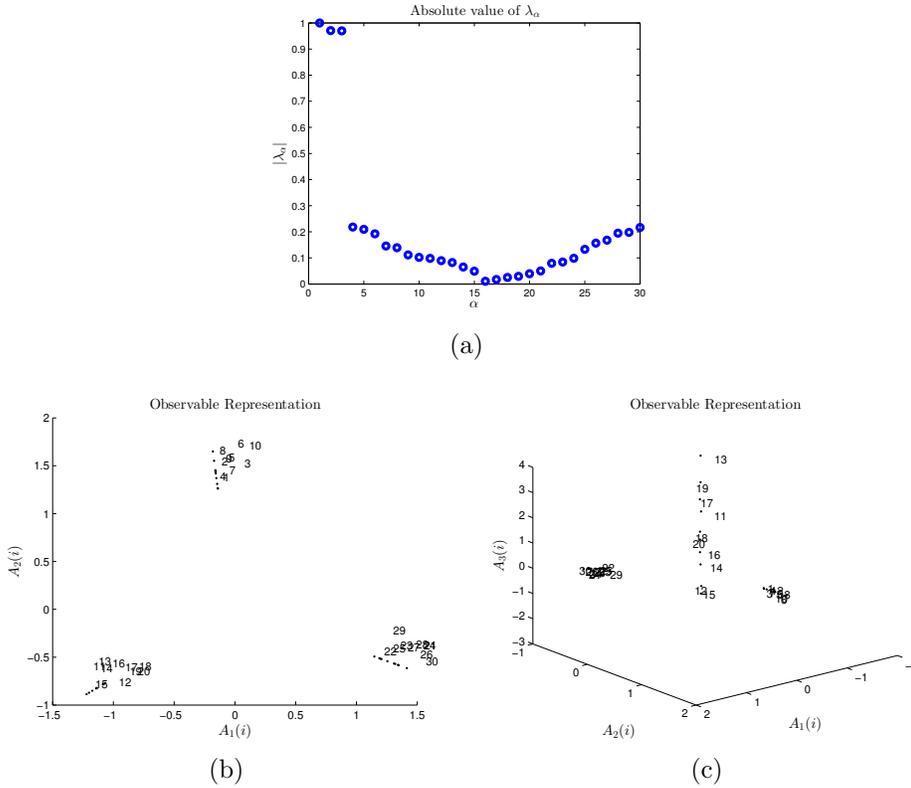


Figure 2.4: (a) shows the eigenvalues for our three cluster system.  $\lambda_2 \gg \lambda_3$  meaning we have a spectral gap. This separation means that in (b) when plotting the first two left eigenvectors we have the correct separation of the system into three clusters. Including a third left eigenvector distorts this picture possibly confusing the distribution of states into each cluster.

## Distances in OR

One of the most interesting aspects of the OR is that we can directly relate distances in the OR to distances in  $R$ . This allows us to understand the often complex relationships inherent in our systems with a simple Euclidean distance. There are no unique choices of which metric to choose, see [48, 49, 50] for some examples. In this work we have focused on the metric from [49]. For completeness we will show how one derives the metric which will also illustrate why so far we have required detailed balance.

The crux of everything to follow is the symmetric matrix  $S$ . For a detailed balance system,  $S$  is defined as,

$$S_{ij} = \frac{1}{\sqrt{p_0(i)}} R_{ij} \sqrt{p_0(j)}. \quad (2.23)$$

Plugging Eq. (2.23) into Eq. (2.18)

$$R_{ij} p_0(j) = R_{ji} p_0(i).$$

and using detailed balance, we arrive at,  $S_{ij} = S_{ji}$ . Being symmetric,  $S$  allows us to make a connection to the eigenvectors of  $R$ ,

$$\begin{aligned} \sum_j R_{ij} P_\alpha(j) &= \lambda_\alpha P_\alpha(i), \\ \sum_j S_{ij} \frac{P_\alpha(j)}{\sqrt{p_0(j)}} &= \lambda_\alpha \frac{P_\alpha(i)}{\sqrt{p_0(i)}}. \end{aligned} \quad (2.24)$$

Defining  $\psi_\alpha$  as the eigenvector of  $S$  we see that,  $\psi_\alpha(x) = \frac{P_\alpha(x)}{\sqrt{p_0(x)}}$ . Likewise it is easily shown that,  $\psi_\alpha(x) = \sqrt{p_0(x)} A_\alpha(x)$ . Since  $S \in \mathbb{R}^{N \times N}$  is guaranteed to have  $N$  linearly orthogonal eigenvectors [26], then through the above argument  $R$  is also guaranteed to have  $N$  linearly orthogonal eigenvectors.

To show how  $R$  relates to  $\mathcal{A}$  (2.20), first use the spectral expansion of  $R$ ,

$$R_{ij} = \sum_\alpha \lambda_\alpha P_\alpha(i) A_\alpha(j) \quad (2.25)$$

We are interested in measuring the difference between any two distributions,  $R_{ui}$  and  $R_{uj}$ . With this in mind we divide their difference by  $\sqrt{p_0}$  and sum over

$u$ ,

$$\begin{aligned} \left| \frac{R_{ui} - R_{uj}}{\sqrt{p_0(u)}} \right| &= \left| \sum_{\alpha} \lambda_{\alpha} \psi_{\alpha}(u) (A_{\alpha}(i) - A_{\alpha}(j)) \right|, \\ \sum_u \left| \frac{R_{ui} - R_{uj}}{\sqrt{p_0(u)}} \right|^2 &= \sum_u \left| \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \Delta A_{\alpha} \Delta A_{\beta} \psi_{\alpha}(u) \psi_{\beta}(u) \right|. \end{aligned} \quad (2.26)$$

From Eq. (2.26) we use the orthogonality of  $\psi$ 's to get the desired result,

$$\sqrt{\sum_u \left| \frac{R_{ui} - R_{uj}}{\sqrt{p_0(u)}} \right|^2} = \sqrt{\sum_{\alpha} |\lambda_{\alpha}|^2 |A_{\alpha}(i) - A_{\alpha}(j)|^2}. \quad (2.27)$$

In [49], they only take the  $L_2$  norm of the right side, giving the  $L_1 \geq L_2$  inequality. Since the eigenvalues are monotonic we can divide through by the  $m^{\text{th}}$  eigenvalue giving the original version of the result,

$$\frac{1}{|\lambda_m|} \sum_u \left| \frac{R_{ui} - R_{uj}}{\sqrt{p_0(u)}} \right| \geq \sqrt{\sum_{\alpha=1}^m |A_{\alpha}(i) - A_{\alpha}(j)|^2}. \quad (2.28)$$

If one wishes to use Eq. (2.28) instead of (2.27) then we can improve upon the  $L_1, L_2$  inequality by bounding the  $L_1$  norm of  $R$ ,  $\|R\|_1$  from above. We start with,

$$\sum_u \left| \frac{R_{ui} - R_{uj}}{\sqrt{p_0(u)}} \right|, \quad (2.29)$$

taking the Cauchy Schwarz inequality using  $1^2$  which gives,

$$\sum_i 1^2 \sum_u \left| \frac{R_{ui} - R_{uj}}{\sqrt{p_0(u)}} \right|^2 \geq \left( \sum_u \left| \frac{R_{ui} - R_{uj}}{\sqrt{p_0(u)}} \right| \right)^2. \quad (2.30)$$

Labelling the right hand side of Eq. (2.27) as  $D_{OR}$  and Eq. (2.29) as  $D_1$  we

have,

$$\sqrt{N}D_{OR} \geq D_1 \geq D_{OR}. \quad (2.31)$$

With the beauty of hindsight we might have guessed in the beginning that we have merely re-derived the famous  $L1$ ,  $L2$  inequality. Given a vectors  $x \in \mathbb{R}^N$ , the  $L1$  norm,  $\|\cdot\|_1$  and the  $L2$  norm  $\|\cdot\|_2$  follow,  $\sqrt{N}\|x\|_2 \geq \|x\|_1 \geq \|x\|_2$ , meaning we could have written from the start that the  $L2$  norm of  $\mathcal{A}$  provides an upper and lower bound to relationships in  $R$ . A question for future work is how does the function  $\sqrt{N}D_{OR}$  change as  $\sqrt{N}$  grows for a system of fixed volume, i.e. will  $D_{OR}$  go to zero for neighbouring states faster then  $\sqrt{N}$ ?

## 2.4 B matrix and Irreversible systems

All of the previous section hinged on  $J_{ij}^{st} = 0$  i.e. the system being in equilibrium. Since in general non-equilibrium systems are irreversible, ideally we would like to be able to extend the distance relationships in the OR to account for these systems, in essence forming a non-detailed balance Observable Representation (NOR). In this section we will first deal with systems in or near a stationary state  $p_0$ . Though as we will see later on, this assumption will still give us enough information to form periodic orbits in fully chaotic non-equilibrium systems. Later we will relax the condition of being near  $p_0$  and show that we can define both the operator and its accompanying NOR for any arbitrary PDF  $p_t(i) \neq p_0(i)$ .

To extend the previous results for  $R$ , a new matrix was defined in [51, 52],

$$B_{ij}^{st} = R_{ij} - \frac{J_{ij}^{st}}{2\sqrt{p_0(j)}}. \quad (2.32)$$

The superscript  $st$  is to differentiate it from the generalized matrix defined in a later section.  $B^{st}$  has some very useful properties:

1.  $\sum_i B_{ij}^{st} = 1$ .

2. If  $R$  is irreducible,  $B^{st}$  is irreducible.

3.  $B^{st}$  and  $R$  share the same stationary distribution.

(1) follows from Kirchoff's loop rule that the sum of current into a state is equal to the sum of current out of a state, meaning  $\sum_i J_{ij}^{st} = \sum_j J_{ij}^{st} = 0$ . (2) follows by re-writing  $B^{st}$  as,

$$B_{ij}^{st} = \frac{R_{ij}}{2} + \frac{R_{ji}p_0(i)}{2p_0(j)}. \quad (2.33)$$

Since  $R$  is a non-negative matrix and  $p_0$  is strictly positive, the number of zeros in  $B^{st}$  cannot exceed the number in  $R$ . Thus if  $R$  is irreducible,  $B^{st}$  must be irreducible. (3) follows from putting the definition of  $B^{st}$  into  $Rp_0 = p_0$ . These properties mean that just like  $R$ ,  $B^{st}$  has a spectral radius  $\lambda_0 = 1$ , has a corresponding left eigenvector  $\Gamma_0(i) = 1, \forall i$  and all of the previous distance relations can be re-cast using the eigenvectors and eigenvalues of  $B^{st}$  in place of those from  $R$ . The left and right eigenvectors and eigenvalues of  $B^{st}$  are defined respectively as,

$$\sum_i \Gamma_\alpha^\dagger(i) B_{ij}^{st} = \nu_\alpha \Gamma_\alpha(j), \quad \sum_j B_{ij}^{st} \phi_\alpha(j) = \nu_\alpha \phi_\alpha(i). \quad (2.34)$$

Using Eq. (2.27) we can define the same relationship between the eigenvectors of the NOR and  $B^{st}$ ,

$$\sqrt{\sum_u \left( \frac{B_{ui}^{st} - B_{uj}^{st}}{\sqrt{p_0(u)}} \right)^2} = \sqrt{\sum_\alpha \nu_\alpha^2 (\Gamma_\alpha(i) - \Gamma_\alpha(j))^2}. \quad (2.35)$$

$B^{st}$  is a slightly more complicated object than  $R$  and as a result has a slightly different meaning. Taking, Eq. (2.33), the second term can be identified as the probability of transitioning, from state  $j \rightarrow i$  in backwards time  $t' - \Delta t \rightarrow t$  through the time reversal matrix  $\hat{R}_{ij}$  of  $R$  [53],

$$\hat{R}_{ij} = \frac{1}{p_0(j)} R_{ji} p_0(i). \quad (2.36)$$

This definition lets us re-write  $B_{ij}^{st}$  as,

$$2B_{ij}^{st} = R_{ij} + \hat{R}_{ij} \quad (2.37)$$

The NOR can then be thought of as a measure of how related forward and reverse transitions are, if two states  $i$  and  $j$  are close in the NOR then the forward and backwards evolutions are similar. This is clearly seen by taking Eq. (2.37) into account when writing Eq. (2.35) for one time step,

$$\left( \sum_u \left| \frac{R_{ui} - R_{uj}}{2\sqrt{p_0(u)}} + \frac{\hat{R}_{ui} - \hat{R}_{uj}}{2\sqrt{p_0(u)}} \right|^2 \right)^{1/2} = \sqrt{\sum_{\alpha=1}^m \lambda_{\alpha}^2 (\Gamma_{\alpha}(i) - \Gamma_{\alpha}(j))^2}. \quad (2.38)$$

Since the publication of [52] and [51] we have found several other examples of the  $B^{st}$  being used in different contexts. Eq. (2.33) is written in [54] where they site it originally coming from, [55]. Neither of these papers apply it to distances in an eigenvector space, and neither of these papers apply these ideas to controlling chaos.

## Coordinate Representation

Now that we have defined a matrix which allows the creation of the NOR for irreversible systems, we can analyze a huge class of new systems. To illustrate this we will use the NOR to re-create the state space for the base of the Sierpinski gasket, which requires a non-detailed balance matrix of transition probabilities. The Sierpinski gasket is an equilateral triangle that is sub divided into ever smaller copies of itself. Fig. (2.5) shows an approximation of the full fractal. This fractal was generated from Pascal's triangle where every odd number is plotted and every even number is not. We will use the NOR to generate the base of the Sierpsinki fractal, namely an equilateral triangle with a smaller one, rotated 180°. The left image in Fig. (2.6) is an illustration of the base of the Sierpinski gasket while the right hand image is generated from the NOR. To

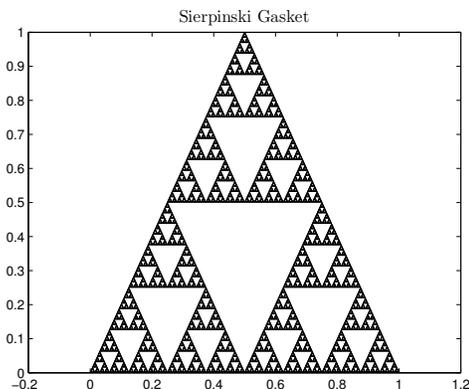


Figure 2.5: The Sierpinski gasket generated from Pascal's Triangle.

generate this states space, each node of the fractal is thought of as a state, while the edges denote possible transitions between states. The transition probability is given by one over the total number of connecting edges.

We can extend this to the base of the 3-D Sierpinski gasket as seen in Fig. (2.7). This image was generated in the same way as the two dimensional version only now including the third left eigenvector. We have overlaid the convex hull to help emphasise the three dimensional nature of the shape.

One should be able to extend this approach to generate more refined versions of fractals but this is left for future work. The next example will be how information embedded in the NOR can be used to form periodic orbits in chaotic systems.

There is more information encoded in the NOR than simply the coordinate representations. It also seems to hold information on return times for orbits. This will be illustrated first with the Logistic map Eq. (1.38). The eigenvector equation for  $B$  can be taken for multiple time steps, where the  $\tau$  time step is given by

$$\sum_i \Gamma_\alpha(i)^\dagger (B_{ij}^{st})^\tau = \nu_\alpha^\tau \Gamma_\alpha(j). \quad (2.39)$$

This shows another advantage to working in the eigenspace of  $B$  instead of

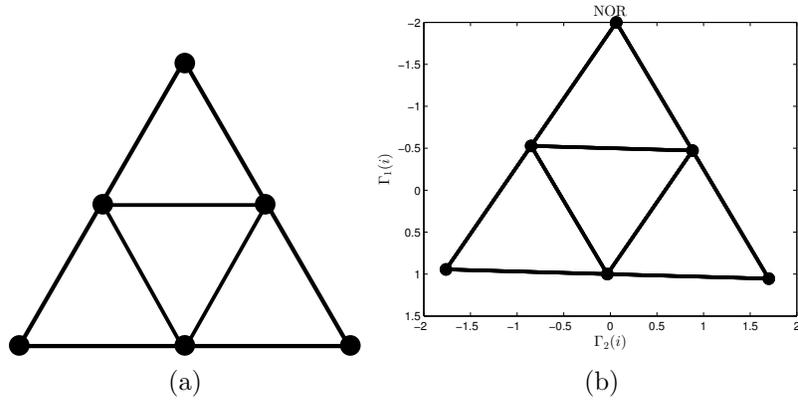


Figure 2.6: (a) is an illustration of the base of the Sierpinski gasket while (b) was made from the first two dimensions of the NOR.

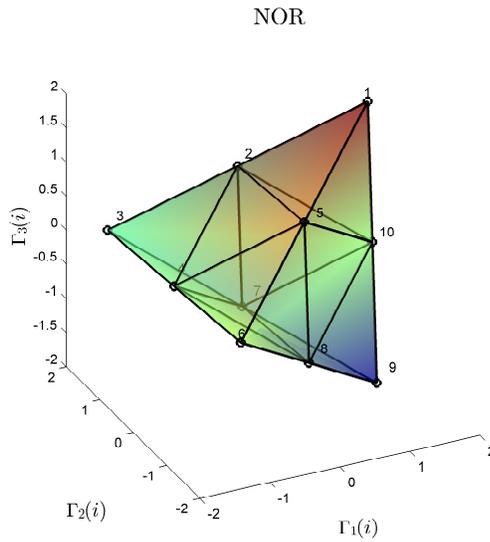


Figure 2.7: The base of the three dimensional Sierpinski gasket.

with  $B$  directly. The left hand side of the above equation is extremely messy to calculate analytically, since  $B = \frac{1}{2}(R + \hat{R})$ , while the right hand side is simply

$\nu$  raised to a power. The generalization of Eq. (2.35) to higher time steps is,

$$\sqrt{\sum_u \left( \frac{(B_{ui}^{st})^\tau - (B_{uj}^{st})^\tau}{\sqrt{p_0(u)}} \right)^2} = \sqrt{\sum_\alpha \nu_\alpha^{2\tau} (\Gamma_\alpha(i) - \Gamma_\alpha(j))^2}. \quad (2.40)$$

To look at the distances for the logistic map using discrete time steps  $n$  we plot the value of the right hand side of Eq. (2.40) such that the distance is normalized and given by the color of the plot, Fig. (2.8). For  $n = 1$  in (2.8) (a) we have the  $f(x) = x$  line, which is zero as we would expect. But there is also a parabola with a copy of itself shifted  $90^\circ$ . For  $n = 3$  in (b) we start to see a higher frequency, low distance, function being added to the surface. (c)  $n = 12$  shows a lot of finer detail forming. For (d) almost all initial distributions would have come to  $p_0$ , and the surface has essentially become constant, meaning no new details form as we further increase  $n$ .

So what do these minimum and maximum distances mean? Firstly, since the distance is symmetric, every relationship is repeated. The vertical line at  $x = 0.75$  in Fig. (2.8) (c) is giving the identical information to the horizontal line at  $x = 0.75$ . These lines correspond to a fixed point in the system. Thus the red is showing that starting at  $x = 0.75$  the system never occupies any other state other than  $x = 0.75$ . If we plot  $x_1$  as a function of  $x_2$  over the  $n = 12$  surface in Fig. (2.9) we see that the parabola in Fig. (2.8) (a-c) exactly line up. Continuing this, we have also plotted  $x_1$  against  $x_3$  and again this function matches up to a minimum line in (2.8) (c). So small distances in the forward and reverse evolutions correspond to the second and third step evolutions. Unfortunately you also have all of the extraneous information and the repetitions of information which at this point we do not know how to separate out.

### 2.4.1 Chaos control in the logistic map

As an application of the NOR that is closer to something applicable to the real world we will continue to use distances in the NOR to form non-chaotic periodic

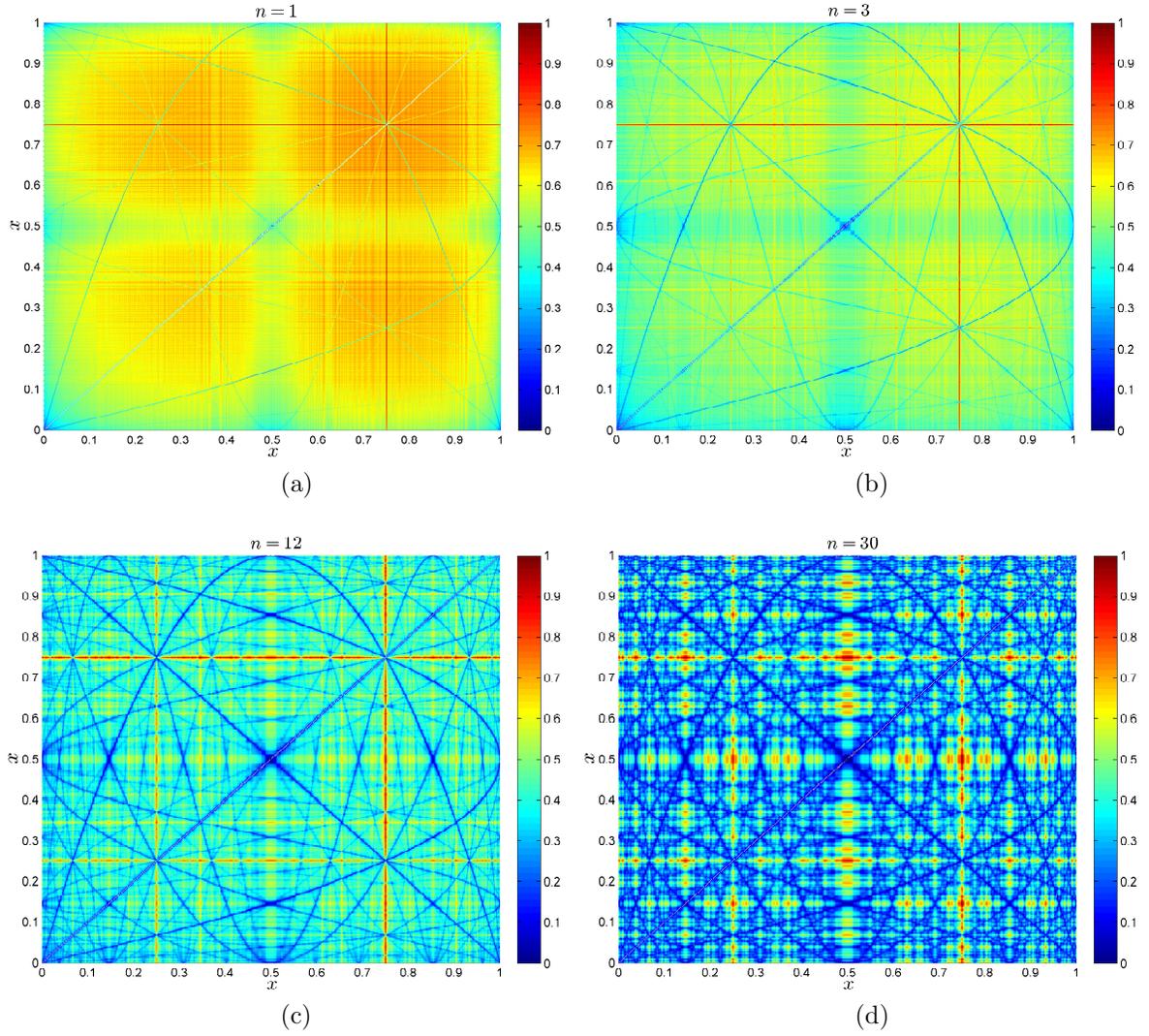


Figure 2.8: The surface of Eq. (2.40) as  $n$  is increased from  $n = 1$  to  $n = 30$ . The  $y = x$  line in (a) has distance zero and the non-blue coloring is due to the limitations of the plotting routine.

orbits in the logistic map (1.38). The control parameter  $a = 4$  will be used so that the system is fully chaotic. The domain,  $X = [0, 1]$  will be divided up into  $N$  states  $G_k = [k, k + 1]/N$ ,  $k = 1, 2, \dots, N - 1$  where  $G_k \cap G_j = \emptyset$  and

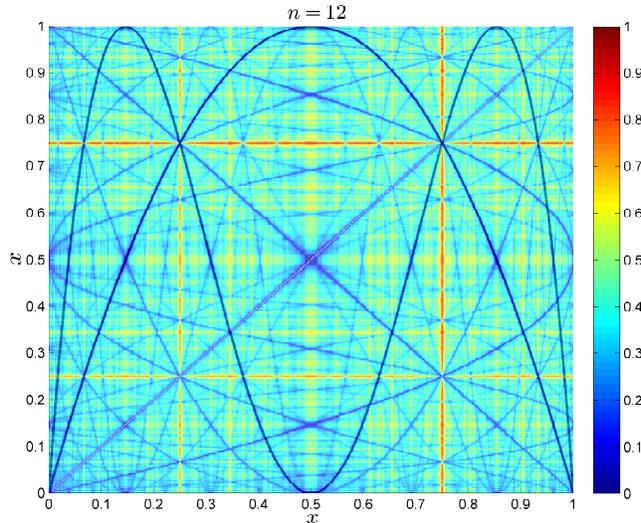


Figure 2.9: Plotting  $x_2$  and  $x_3$  in black as a function of  $x$  for the logistic map, we see that these functions exactly overlap the blue minimum lines seen in Fig. (2.8). Note the white markings are artefacts of the surface algorithm and should be ignored.

$\cup_k G_k = X$ . The position of any orbit  $x_n$  will then belong to a bin  $G_k$  meaning  $R_{jk}$  is,

$$R_{jk} = \Pr[x_{n+1} \in G_j \mid x_n \in G_k]. \quad (2.41)$$

Thus,  $R$  is found from following a large ensemble of orbits as they evolve through the state space  $X$ . From  $R$  we build  $B^{st}$  and  $J^{st}$  using Eq. (2.32) and Eq. (2.18) respectively.

Probably the greatest asset of this approach is that we do not require any knowledge of the equations of motion. This is especially advantageous for large complex real world systems where we do not know the equations of motion and are often forced to use a set of solvable candidate equations which are gross approximations of the fundamental equations, such as approximations of the Naiver Stokes equations. By working with solvable approximate equations, we throw away much of the physics of the system. Here we hope to keep much of this inherent physical information but at a cost. As previously mentioned how

we coarse grain our system is rather arbitrary and can possibly influence the results. Also we usually end up with very large sparse matrices which quickly reach a size that creates computational challenges. For the logistic map our domain is divided uniformly into grains and though not pictured, adjusting the size of grains by small amounts does not seem to greatly change the resulting NOR.

The original idea in [52] was that given two grains  $G_k$ , and  $G_{k+1}$  which are adjacent to each other in  $X$ , if these two have a small distance in the NOR, then orbits originating from them will be mapped to similar areas, meaning the union of  $G_k \cup G_{k+1}$  will have a small divergence of trajectories. To implement this approach, we first evolve an initial condition for  $n$  iterations. Then we define the set  $\mathcal{G}$  of all grains within some distance  $\epsilon$  to the orbit. This distance  $\mathcal{G} = \{\|x_n - G_k\|_2 < \epsilon\}$  is a free parameter of the system. This set of grains corresponds to a set of points in the NOR. The grain we perturb the position of the orbit to is the one in  $\mathcal{G}$  which has the smallest off diagonal. Though since the original publication it has become clear that this picture of direct correspondence between forward Lyapunov exponents and distances between off diagonal states in the NOR is not quite correct, though it can lead to similar results.

For the results that follow, we used,  $m = 3$  in Eq. (2.38), the right hand of Eq. (2.38) side will now be designated  $D_{NOR}$ . Plotting  $D_{NOR}$  as a function of  $x$ , we see that there is an apparent connection between the Lyapunov exponent and distance between nearest neighbours,  $D_{NOR}(k, k + 1)$ . Choosing the distance between nearest neighbors can also be thought of as a kind of gradient at state  $k$  in the NOR. Fig (2.10) shows  $D_{NOR}$  for  $m = 1500$  in black with the minimum distance denoted with the red square. The minimum value is very close to  $x = 0.5$ , which is also the minimum finite time Lyapunov exponent given by  $l_i = \log(|a(1 - 2x)|)$  (shown in red). Though the curve of  $D_{NOR}(k, k + 1)$  becomes more like the curve of  $l_i$  as we increase  $m$ , the two are never equal. Using only  $m = 2$  dimensions is enough to already find the minimum around

$x = 0.5$ . Two of the other local minima we see in Fig. (2.10),  $f(0.1452)$  and  $f(0.856)$  are both points that map in one time step to the unstable fixed point,  $f(0.1452) \approx f(0.856) \approx 0.5$ . The local minimum at  $x = 1$  maps to the local minimum and unstable fixed point at  $x = 0$ .

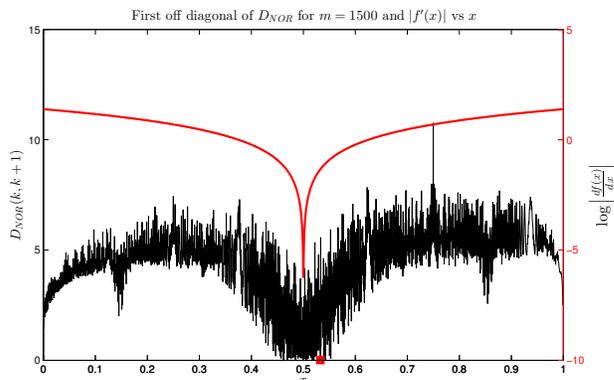


Figure 2.10:  $D_{NOR}(k, k+1)$  using  $m = 1500$  is shown in black with the minimum value shown with the red square at  $x = 0.513$ . The finite time Lyapunov exponent is shown in red. We can see both functions have similar behaviour.

In Fig. (2.11) (a) an ensemble of 100 initial conditions are evolved freely under the logistic map for the first  $n = 15$  iterations. Control is then implemented as explained above on each time step using  $\epsilon = 0.1$ . We see that before control the orbits fill out the entire state space, after control they very quickly form one periodic orbit, though each orbit may be out of phase with each other. Fig. (2.11) (b) shows the Lyapunov exponents for the same evolution and as expected upon implementing control they all steadily fall until for Eq. (1.30)  $L < 0$ , and the system is not chaotic.

It is fairly obvious that the size of  $\epsilon$  determines how quickly we can bring the logistic map out of chaos, if at all. Due to the chaotic nature of the system, the cut off where  $\epsilon$  becomes too small to control chaos is very sensitive to both the initial conditions of the system and how  $R$  and in turn  $B^{st}$  are built. To show the sensitivity of  $\epsilon$ , two initial conditions were randomly chosen within the contracting region,  $\frac{3}{8} < x < \frac{5}{8}$  and two from the expanding region,  $x \in$

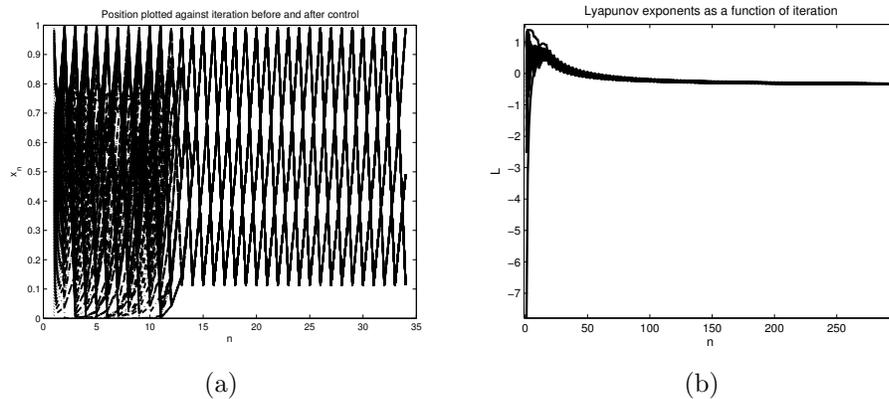


Figure 2.11: (a) shows the evolution of 100 orbits with respect to time. After evolving freely for  $n = 15$  iterations, we implement control and all trajectories fall into the same periodic orbit. (b) shows the lyapunov exponent as a function of time for all 100 orbits. After control  $L$  continually falls until  $L < 0$  and the system is no longer chaotic.

$[0, \frac{3}{8}] \cup [\frac{5}{8}, 1]$ . The initial conditions  $x = [0.1576, 0.4854, 0.6324, 0.9134]$  were evolved using  $\epsilon = 0.9$  and  $\epsilon = 0.8$ . In Fig. (2.12) we can see that, for  $\epsilon = 0.9$  the system is brought out of chaos, while  $\epsilon = 0.8$  the system is not, since,  $L > 0$ .

Requiring  $L < 0$  is actually an unnecessarily strict condition for the system to formally not be considered chaotic. Since requirement (2) for chaos that the system is mixing is immediately violated as soon as the orbits become periodic through state space. In the next section, we will again form periodic orbits though this time from a continuous 3-dimensional system. For the Lorenz system we will always have one positive Lyapunov exponent despite the formation of periodic orbits.

## Periodic orbits in the Lorenz system

Since we have shown we can stop chaotic behaviour in the logistic map, we now tackle a more realistic continuous system. The Lorenz system introduced

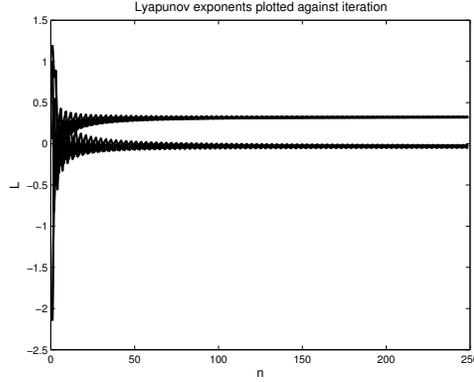


Figure 2.12: Using  $\epsilon = 0.9$  the initial conditions  $x = [0.1576, 0.4854, 0.6324, 0.9134]$  are brought out of chaos.  $\epsilon = 0.8$  for the same initial conditions is still chaotic as seen with the upper line being greater than zero.

in chapter (1) has three control parameters which will take the values  $\sigma = 10$ ,  $\beta = 8/3$  and  $\rho = 28$  in the work that follows. See Section (1.3.2) for more details. In what follows we will not make all three Lyapunov exponents negative, but we will negate the second requirement for chaos of topological transitivity, i.e. we will form periodic orbits.

To generate  $R$  in, Eq.(2.9) and thus  $B^{st}$ , Eq. (2.32) we first must divide the Lorenz attractor  $\Omega$  into  $N$  coarse grains. Each grain will be a disjoint subset of the original space labelled,  $\mathbf{G}(i, j, k) \subset \Omega$ , while the individual grain dimensions are,  $x_i = 1, 2, \dots, N$ . To keep notation to a minimum the position along an orbit of the system is given by the coordinates  $\mathbf{x} = [x_1, x_2, x_3]$  without subscript, while coarse grains are given with subscripts  $[x_{1,i}, x_{2,j}, x_{3,k}]$ , meaning every orbit belongs to a specific coarse grain,  $[x_{1,i}, x_{2,j}, x_{3,k}] \in \mathbf{G}(i, j, k)$ . Along with our coordinates in the attractor's space, we also have a corresponding set of linear coordinates in the NOR space. These will simply be given by indices,  $Q_l$ . For example if we divide our attractor up into  $r_i^{1 \times \mathcal{N}}$ ,  $s_j^{1 \times \mathcal{M}}$ ,  $t_k^{1 \times \mathcal{P}}$  grains (note order matters here). Then our linear coordinates  $Q_l^{1 \times \mathcal{NMP}}$  are found through the

transform,

$$l = (k - 1)\mathcal{N}\mathcal{M} + (j\mathcal{N} - \mathcal{N} + i). \quad (2.42)$$

Likewise the back transform from linear coordinates to the indices of the attractor are defined through,

$$\begin{aligned} i &= l - \text{floor}\left(\frac{l-1}{\mathcal{N}}\right)\mathcal{N}, \\ j &= 1 + \text{floor}\left(\frac{l-1}{\mathcal{N}}\right) - \text{floor}\left(\frac{l-1}{\mathcal{N}\mathcal{M}}\right)\mathcal{M}, \\ k &= 1 + \text{floor}\left(\frac{l-1}{\mathcal{N}\mathcal{M}}\right). \end{aligned} \quad (2.43)$$

Floor in the above equations rounds the argument down to the nearest integer. Using  $n = 40$  divisions gives a matrix  $R$  with  $40^6$  elements, which can be stored as a sparse matrix but is too large to be stored as a full matrix. Thankfully we can eliminate much of our matrix as we only want the transitions that generate our irreducible operator which will represent only the connected attractor. Thus taking our large matrix we eliminate any rows or columns consisting of zero values or absorbing states until  $R$  is irreducible. This means we are left with a greatly reduced matrix  $R^{N \times N}$ ,  $N = 3171$ , that still gives us a faithful representation of the underlying attractor. The set of irreducible states is found by simply evolving orbits from randomly chosen initial conditions and then using this to form the irreducible set. To construct  $R$  we can either directly calculate the transition probabilities by calculating Eq. (2.9), or we can evolve a large ensemble of orbits  $N_{ens}$  which is uniformly distributed over each state  $G$  and calculate Eq. (2.13) by assuming each orbit occupies an equal initial volume  $\Delta v_j = Vol(I_j)/N_{ens}$ . The convex hull is used to approximate the initial total volume of the ensemble and Eq. (2.13) is approximated as,

$$R_{ij} = \sum_{k=1}^{N_{ens}} \frac{\delta(u(y_k^*) \in I_i)}{|\det J^t(y_k^*)|} \Delta v_j, \quad (2.44)$$

where we sum over each test orbit out of the total number  $N_{ens}$ . For the transition matrix used to control the Lorenz system later on we used,  $N_{ens} = 5000$  orbits. So far we also seem to generate at least subjectively the same NOR for simulations using different initial conditions but the same  $N_{ens}$ .

To get a better feel for what the NOR looks like for the Lorenz system, we plot the NOR for  $2 \leq \alpha \leq 9$  in Fig. (2.13) in sets of three. Each image

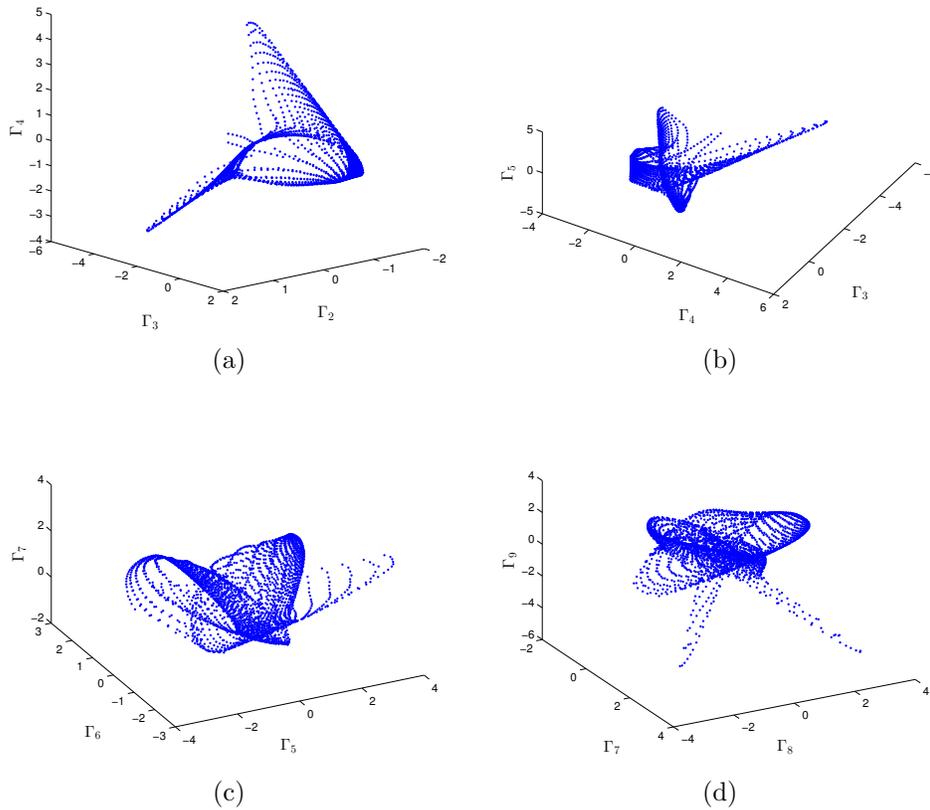


Figure 2.13: Different slices of the NOR, made using the  $\alpha$ -th left eigenvectors.

can be thought of as a three dimensional slice of a higher dimensional shape. The distance between each point represents the relationship between two course grained approximations of the phase space. This relationship is measured using Eq. (2.40).

To implement control, we choose an orbit which has freely evolved from time  $t_0$ . The first parameter we use is the perturbation time  $\tau = t\Delta t$  where  $\tau$  is the time span between perturbations of the system;  $\tau$  is a free parameter determined by the system under study. From the orbit at  $\tau$ ,  $\mathbf{x}(\tau)$  we find all  $\mathbf{G}(i, j, k)$  within a distance  $\epsilon$  from  $\mathbf{x}$  as shown in Fig. (2.14) (a). Again,  $\epsilon$  is a free parameter of the system. Fig. (2.14) (b) shows in red the corresponding grains within  $\epsilon$  of  $\mathbf{x}(\tau)$  in the NOR. Out of all red NOR states, we select one of the two points which are closest to each other.

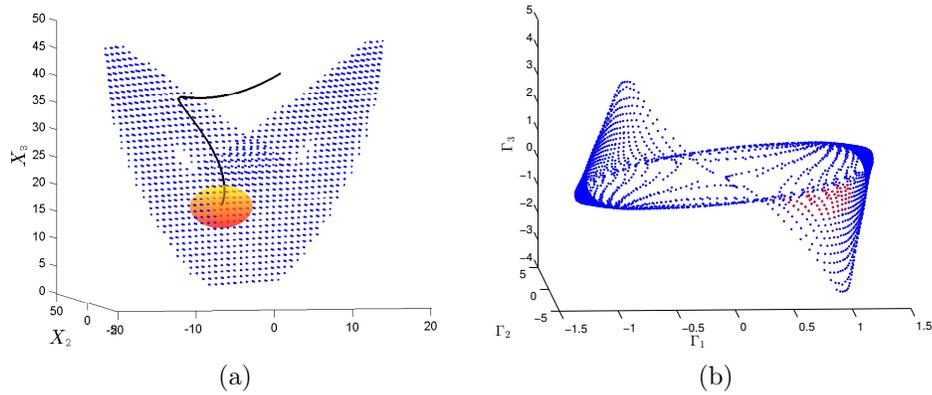


Figure 2.14: (a) A typical orbit is evolved in black up to time  $\tau$ . Then all orbits within a distance  $\epsilon$  of  $G_{i,j,k}$  are found. These are surrounded by the sphere. The corresponding states are then shown in (b) where  $\Gamma_1$ ,  $\Gamma_2$  and  $\Gamma_3$  are plotted in the NOR with the red dots.

As one might expect the range of values for  $\tau$ ,  $\epsilon$  and  $m$  where the NOR is effective appears to be limited. If we make  $\tau$  too large, the system continues to be chaotic. If we make  $\epsilon$  too small, again the system remains chaotic. Probably the most interesting fact is that adjusting the number of dimensions used has a drastic effect on the NOR. The key to the NOR being effective is for small distances in the NOR to correlate with orbits remaining nearby to each other. To try and get a hold on these free parameters of the system we first look at  $\tau$ . Fig. (2.15) again shows  $D_{NOR}$  on the x-axis. On the y-axis we show the distance between a fixed grain  $G_f$  and an initial state,  $\|\mathbf{x}(t_0) - G_f\|_2$ . The z-

axis shows the distance between the evolution of the reference grain  $G_f$  and the state after time  $\tau$ ,  $\|f^\tau(\mathbf{x}(t_0)) - f^\tau(G_f)\|_2$ . Ideally we hope for small distances in  $D_{NOR}$  to correspond to small distances between orbits after evolving for time  $\tau$ . For short times of  $\tau = .1$  and  $m = 3$ , shown in top left of Fig. (2.15) we see just this. As we move left and increase  $\tau$  from  $\tau = 0.3$  to  $\tau = 0.4$ , we see that small distances in the NOR sometimes lead to large final distances. Increasing  $m$  remedies this slightly, as seen using the same time increments for the bottom three images but now  $m = 119$ .  $m = 119$  was chosen since each  $\alpha$  for  $\alpha > m$  contributes less than  $1 \times 10^{-4}$  to the over all distance. Using more dimensions does seem to give us better results with small distances in the NOR leading to small distances between evolved states. The picture possibly isn't totally clear though, for anecdotal evidence suggests that using  $m = 3$  dimensions gives the best chance of forming a periodic orbit, leaving us with a kind of paradox. The following are some examples of control using the NOR. For the first example Fig. (2.16) we used  $\tau = 0.2$ ,  $m = 3$  and  $\epsilon = 4$ . Starting in (a) we have both the controlled orbit in red and a second orbit which has the identical initial condition but is allowed to evolve freely in blue. The second figure (b) shows just the controlled trajectory in black which forms its periodic orbit. The third figure (c) shows a time series of  $x_1$  and  $x_2$  and more clearly shows the apparent periodic orbit. To check that this is truly a periodic orbit the bottom right figure (d) is the frequency power spectrum of the trajectory. The single large peak corresponds to the orbit with one period proving the system's orbit is indeed periodic. If the system was chaotic instead, there would be a continuous distribution of peaks. The second example of control is using  $\tau = 0.3$  and increasing  $m = 119$ , we also find that we have formed a periodic orbit, though this time having a longer period. The resulting power spectrum plot Fig. (2.17) (d) does not show as defined a peak due to the longer period not being made up of as many cycles. Another possible contribution to the spread of the peak is seen in Fig. (2.17) (c) where on the fourth to last orbit, the system appears to escape the periodic cycle only to return one cycle later. Given the finite length

of our simulation one can never know if this is part of a longer periodic orbit or a deviation from the current one.

The final example is using  $m = 7$ ,  $\epsilon = 4$  and  $\tau = 0.3$ , though here we have evolved the system for total time  $T = 40$  instead of the previous examples where the system was evolved for either,  $T = 10$  or  $T = 12$ . In this example we see again we have generated a periodic orbit and it appears stable for the entire evolution once it has formed. What is more interesting is that the Lyapunov exponents for the controlled orbit are still greater than zero, indicating sensitive dependence on initial conditions as shown in Fig (2.19). To understand how the control works, recall that when a trajectory is perturbed it is randomly placed in the selected state. As it then freely evolves, the system is still unstable and thus orbits still diverge. The distance they diverge is small enough to have the control select the same state to perturb the system back into after time  $\tau$ . The regular perturbations then generate the periodic orbit seen in Fig. (2.18).

The above examples of control over the system are very appealing yet limited. Though we do in fact generate periodic orbits for most randomly chosen initial conditions, we do not yet know how  $\epsilon$ ,  $\tau$  or  $m$  affects the chances of forming an orbit or which orbit is chosen. In the end, this approach holds much promise, but in its current form is probably only applicable in the worst case scenario, i.e. when one does not know the equations of motion of the system. The next section we will nonetheless continue to push forward and show that we can in fact define a matrix  $B$  for any arbitrary PDF.

## **$B$ and NOR for arbitrary PDFs**

For all of the previous work on the OR,  $p_0$  has been used to relate the NOR space to transitions in state space. In this sense we are assuming our system is near or well represented by  $p_0$ . Though we have just seen this assumption is still useful for some non-equilibrium systems, we would like to define the NOR for any arbitrary PDF. The main assumption is that there exists a matrix  $R_{ij}$  which is fixed for the evolution of the system. If we start with an ensemble of

orbits in a subset of phase space then the evolution of these orbits only feels the dynamics of this subset of phase space. That is some matrix of transition probabilities  $r_{ij} \subseteq R_{ij}$ . Similar to Sect. (2.4.1) where the system was evolved by a mapping  $f^t(x)$  which was unknown, the probability distributions  $p(i, t)$  will be evolved by the complete operator  $p(i, t') = \sum_j R_{ij}p(j, t)$  for which we only know part of at any time  $t$ . Since we assume  $R_{ij}$  is still irreducible  $r_{ij}$  will eventually grow to equal  $R_{ij}$ .

The goal of this finite time evolution view point of the system is to shed light on the relationships inherent in the system between one non-equilibrium distribution and its evolution to another non-equilibrium distribution. This is in contrast to what is done historically where the system either begins or ends in equilibrium, see [56, 57, 1, 58] for several examples.

For any arbitrary distribution  $p(i, t)$ ,  $\sum_i p(i, t) = 1$ ,  $p(i, t)_{\bar{n}} \neq 0 \forall i$  is the subset of states with non-zero probabilities, This is defined with its zero counter part as,

$$\begin{aligned}\bar{n} &\equiv \{p(i, t) \neq 0\} \\ \bar{n}_z &\equiv \{p(i, t) = 0\}.\end{aligned}\tag{2.45}$$

$B$  is then defined using this non-zero set as,

$$B_{ij} = r_{ij} - \frac{J_{ij}}{2p(j, t)_{\bar{n}}}.\tag{2.46}$$

$r_{ij}$  is a sub-matrix of  $R_{ij}$  which results from using only the values of  $R_{ij}$  which correspond to non-zero values of  $p(i, t)$  as shown in Fig. (2.20). Because  $r_{ij}$  does not always use all of the transition probabilities, the matrix  $B$  to be defined shortly will have some different properties from  $B^{st}$ .

$J_{ij}$  will then be defined as the current of probability between states  $j$  and  $i$  defined with no summation as,

$$J_{ij} = r_{ij}p(j, t)_{\bar{n}} - r_{ji}p(i, t)_{\bar{n}}.\tag{2.47}$$

When  $p(i, t) = p_0(i)$ ,  $J_{ij} \rightarrow J_{ij}^{st}$ .  $J^{st}$  follows a conservation of current rule,  $\sum_i J_{ij}^{st} = \sum_j J_{ij}^{st} = 0$ . This is no longer the case out of equilibrium,  $J_{ij}$  is instead

$$\sum_j J_{ij} = p(i, t')_{\bar{n}} - p(i, t)_{\bar{n}}, \quad t' = t + \Delta t. \quad (2.48)$$

This follows from  $\sum_j r_{ij} p(j, t)_{\bar{n}} = p(i, t')_{\bar{n}}$  where  $p(i, t')_{\bar{n}} \subseteq p(i, t')$ ,  $p(i, t')$  is the full distribution evolved by  $p(i, t') = \sum_j R_{ij} p(j, t)$ .

Given our construction of  $B$  we are guaranteed that it is a non-negative square matrix. So to be able to define the NOR we also need  $B$  to be irreducible and similar to a symmetric matrix. Given that we are assuming  $R_{ij}$  is irreducible then as long as  $p(i, 0)_{\bar{n}}$  corresponds to only non-zero transition probabilities in  $R$ ,  $r_{ij}$  will also be irreducible. The relation between  $B$  and its symmetric transform will be demonstrated using a slightly different but equivalent way to definitions from Sect (2.4). Again using similarity transforms we define the matrix  $U_{ij}$  which is similar to  $r_{ij}$  ( $U_{ij} \sim r_{ij}$ ) through  $U_{ij} = \frac{1}{\sqrt{p(i, t)}} r_{ij} \sqrt{p(j, t)}$ . We can re-write  $U_{ij}$  as,

$$U = \frac{1}{2} (U + U^\dagger) + \frac{1}{2} (U - U^\dagger) = S + U^{as}, \quad (2.49)$$

where  $S$  and  $U^{as}$  are the symmetric and anti-symmetric parts. Looking only at the symmetric part we have,

$$\begin{aligned} S &= \frac{1}{2\sqrt{p(i, t)}} r_{ij} \sqrt{p(j, t)} + \frac{1}{\sqrt{p(j, t)}} r_{ji} \sqrt{p(i, t)}, \\ &= \frac{1}{2\sqrt{p(i, t)}} r_{ij} \sqrt{p(j, t)} + \frac{1}{2\sqrt{p(i, t)}} \left( \frac{1}{p(j, t)} r_{ji} p(i, t) \right) \sqrt{p(j, t)}, \\ &= \left( \frac{1}{\sqrt{p(i, t)}} \left( \frac{r_{ij} + \hat{r}_{ij}}{2} \right) \sqrt{p(j, t)} \right). \end{aligned} \quad (2.50)$$

Thus we see that  $S_{ij} \sim B_{ij}$  are similar and likewise it is easy to show that  $U^{as} \sim \frac{J_{ij}}{2p(j, t)}$ . Since  $S$  is guaranteed to have a set of  $N$  orthogonal eigenvectors

(see Thm 2.5.6 [26]), meaning  $B$  is also guaranteed this set through the relation,  $\psi = \phi_\alpha(i)/\sqrt{p(i, t)}$ ,  $\psi = \Gamma_\alpha(i)\sqrt{p(i, t)}$  again where  $\psi_\alpha$  are the eigenvectors of  $S$ . Therefore we can construct the NOR for any arbitrary PDF. The trivial case which we are ignoring is  $r$  being a matrix of all zeros. Though we have a set of eigenvectors for  $B$ , it is no longer a stochastic matrix, instead,

$$\begin{aligned} \sum_i B_{ij} &= \sum_i \frac{r_{ij}}{2} + \frac{p(j, t')_{\bar{n}}}{2p(j, t)_{\bar{n}}}, \\ &= \frac{1}{2} \left( 1 - \sum_{k \in \bar{n}_z} R_{kj} \right) + \frac{p(j, t')_{\bar{n}}}{2p(j, t)_{\bar{n}}}. \end{aligned} \quad (2.51)$$

This means that although we still have a single spectral radius, it is not guaranteed to be equal to one,  $\lambda_0 \neq 1$ . Next we will look at a basic example for  $B$ . We will see that it is relatively easy to generate the NOR at each time step for a Markovian toy problem.

## **$B$ for two state system**

Imagine two clusters of states as was previously defined in Sect. (2.3). The system can transition from any state in a cluster to any other state in that same cluster in one time step. Each cluster is represented by a matrix  $r_1$ ,  $r_2 \in \mathbb{R}^{n \times n}$ . Each matrix  $r_1$  or  $r_2$  is a sub-matrix of the larger matrix of transition probabilities  $R$ . In this example  $r_1$  is made up of 20 states while  $r_2$  is made up of 15 states. In the figures to follow we have labelled states belonging to  $r_1$ ,  $i = 1, 2, \dots, 20$  and states belonging to  $r_2$ , are  $i = 31, 32, \dots, 40$ . Though it needs to be noted that the order of labelling is completely arbitrary with only the distance between states being of consequence. There is a small probability of transitioning along a set of states from  $r_1$  to  $r_2$  and vice versa. These strings of states we will label as corridors as illustrated in Fig. (2.21) and consist of 10 states each. These states are labelled  $i = 21, 22, \dots, 30$  and  $i = 46, 47, \dots, 55$ . Together this creates the irreducible system. To give a feel for the NOR based

on an arbitrary distribution, we initially start with a delta function in one state of  $r_1$ , though one can pick any non-zero value of  $R$  to start in. Each initial condition gives a different set of probability distributions as the system evolves to  $p_0$ . Unlike using  $B^{st}$  the NOR now only gives us information about the state space explored by the system up to time  $t$ .

By the second time step the delta function has spread out over  $r_1$ . The surface of  $D_{NOR}$  Eq. (2.40) is given in Fig. (2.22) (a). The variation in distance is due solely to the different random values in  $r_1$ . For  $n = 3$  to  $n = 8$  (b) the system has expanded its occupied state space to now cover most of the corridor from  $r_1$  and  $r_2$  but not occupying  $r_2$  itself. These transitions are on each side of the diagonal where there is a small distance. States further out from the diagonal for which the system cannot reach have larger distances. We also see that the detail in  $r_1$  shown in (a) has been smeared over as more of the state space is explored. In essence every state in  $r_1$  has become equivalent. On  $n = 14$  (c), the system spreads over  $r_2$  and we again see variation in the details of  $r_2$ . Especially the large distances for states 43 and 44.  $n = 15$  (d) the system has found the beginning of the other corridor back to  $r_1$  and interestingly the variation in  $r_2$  has mostly disappeared, now the metric is treating almost all of the previously explored states space as one distance and thus one state. This smearing of all states into the same distance is even more pronounced for  $n = 19$  (e). One can just make out the different color blue between the allowed transitions in the irreducible operator and the excluded regions between 35 and 45. The reason the previously excluded regions look similar to the allowed transitions is that the maximum distance has doubled from (d) leaving all smaller distances lumped into the same color.  $n = 24$  (f), the system finally connects back to  $r_1$  and we start to see a change in the NOR. Now there is a definite differentiation between distances in  $r_1$ ,  $r_2$  and inaccessible areas of state space. Between  $n = 25$  and  $n = 59$  this differentiation increases along with two more paths of small distance which form between  $r_2$  and  $r_1$ . These represent the time reversal probability of travelling back down the corridors.

From  $n = 60$  onwards the NOR surface does not appreciably change, but curiously the stationary distribution of  $B$  is not equal to the stationary distribution of  $B^{st}$  as one would expect. Both are plotted in Fig. (2.23) when  $n = 60$  and we see a very large difference, especially in the  $r_2$  area. The distance between any two probability distributions can be defined using Wootter's distance [59],

$$d(p_0(i), p(i, t)) = \cos^{-1} \left[ \sum_i \sqrt{p_0(i)} \sqrt{p(i, t)} \right]. \quad (2.52)$$

See the next chapter for more details. Fig. (2.24) shows that the distance between  $p(i, t)$  and  $p_0$  of  $B^{st}$  starts decreasing very rapidly at first, then becomes extremely slow for longer times. The linear nature of the logarithm means the distance decays like,  $d(p_0, p(i, t)) \propto e^{\gamma t}$  with  $\gamma = -3.83 \times 10^{-4}$ . Finally, as remarked earlier, the summation over  $B_{ij}$ ,  $\sum_i B_{ij} \neq 1$ , much of the time. When the summation is one, this signifies conservation of probability. By plotting  $\langle \sum_i B_{ij} \rangle$  (averaged over all states  $j$ ) as a function of time, we can see how the total probability of the system changes. Fig (2.25) shows this average probability and right away, on the second time step, the average total probability greatly exceeds one meaning our analogy of  $B_{ij}$  as a stochastic operator no longer holds. From  $n = 3$  to  $n = 13$  the system is evolving across the first corridor to  $r_2$  and  $\langle \sum_i B_{ij} \rangle \approx 1$ . Then the system reaches  $r_2$  and again there is a large increase in total probability before approaching unity and the stationary marginal distribution  $p_0(i)$ .

We are clearly at the beginning of studying the distance  $D_{NOR}$  for an arbitrary probability distributions. But in the quest to understand non-equilibrium systems generality is the holy grail, this method then shows tantalizing promise at increasing our understanding of non-equilibrium systems and their evolutions. Future work will be to look into defining connections between distances in the NOR and traditional physical measures such as work and dissipation of the system.

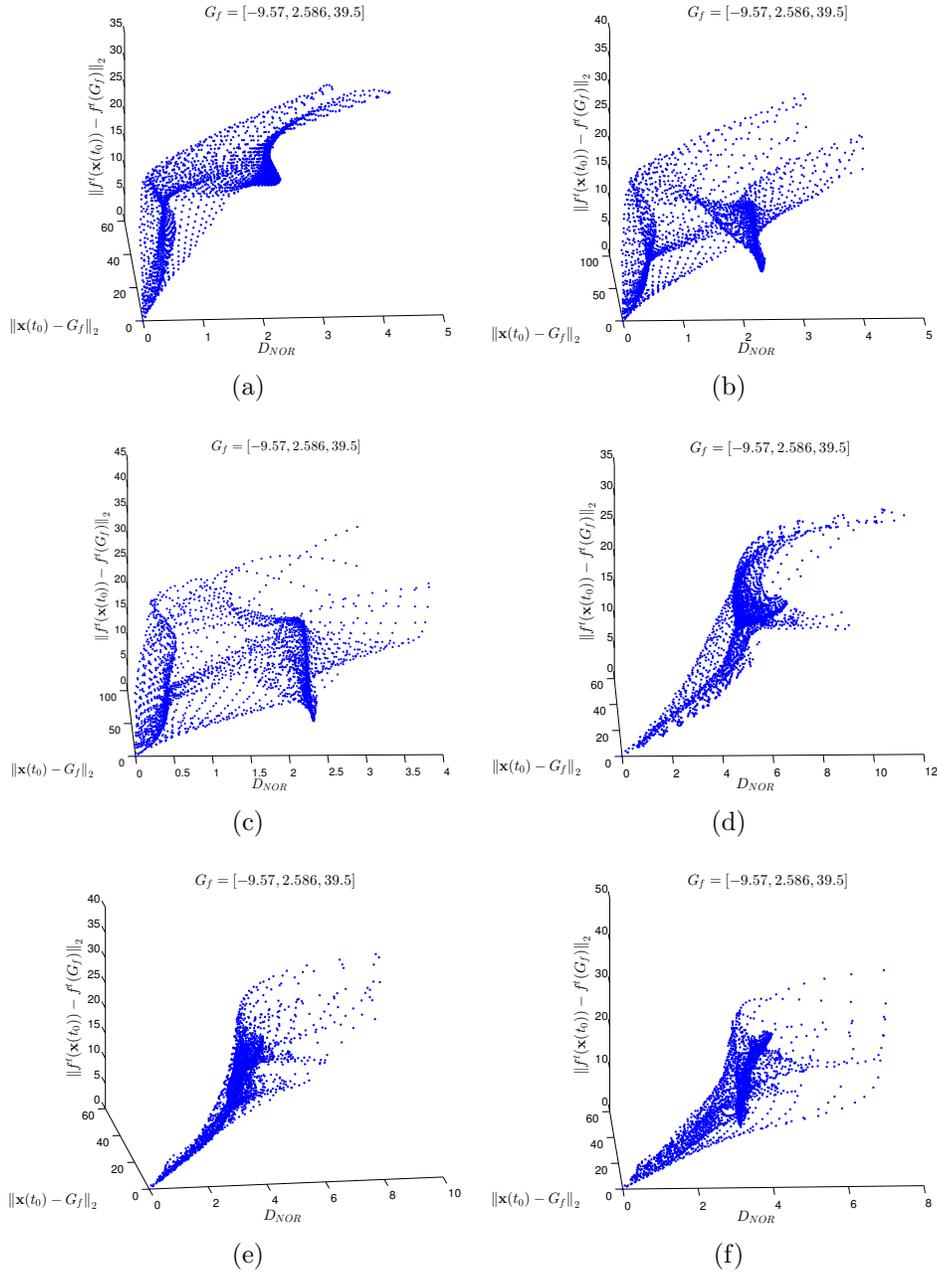


Figure 2.15: (a-c)  $m = 3$  is fixed, while  $\tau$  increases from (a-c) by,  $\tau = .1$ ,  $\tau = 0.3$  and  $\tau = 0.4$ . (d-f) same  $\tau$  increments but now  $m = 119$ .

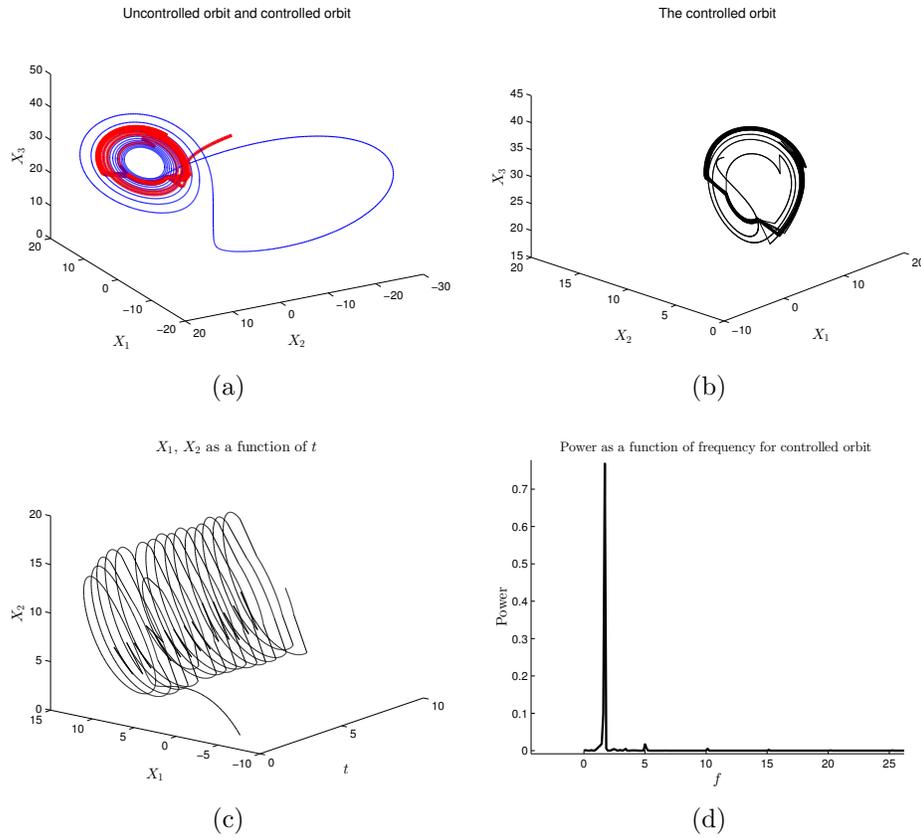


Figure 2.16: The above example of control uses  $\tau = 0.2$ ,  $m = 3$  and  $\epsilon = 4$ . (a) is the controlled orbit in red and the un-controlled orbit in blue. Both start from identical initial conditions. (b) shows just the controlled orbit and (c) is the time series of  $x_1$  and  $x_2$ . (d) gives the power spectrum whose single peak proves this is a periodic orbit.

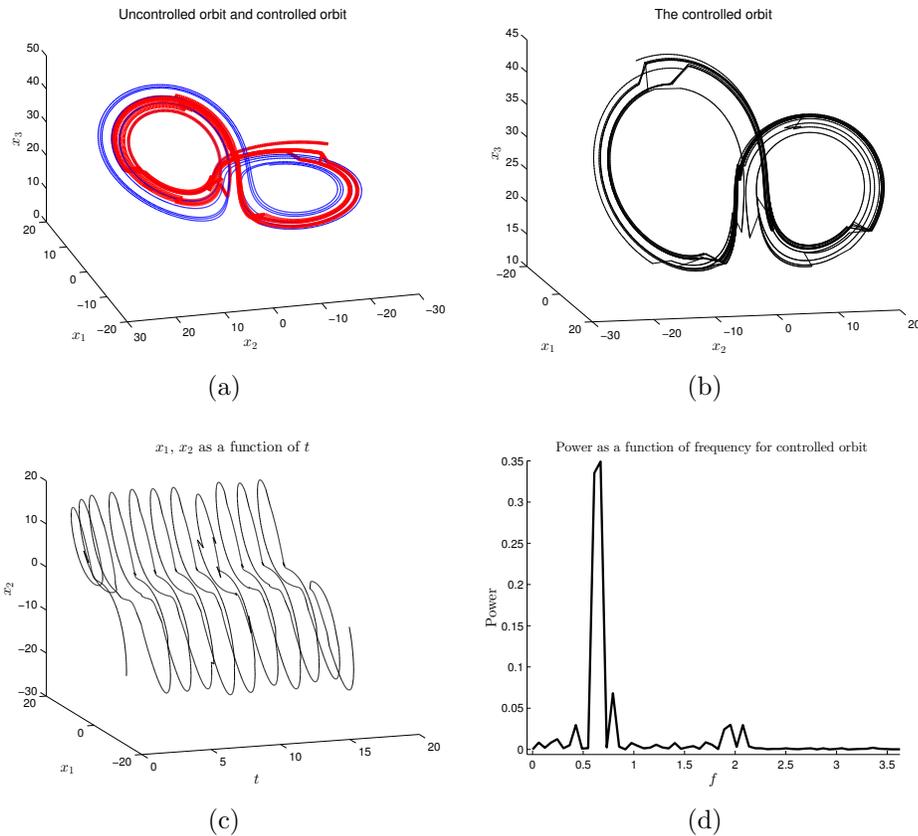


Figure 2.17: The above example of control uses  $\tau = 0.3$ ,  $m = 119$  and  $\epsilon = 4$ . (a) is the controlled orbit in red and the un-controlled orbit in blue. Both start from identical initial conditions. (b) shows just the controlled orbit and (c) is the time series of  $x_1$  and  $x_2$ . (d) gives the power spectrum whose single peak proves this is a periodic orbit though the longer period means this peak is less precise.

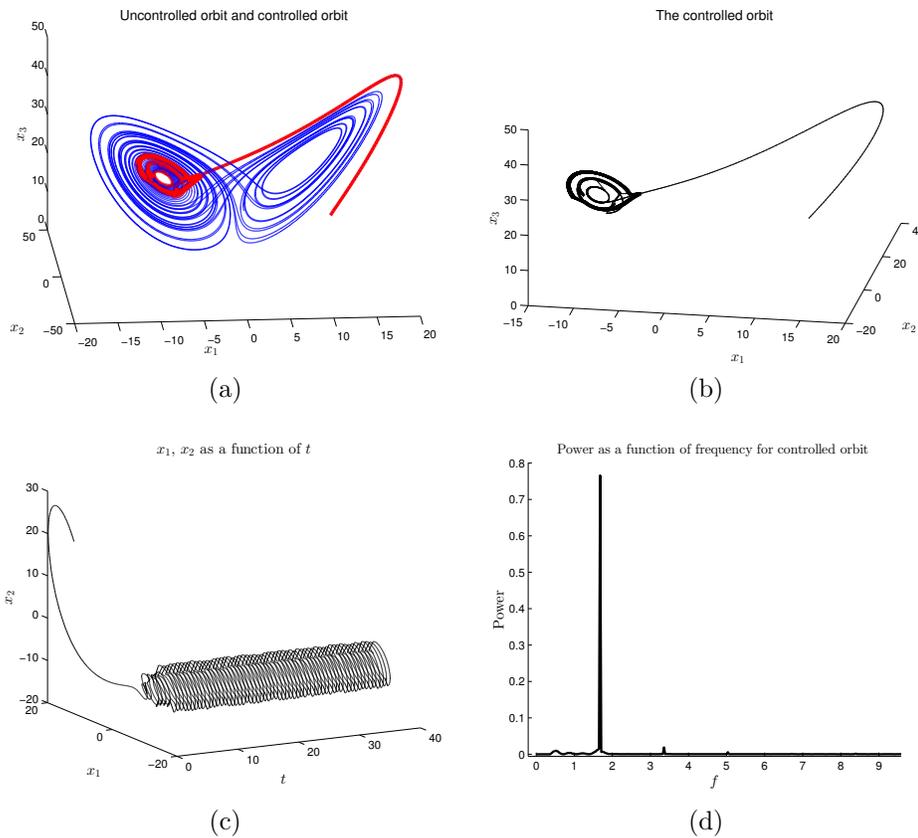


Figure 2.18: The above example of control uses  $\tau = 0.3$ ,  $m = 7$  and  $\epsilon = 4$  but the system has been evolved for twice as long as the previous examples. (a) is the controlled orbit in red and the un-controlled orbit in blue. Both start from identical initial conditions. (b) shows just the controlled orbit and (c) is the time series of  $x_1$  and  $x_2$ . We see that this orbit seem extremely stable with vary little variation over the systems evolution. (d) gives the power spectrum whose single peak proves this is a periodic orbit.

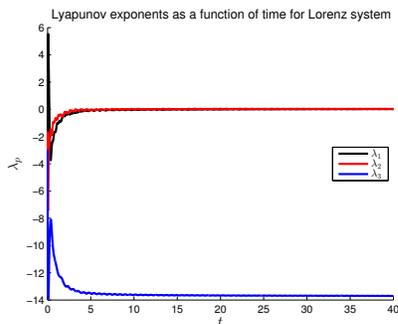


Figure 2.19: The lyapunov exponents for the Lorenz system with control. The final values are  $\lambda_p = [0.0111, 0.0310, -13.6959]$ .

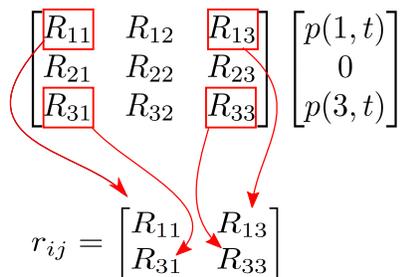


Figure 2.20: An example of how we form  $r_{ij}$ . For the  $j$ th column of  $r_{ij}$  and all non-zero values of  $p(i, t)$  we select the corresponding values from  $R_{ij}$ . This guarantees us an  $n \times n$  matrix  $r_{ij}$  where  $n$  are the number of non-zero values in  $p(i, t)$ .

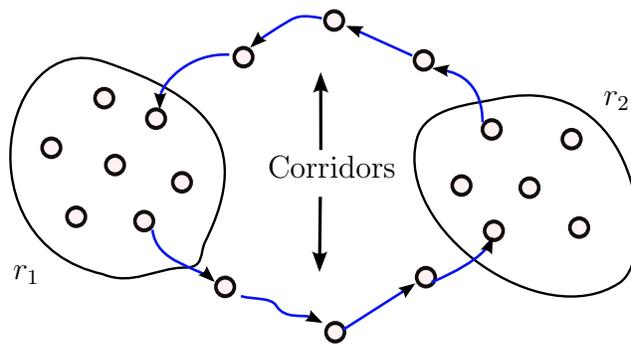


Figure 2.21: An illustration of our two cluster system. States in  $r_1$  and  $r_2$  can transition between each other freely while there is a small probability of transitioning along the corridors of states either to or from  $r_1$  and  $r_2$ . The number of states above are just for illustration purposes with exact numbers differing from the actual calculations.

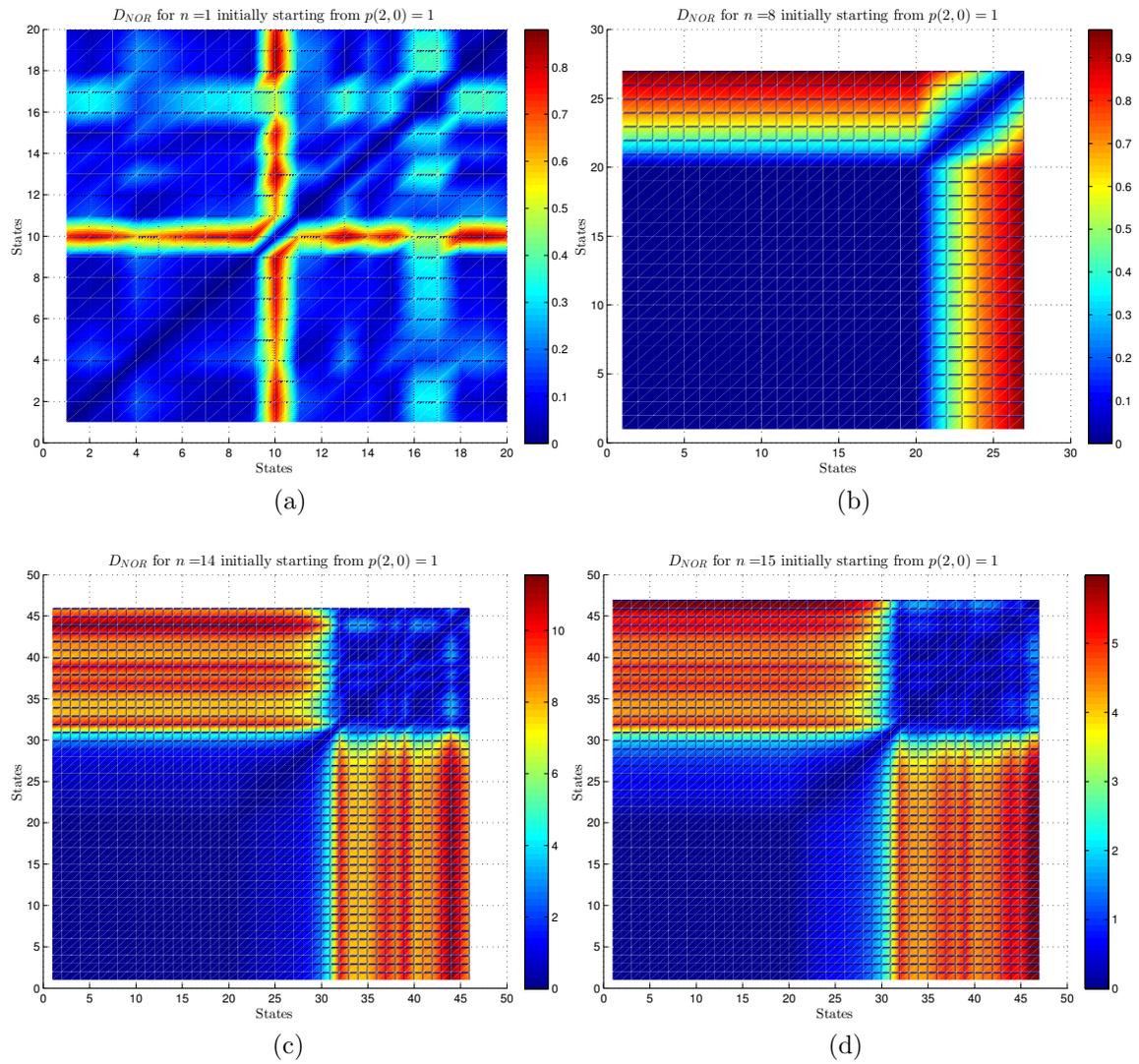


Figure 2.22: The NOR surface from  $B$  as the system evolves from a delta function at  $p(2,0) = 1$ .

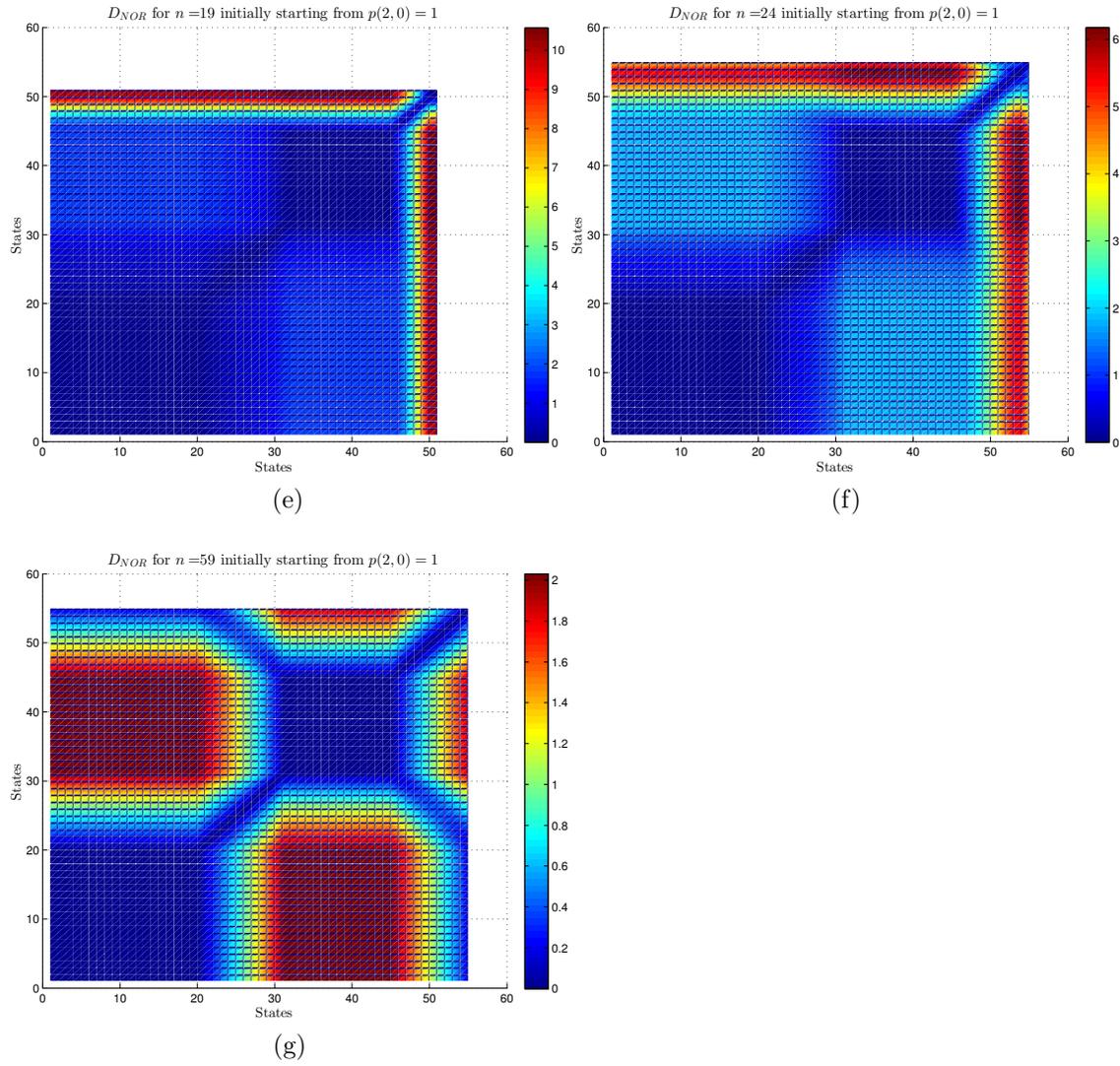


Figure 2.22: *Continued*: The NOR surface from  $B$  as the system evolves from a delta function at  $p(2,0) = 1$ .

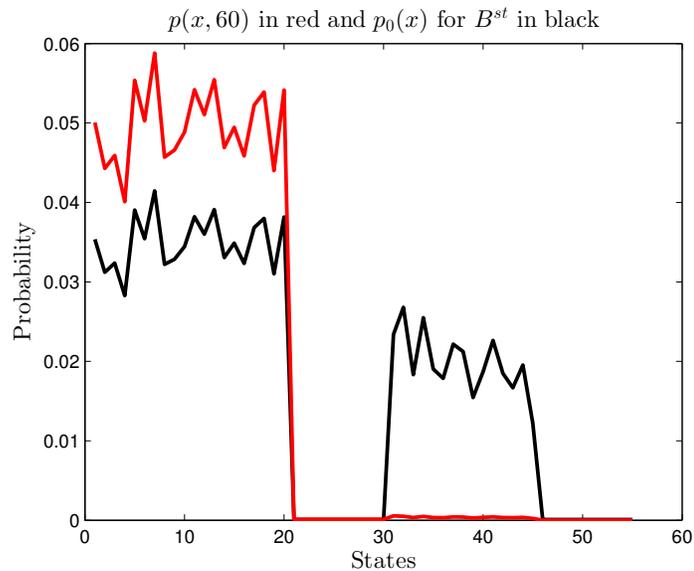


Figure 2.23:  $p_0(i)$  for  $B^{st}$  in black with  $p(i, 60)$  in red. Though the NOR doesn't change drastically on each time step, both distributions are still very far apart.

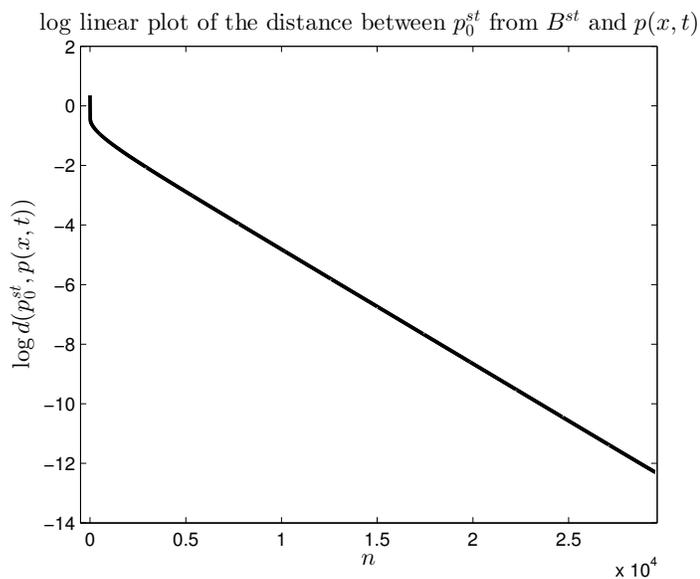


Figure 2.24: log linear plot of the distance between  $p_0$  from  $B$  and  $p(i, t)$ . The linear slope is  $\gamma = -3.83 \times 10^{-4}$

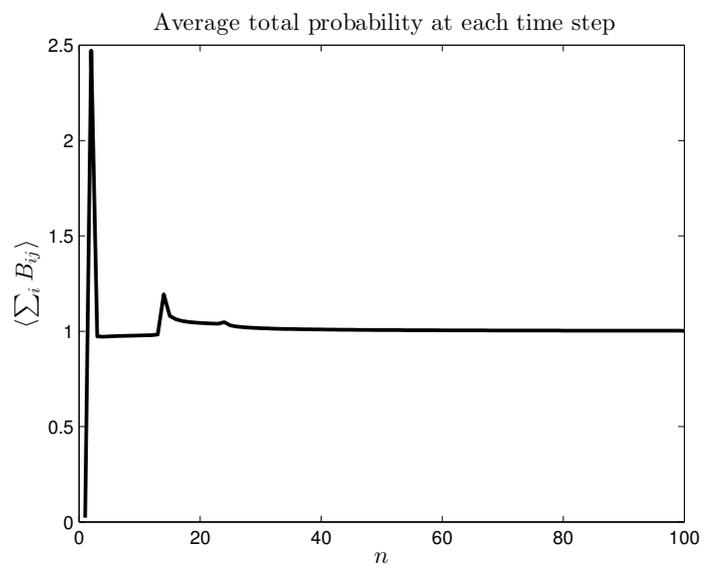


Figure 2.25: The average total probability as a function of time. We see that average total probability greatly exceeds one twice in the evolution. The average total probability approaches one in its long time evolution.

# Chapter 3

## Statistical Mechanics Through Distances

### 3.1 Introduction

The field of statistical mechanics has used abstract spaces as mainstays in the theory since its inception with the use of phase spaces. The phase space allows the immense and often infinite amount of information in a system to be organized in a meaningful way. The beauty of equilibrium mechanics is that the infinite number of configurations of the particles say in a room of gas, take on the simple Boltzmann Gibbs distribution Eq. (1.1). Once we leave equilibrium this is not longer guaranteed, and it is still an open question in general what distributions the system will take in time. We still do not even know if there is an equivalently simple solution, as there is in equilibrium with the Boltzmann distribution. The thermodynamic length, and in particular its non-equilibrium generalization, the information length, aims to provide an answer for this. Instead of focusing on the individual probability density functions of a system, they seek to instead take a step back and look at the distances between PDFs. This shift from measuring specific quantities to relations between quantities is very similar to the the previous Chapter on the Observable Representation

where the dynamics of the system were the focus. Here we will measure the marginal distributions of a system  $p(x, t')$  instead of the conditional distributions  $R(x, t'|y, t)$  as we did with the matrix of transitional probabilities.

$$p(x, t') = \int_{\Omega} R(x, t'|y, t)p(y, t)dy, \quad t' = t + dt \quad (3.1)$$

This approach however, will be shown to not only be able to reproduce fundamental relations in equilibrium statistical mechanics such as work but also provide a mathematical foundation which will allow us to move arbitrarily far away from equilibrium. Upon moving away from equilibrium we will find at least one group of systems, namely music and sound, that follow an elegant behaviour. Thus at least in a sense giving us comfort that simple relations may still be found arbitrarily far from equilibrium.

We will start by introducing the thermodynamic length in chronological order which in turn will mean beginning with equilibrium statistical mechanics. Next we will show how the Fisher information can be used as the metric over a set of generalized coordinates. In this interpretation the thermodynamic length is the distance between discernible states of our system.

## 3.2 Thermodynamic length

Since the time of Gibbs and his now coined Gibbs spaces [60], abstract spaces and geometric measures in them have been of interest. Uniqueness of these measures has always provided an issue though. For instance taking one norm between a quantity may give similar results to taking another norm of the same quantity. So which is the “correct” measure? Or is there a correct measure? In 1970 Weinhold showed how equilibrium statistical mechanics can be represented over a space of equilibrium states, complete with an inner product structure [1]. A vector space with an inner product is defined [27]: Given three vectors  $x, y, z \in \mathbb{R}$ ,

1.  $\langle x|y\rangle = \langle y|x\rangle$ .
2.  $\langle \alpha x + \beta x|z\rangle = \alpha\langle x|y\rangle + \beta\langle x|y\rangle$ , where  $\alpha$  and  $\beta$  are scalars.
3.  $\langle x|x\rangle \geq 0$  and equality only if  $x = 0$ .

Applying these definitions to the vectors  $X_i$  we can identify the vectors,

$$dV_i \leftrightarrow |\mathcal{V}_i\rangle,$$

$$\langle \mathcal{V}_i|\mathcal{V}_j\rangle = \frac{\partial V_i}{\partial X_j} = \frac{\partial^2 U}{\partial X_j \partial X_i}.$$

$dV_i$  is the differential of the conjugate variables of the system  $X_i$  and  $U(X_i)$  is the internal energy of the system which depends only on the extensive variables of the system. The last equation in the second line is simply re-writing the inner product, as the more traditional form of Weinhold's metric. To calculate the distances between states we define the thermodynamic length Using Weinhold's metric as,

$$\mathcal{L}_{th}^u = \int_0^\tau dt \sqrt{\frac{dX^i}{dt} g_{ij} \frac{dX^j}{dt}} = \int_0^\tau d\mathcal{L}_{th}. \quad (3.2)$$

$dX^j/dt$  is the change in the  $i$ -th extensive variable. The  $u$  superscript is to designate this length using Weinhold's metric. We shall define the thermodynamic length using multiple metrics and show how they are related to each other. Each  $g_{ij}$  gives the manifold of states its meaning of distance. Among other choices is Rupeiner's metric, who showed that we can define the same structure for an equilibrium system using,  $g_{ij} = \partial^2 S / \partial X_i \partial X_j$  [61]. Both metrics were shown by P. Salamon et al. [62] to be equivalent for infinitesimal changes in  $X_i$  up to a factor of the temperature  $T$ . Using Rupeiner's metric, we can relate the dissipation a system undergoes for a  $N$ -step quasistatic process [56], meaning the system evolves through finite time steps which are long enough that the system equilibrates after each time step. The distance the system travels under

one time step is related through,

$$\Delta\mathcal{L}^s = \frac{1}{\sqrt{2}} \sqrt{\Delta X^i \frac{\partial^2 S}{\partial X_i \partial X_j} \Delta X^j}, \quad (3.3)$$

where  $\Delta X_i$  is the change in  $X_i$ . The dissipation is identified as,  $\Delta S_t = (\Delta\mathcal{L}^s)^2/2 = \Delta\mathcal{J}/2$ , meaning that through the Cauchy Schwarz inequality, the dissipation and distance a system travels is given by,

$$\Delta S = \frac{1}{2} \sum_{t=1}^N \Delta\mathcal{J}_{th}^s \geq \frac{1}{2N} \sum_{t=1}^N (\Delta\mathcal{L}_{th}^s)^2 = \frac{1}{2N} (\mathcal{L}_{th}^s)^2. \quad (3.4)$$

$\mathcal{L}_{th}^s$  is the discrete version of Eq. (3.2) and the inequality between  $\mathcal{J}_{th}$  and  $\mathcal{L}_{th}$  will play a fundamental role in the following work. The minimum dissipation in Eq. (3.4) is given when each contribution to  $\mathcal{L}_{th}^s$  is constant. If the system is in equilibrium then Eq. (3.4) holds for the continuous case as well.

We should note that when we talk about dissipation and lengths over equilibrium states, we are assuming two conditions from Sect. (1.2) hold:

1. The PDF of the system is that which maximizes  $S = - \int_{\Omega} p(x) \log p(x) dx$ .
2. Subject to the constraints,  $\int_{\Omega} p(x) X_i(x) dx = \langle X_i \rangle$ .

These two conditions lead to  $p_0(x) = \frac{e^{-\lambda^i X_i}}{\mathcal{Z}}$ , through maximizing the corresponding lagrangian. Both conditions are known in equilibrium but must be assumed if the above equilibrium results hold when  $p(x, t) \neq p_0(x)$ . If both conditions hold for a general system then in the following section, we will see that there is a connection between the microscopic dynamics, i.e. the probability density functions to the macroscopic dissipation of the system.

### 3.3 Information length

A third choice of metrics gives us  $\mathcal{L}$ . This is a distance over a space of probability distributions or a statistical space as we will refer to it. Each “state” is now a probability distribution and as the system evolves it traces out a path  $\gamma$  through space. So that  $\mathcal{L}$  measures the distance between probability distributions over the path  $\gamma$ . To visualize this, see Fig. (3.1) where we have plotted a typical trajectory over a sample, three dimensional statistical space meaning the PDF only has three outcomes  $p = \{p_1, p_2, p_3\}$ . This distribution is only constrained by conservation of probability.  $\mathcal{L}$  is formed from a metric based on the Fisher information [63, 64, 65],

$$g_{ij} = \int_{\Omega} p(x, t) \frac{\partial \log p(x, t)}{\partial \lambda^i} \frac{\partial \log p(x, t)}{\partial \lambda^j} dx. \quad (3.5)$$

If we plug, Eq. (3.5) into Eq. (3.2) using the intensive variables  $\lambda^i$  instead of the extensive ones,  $X^i$  and sum over  $i$  and  $j$  we have,

$$\mathcal{L} = \int_0^{\tau} dt \sqrt{\int_{\Omega} \frac{1}{p(x, t)} \left( \frac{dp}{dt} \right)^2 dx}. \quad (3.6)$$

The advantage of Eq. (3.6) is that we are no longer confined to a manifold of equilibrium states. The statistical space can be defined for any system including outside of statistical mechanics. Indeed Wooteer defined  $\mathcal{L}$  over Hilbert space in quantum mechanics [59]. This distance we have already used in an equivalent form in Sect (2.4.1) where we measured the distance between two probability distributions as  $d(p(x, t_1), p(x, t_2))$  in Eq. (2.52). Therefore to differentiate between both situations, we shall refer to the distance over general states the information length, thus dropping the *th* subscript in Eq. (3.6).

Next we will demonstrate how the Fisher information and thus the information length is related to the relative entropy from Sect. (1.2), as was originally shown in [66, 67].

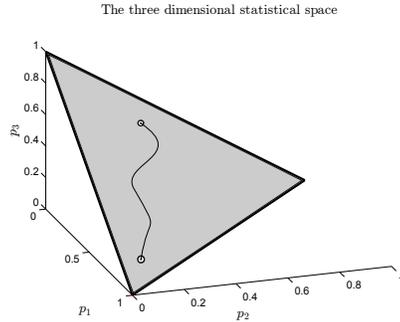


Figure 3.1: The three dimensional statistical space with conservation of probability. As the system evolves it traces out a trajectory  $\gamma$  through the space. Conservation of probability limits the space from a plane in  $\mathbb{R}^3$  to a triangle in  $\mathbb{R}^3$ .

Define two distributions,  $p_1 = p(x, t)$  and  $p_2 = p(x, t + \epsilon)$ . The relative entropy is defined as,

$$DS[p_1|p_2] = \int_{\Omega} p_1 \log \left( \frac{p_1}{p_2} \right) dx$$

The Taylor series to second order in  $\epsilon$  for  $DS[p_1|p_2]$  is given by,

$$DS[p_1|p_2] = DS[p_1|p_1] + \frac{dDS[p_1|p_2]}{d\epsilon} \epsilon + \frac{d^2DS[p_1|p_2]}{d\epsilon^2} \frac{\epsilon^2}{2} + \dots \quad (3.7)$$

Next we work out the first and second derivatives. The first derivative is,

$$\frac{dDS[p_1|p_2]}{d\epsilon} = - \int_{\Omega} \frac{p_1}{p_2} \frac{dp_2}{d\epsilon} dx. \quad (3.8)$$

But by expanding  $p_2 = p(x, t + \epsilon)$  in powers of  $\epsilon$ ,

$$p(x, t + \epsilon) = p(x, t) + \frac{\partial p_1}{\partial t} \epsilon + \frac{\partial^2 p_1}{\partial t^2} \frac{\epsilon^2}{2!} + \dots, \quad (3.9)$$

we have,

$$\begin{aligned}\frac{dDS[p_1|p_2]}{d\epsilon} &= - \int_{\Omega} \frac{p_1}{p_2} \left( \frac{\partial p_1}{\partial t} + \frac{\partial^2 p_1}{\partial t^2} \epsilon + \frac{\partial^3 p_1}{\partial t^3} \frac{\epsilon^2}{2} + \dots \right) dx. \\ &= - \int_{\Omega} \frac{p_1}{p_2} \frac{\partial p(x, t + \epsilon)}{\partial t} dx.\end{aligned}\quad (3.10)$$

The second derivative is,

$$\begin{aligned}\frac{d^2 DS[p_1|p_2]}{d\epsilon^2} &= - \int_{\Omega} \left[ \frac{d^2 p_2}{d\epsilon^2} \frac{p_1}{p_2} - \frac{p_1}{p_2^2} \left( \frac{dp_2}{d\epsilon} \right)^2 \right] dx, \\ &= \int_{\Omega} \left[ \frac{p_1}{p_2^2} \left( \frac{\partial p}{\partial t} \right)^2 - \frac{\partial^2 p}{\partial t^2} \frac{p_1}{p_2} \right] dx.\end{aligned}\quad (3.11)$$

Evaluating at the point  $\epsilon = 0$  we see that,

$$\frac{dDS[p_1|p_2]}{d\epsilon} \Big|_{\epsilon=0} = 0, \quad \frac{d^2 DS[p_1|p_2]}{d\epsilon^2} \Big|_{\epsilon=0} = \int_{\Omega} \frac{1}{p_1} \left( \frac{\partial p_1}{\partial t} \right)^2 dx = I_F. \quad (3.12)$$

The Fisher information  $I_F$  is then equal to  $\frac{d^2 DS[p_1, p_2]}{d\epsilon^2}$ . Thus we can write,

$$\mathcal{L} \approx \sum_{i=1}^N \sqrt{I_F(t_i)} \Delta t = \sum_{i=1}^N \sqrt{2DS[p(x, t_i)|p(x, t_i + \epsilon)]}. \quad (3.13)$$

if we take the limit as  $\Delta t \rightarrow 0$ , we have,

$$\mathcal{L} = \lim_{\Delta t \rightarrow 0} \sum_{i=1}^N \sqrt{I_F(t_i)} \Delta t = \int_0^T \sqrt{I_F(t)} dt = \sum_{i=1}^N \sqrt{2DS[p_1|p_2]}. \quad (3.14)$$

What is more interesting, is that this seems to imply  $DS[p_1|p_2]$  must sometimes be symmetric. This is due to  $I_F$  being a metric. But to be a metric, a quantity has to be symmetric, meaning in this case,  $DS[p_1|p_2] = DS[p_2|p_1]$ . The trouble with this is it is well cited that  $DS$  isn't symmetric [41], and thus does not constitute a true metric. Where have we gone wrong then? It turns

out, to second order  $DS[p_1|p_2] = DS[p_2|p_1]$ , meaning that the terms must only differ at higher orders.

To show this, following as before,

$$DS[p_2|p_1] = \int_{\Omega} p_2 \log\left(\frac{p_2}{p_1}\right) dx. \quad (3.15)$$

The first derivative with respect to  $\epsilon$  is,

$$\begin{aligned} \frac{dDS[p_2|p_1]}{d\epsilon} &= \int_{\Omega} \left[ \frac{dp_2}{d\epsilon} + \frac{dp_2}{d\epsilon} \log(p_2) - \frac{dp_2}{d\epsilon} \log(p_1) \right] dx, \\ &= \int_{\Omega} \left[ \frac{\partial p_1}{\partial t} + \frac{\partial p_1}{\partial t} \log(p_2) - \frac{\partial p_1}{\partial t} \log(p_1) \right] dx. \end{aligned} \quad (3.16)$$

The second derivative works out to be,

$$\frac{d^2 DS[p_2|p_1]}{d\epsilon^2} = \int_{\Omega} \left[ \frac{\partial^2 p_1}{\partial t^2} + \frac{1}{p_2} \left( \frac{\partial p_1}{\partial t} \right)^2 + \log(p_2) \frac{\partial^2 p_1}{\partial t^2} - \log(p_1) \frac{\partial^2 p_1}{\partial t^2} \right] dx. \quad (3.17)$$

If we evaluate our derivatives at  $\epsilon = 0$ , we see that,

$$\frac{dDS[p_2|p_1]}{d\epsilon} \Big|_{\epsilon=0} = 0, \quad \frac{d^2 DS[p_2|p_1]}{d\epsilon^2} \Big|_{\epsilon=0} = \int_{\Omega} \frac{1}{p_1} \left( \frac{\partial p_1}{\partial t} \right)^2 dx. \quad (3.18)$$

Meaning that up to second order,  $DS[p_2|p_1] = DS[p_1|p_2]$ . These relations can be used to illustrate how  $\mathcal{L}$  is related to  $\mathcal{L}_{th}^s$ . Each version of the thermodynamic length is equivalent for a system evolving to equilibrium, given also the path through probability space is one of constant velocity. The minimum value of  $\mathcal{L}$  can be seen using the substitution introduced by Wooter [59],  $p(x, t) = q(x, t)^2$ . Using this in Eq. (3.6) gives,

$$\mathcal{L} = 2 \int_0^T dt \sqrt{\sum_x \left( \frac{dq}{dt} \right)^2}. \quad (3.19)$$

This equation is the distance of a curve on the unit sphere and is well defined for any arbitrary distribution. The minimum is achieved when,

$$\sqrt{\sum_x \left(\frac{dq}{dt}\right)^2} = \text{const.} \quad (3.20)$$

Taking the Cauchy Schwarz inequality of Eq. (3.14) we can also see how the thermodynamic divergence enters into the picture,

$$\mathcal{J} \geq \frac{\mathcal{L}^2}{T} = 2 \left( \sum_{i=1}^N \sqrt{DS[p_1|p_2]} \right)^2 \quad (3.21)$$

Using the minimum condition in Eq. (3.20) implies  $DS[p_1|p_2]$  is constant and the integral only depends on the beginning and end points. [58] addresses the special case, when our system evolves to equilibrium, i.e. from  $p(x, 0)$  to  $p(x, t') = p_0(x)$ , then  $DS$  is the maximum available work  $DS = \Delta S - \Delta \lambda^i \langle X_i \rangle$ . Since the available work is the dissipation minus the change in extensive variables. Assuming the constant velocity assumption along with evolving to equilibrium we find,

$$\min \mathcal{J}_{th} = \Delta S - \lambda^i \langle X_i \rangle = \frac{\mathcal{L}^2}{T} = \text{const.}$$

Using the same conditions from Eq. (3.4) the minimum of  $\mathcal{J}_{th}^s$  is,

$$\min \mathcal{J}_{th}^s = \Delta S = \frac{(\mathcal{L}_{th}^s)^2}{T} = \text{const.}$$

Thus we conclude that the lengths are related by,  $(\mathcal{L}_{th}^s)^2 = (\mathcal{L})^2 + T \lambda^i \langle X_i \rangle$ . This explicitly shows a connection between the work a system does as it evolves to equilibrium and the distance it travels. For the systems examined later, even though conditions (1) and (2) are not guaranteed we see that they often hold.

### 3.3.1 Fisher information and intensive variables

There is still another connection in equilibrium for the thermodynamic length, namely through the free energy and the set of intensive variables  $\lambda^q$  [57]. These could be the pressure or temperature of a system. They are defined through the extensive variables,  $\lambda^q = \partial S / \partial X_q$ . The free energy  $\Psi$  is defined as,

$$\Psi = S - \lambda^i \langle X_i \rangle. \quad (3.22)$$

Taking the first partial with respect to  $\lambda^i$  gives,

$$\frac{\partial \Psi}{\partial \lambda^i} = - \int_{\Omega} \frac{\partial p(x, t)}{\partial \lambda^i} (1 + \log p(x, t) + \lambda^q X_q) dx - \lambda^q \int_{\Omega} p(x, t) \frac{\partial X_q}{\partial \lambda^i} dx - \langle X_i \rangle. \quad (3.23)$$

We notice that since we have maximized the Lagrangian which resulted from conditions (1) and (2), we have  $1 + \log p(x, t) + \lambda^q X_q = 0$ . We also see that due to independence  $\frac{\partial X_q}{\partial \lambda^i} = 0$  giving,

$$\frac{\partial \Psi}{\partial \lambda^i} = - \langle X_i \rangle. \quad (3.24)$$

Noting that  $p(x, t) = e^{-\lambda^q X_q} / \mathcal{Z}$ , the second partial derivative with respect to  $\lambda^j$  is,

$$\frac{\partial^2 \Psi}{\partial \lambda^j \partial \lambda^i} = \int_{\Omega} dx X_i \left[ \mathcal{Z}^{-1} X_j e^{-\lambda^q X_q} + \mathcal{Z}^{-2} e^{-\lambda^q X_q} \frac{\partial \mathcal{Z}}{\partial \lambda^j} \right]. \quad (3.25)$$

Calculating the partial of  $\mathcal{Z}$  and putting it all together we find,

$$\frac{\partial^2 \Psi}{\partial \lambda^j \partial \lambda^i} = \langle (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle) \rangle. \quad (3.26)$$

The free energy can be related to the covariance of the extensive variables of the system. This is a perfectly good metric, since the covariance is greater than or equal to zero, is symmetric and satisfies the triangle inequality. To show how this is related to the Fisher information, we simply plug the Boltzmann distribution  $p(x, t) = e^{\lambda^q X_q} / \mathcal{Z}$  into Eq. (3.5). The partial derivative with

respect to  $\lambda^i$  gives,

$$\begin{aligned} \frac{\partial \log p(x)}{\partial \lambda^i} &= \frac{\partial}{\partial \lambda^i} (-\lambda^q X_q - \log \mathcal{Z}), \\ &= \left( -X_i + \frac{1}{\mathcal{Z}} \int_{\Omega} X_i e^{-\lambda^q X_q} dx \right). \end{aligned} \quad (3.27)$$

We see that we recover the covariance of the extensive variables,

$$g_{ij} = \langle (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle) \rangle. \quad (3.28)$$

Thus in equilibrium the thermodynamic length based on the free energy and intensive variables is equivalent to using the Fisher information  $\mathcal{L}$ . In fact we have shown that essentially all of the different metrics are equivalent in equilibrium. The fact that it is the Fisher information that ties them all together to form a deep connection to the structure of statistical mechanics has led some to argue that it is the fundamental quantity when measuring the thermodynamic length or divergence [57]. Indeed it is the Fisher information we will use for the remainder of this work.

### 3.3.2 $\mathcal{L}$ for discrete systems

Since Weinhold proposed the thermodynamic length in 1974 [1] it has been extensively studied theoretically, work that continues to this day. But applications of the thermodynamic length have been scarce. This is due in part to being forced to work in discrete measures, which introduces a new set of problems not shared by their continuous counterparts. One such problem is that as a system evolves out of equilibrium it may not occupy all of its phase space meaning  $p(x, t) = 0$  while,  $p(x, t') \neq 0$ ,  $t' = t + \Delta t$ , which possibly leaves  $\mathcal{L}$  and

$\mathcal{J}$  undefined for much of its evolution,

$$\mathcal{L} = \sum_{i=1}^N \Delta t \sqrt{\sum_x \frac{1}{p(x,t)} \left(\frac{\Delta p}{\Delta t}\right)^2}, \quad (3.29)$$

$$\mathcal{J} = \sum_{i=1}^N \Delta t \sum_x \frac{1}{p(x,t)} \left(\frac{\Delta p}{\Delta t}\right)^2. \quad (3.30)$$

The reason this is not an issue for a continuous system is due to the substitution from Eq. (3.19),

$$\mathcal{L} = 2 \int_0^\tau \sqrt{\int_\Omega \left(\frac{dq}{dt}\right)^2 dx}.$$

If we have discrete time steps though, the information length based on  $q$  and  $p$  are not equivalent. This means we do not recover the equilibrium results for discrete systems. To recover the equilibrium results and rid our equations of un-physical infinite lengths we proposed in [68] a set theoretic approach. For two consecutive PDFs define the sets,

$$\begin{aligned} Q_p &= \{x : p(x,t) \neq 0 \mid p(x,t') = 0\}, \\ Q_w &= \{x : p(x,t) \neq 0 \mid p(x,t') \neq 0\}. \end{aligned} \quad (3.31)$$

The third possibility which is not included in the above equation is the case where  $p(x,t) = 0$  and  $p(x,t') \neq 0$ , which if we wish to not have infinite lengths must be excluded from  $\mathcal{L}$ . The subscript  $p$  in  $Q_p$  designates the unused probability of evolving over one time step, while  $Q_w$  is the set that gives a measure of the available work in evolving over one time step, as shown later. Using these sets we can re-write our definition for thermodynamic length as,

$$L_{Q_p} = \sum_{x \in Q_p} \frac{p(x,t)}{(\Delta t)^2},$$

$$L_{Q_w} = \sum_{x \in Q_w} \frac{1}{p(x, t)} \left( \frac{\Delta p(x, t)}{\Delta t} \right)^2,$$

and express Eq. (3.29) as:

$$\mathcal{L} = \sum_{t=1}^{\tau} \Delta t \sqrt{L_{Q_p} + L_{Q_w}}. \quad (3.32)$$

An immediate consequence of our sets is that if  $Q_w = \emptyset$  then  $\Delta \mathcal{L} = 1$ . To help illustrate our interpretation of  $Q_p$  and  $Q_w$  as the sets which deal with probability and work in the systems evolution, we will now show how  $L_{Q_w}$  is related to the discrete relative entropy and thus the available work of the system.

### 3.3.3 Work for discrete non-equilibrium systems

In the previous section we showed the link between  $\mathcal{L}$  and the relative entropy. This is of course already well known for many circumstances [69]. The links between the Fisher Information and the relative entropy are also known and was shown earlier, in [70, 66]. Here for completeness we briefly show how for the discrete case,  $L_{Q_w}$  is related to the relative entropy,

$$DS[p(x, t)|p(x, t')] = \sum_x \mathcal{DS}[p(x, t)|p(x, t')].$$

$\mathcal{DS}[p(x, t)|p(x, t')]$  is the microscopic relative entropy, or the local relative entropy, since it only pertains to individual states. Using  $p(x, t') = p(x, t) + \Delta p(x, t)$ , ( $\Delta p(x, t) = p(x, t') - p(x, t)$ ) in  $\mathcal{DS}$  as:

$$\begin{aligned} \mathcal{DS}[p(x, t)|p(x, t')] &= -p(x, t) \log \left[ 1 + \frac{1}{p(x, t)} \Delta p(x, t) \right] \\ &= -\Delta p(x, t), \end{aligned}$$

where  $\log(1 + x) \approx x$  was used above. Therefore, to leading order in  $\Delta p$ , the substitution of Eq. (3.33) into Eq. (3.32) gives us a new way to express the

thermodynamic length,

$$\mathcal{L}_{\mathcal{DS}} = \sum_t^\tau \Delta t \sqrt{\sum_{x \in Q_p} \frac{p(x, t)}{(\Delta t)^2} + \sum_{x \in Q_w} \frac{1}{p(x, t)} \left( \frac{\mathcal{DS}[p(x, t)|p(x, t')]}{\Delta t} \right)^2}, \quad (3.33)$$

To get a feel for the thermodynamic length we will next look at the logistic map as a case study. This will show us that the system follows the path of minimum dissipation for most of its evolution. Only in the transition from a PDF covering the entire state space but not yet the stationary distribution to the stationary distribution is there a larger dissipation.

### 3.4 The Logistic map and $\mathcal{L}$

The logistic map Eq. (1.39) exhibits much of the interesting properties of the thermodynamic length and thus makes a prime case to study in detail. It is also a perfect system to apply our set relations as it is non-differentiable in time making it particularly difficult to deal with traditionally. The domain  $X = [-1, 1]$  will again be broken up into  $M$  bins  $G_k = 2[k, k + 1]/M - 1$ ,  $k = 0, 1, 2, \dots, M$  where  $G_k \cap G_j = \emptyset$  and  $\cup_k G_k = X$ . In the chaotic regime, ( $a = 2$ ) it is well known that almost every orbit evolves to follow the stationary distribution,  $p_0 = 1/\pi(1 - x^2)^{1/2}$ .

We evolve an ensemble of orbits initially approximating a delta function and at each time step we will calculate the probability of  $x_t \in G_k$  given by  $p(k, t)$ . The PDFs will of course be subject to conservation of probability,  $\sum_k p(k, t) = 1$ . The main questions of interest are how does an initial distribution approach  $p_0$  in probability space? Also how does the approximation of Eq. (3.33) compare with Eq. (3.29) for the logistic map? Finally, what is the most efficient path to the stationary distribution? That is, is there a set of initial conditions which have the shortest distance in probability space? This would pertain to a system where there is an expense associated with the length of time the system evolves.

Likewise one could ask the opposite, are there a set of initial conditions which prolong the approach to the stationary distribution?

As a first example we evolve  $M = 9 \times 10^7$  orbits initially centred around  $x = -0.553$ . We see in Fig. (3.2) that for  $0 \leq t \leq 12$ ,  $\mathcal{L}$  on each time step has a constant slope of  $\Delta\mathcal{L} = 1$  meaning the PDFs do not overlap on each time step. From  $12 < t \leq 16$  the PDFs on each time step overlap with one another decreasing the slope of  $\mathcal{L}$ , but each PDF individually does not cover the entire domain  $X$ . The slope of  $\mathcal{L}$  has decreased from one to  $\Delta\mathcal{L} = 0.41453$ . This phase can be thought of as extremely rapid non-equilibrium evolution since the PDF changes by a large amount on each time step. For  $16 < t \leq 19$  the PDFs cover the entire domain and thus change less on each time step, but has still not reached the stationary distribution. Finally for  $t > 19$  the system has essentially reached  $p_0$  and  $\Delta\mathcal{L} \approx 0$ .

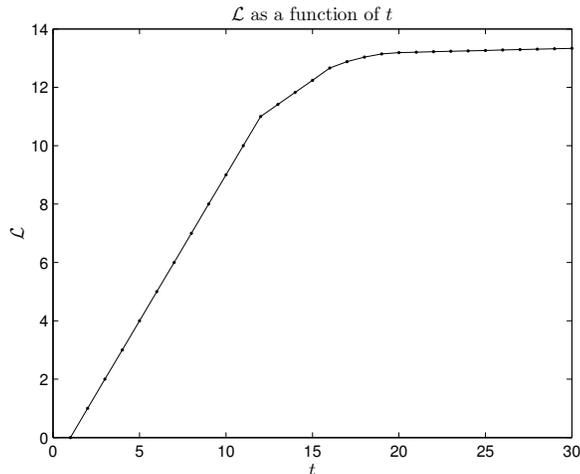


Figure 3.2: The evolution of  $M = 9 \times 10^7$  initial points centred around  $x_0 = -0.553$ . See text for detailed explanation.

Using the same ensemble we can test the agreement between Eq. (3.33) and Eq. (3.29). Fig. (3.3) shows  $\Delta\mathcal{L}$  as a function of  $t$  with the black dots and  $\Delta\mathcal{L}_{\mathcal{D}\mathcal{S}}$  shown with open circles. The independent evolution is even more clear in

this figure for  $0 \leq t \leq 12$ . For  $12 < t \leq 16$  since the PDFs are undergoing rapid change, our naive first order approximation to form  $\mathcal{L}_{DS}$  breaks down. Here one should attempt to incorporate higher order terms [71]. As the PDFs fill out the entire state space and slow their evolution we recover a good agreement between  $\Delta\mathcal{L}$  and  $\Delta\mathcal{L}_{DS}$ . Of course both distributions go to zero as the system becomes stationary.

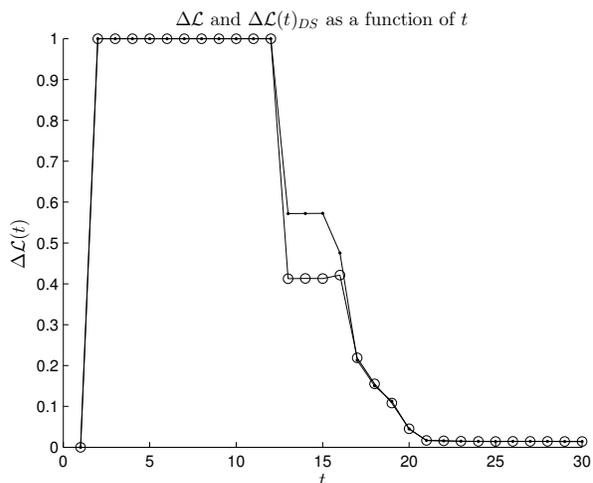


Figure 3.3: Plot of the discrete version of  $\mathcal{L}$  equation (3.32) against time in black which shows a good agreement with equation (3.33) plotted in black with circles. Both use  $M = 9 \times 10^7$  initial points who all start as a delta function around,  $x_o = -0.533$ .

Although the unstable fixed points at  $x = -1$  and  $x = 0.5$  are of measure zero, the discretization of the domain means their influence can create the appearance of fixed points. That is, an orbit may land very near a fixed point and then on the next time step due to their proximity, they again land in the same bin creating the appearance of a fixed point. This is shown in Fig. (3.4) where the absolute distance between  $x_t$  and  $x_{t+1}$  is plotted for the logistic map. The orbits that land in grains containing fixed points have their velocity through state space slowed. It is this slowing that allows the PDFs to overlap and the available work to decrease. The strongest example of this slowing can be seen

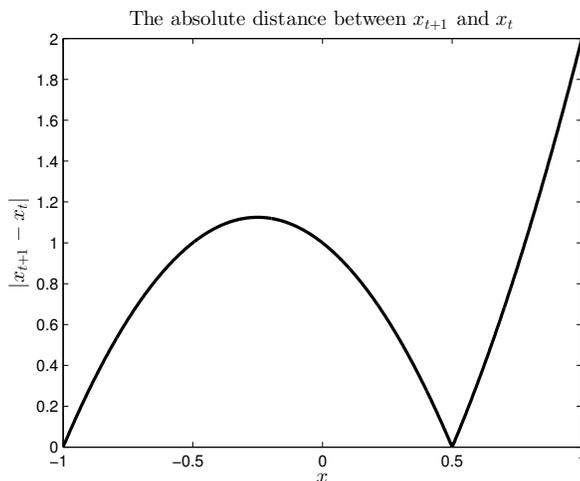


Figure 3.4: The distance between  $x_t$  and  $x_{t+1}$  for the logistic map. We see how the minimums around the two fixed points may lead orbits landing near to them to again land in the same bin on the next time step, thus creating the appearance of a fixed point.

in Fig. (3.5). Here an ensemble of orbits are initially centred at  $x_0 = 0.7071$ . For the first few iterations all orbits only occupy  $Q_p$ . For  $4 \leq t \leq 7$  the entire ensemble evolves into the first bin around  $x = -1$ . All orbits are then trapped, giving the appearance of a fixed point. For  $7 < t \leq 16$  some of the orbits escape the first bin and this gives an overlap in PDFs. This is illustrated in Fig. (3.6) where a typical example of the overlap of two PDFs as one evolves away from the fixed point.  $p(k, 13)$  and  $p(k, 14)$  both occupy  $Q_w$  from approximately  $-1 \leq x < -0.75$ . The region  $-0.75 < x$  is only counted on the next time step,  $t = 15$ . We will see shortly that if  $\Delta\mathcal{L} = 0$  when  $p(k, t) \neq p_0$  then the operator that would be generated from this evolution is reducible. From  $t > 16$  the system has settled into equilibrium.

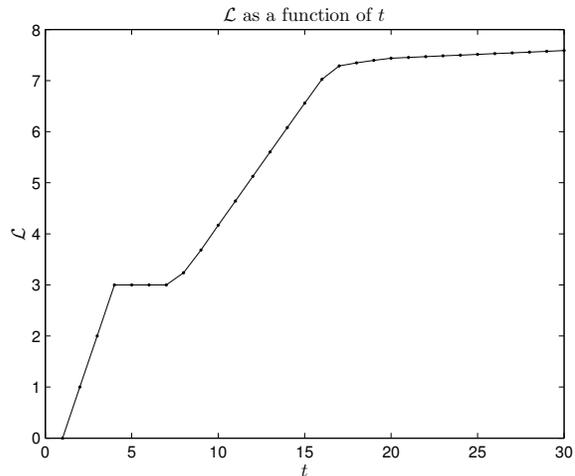


Figure 3.5: The evolution of  $\mathcal{L}$  starting from  $x_0 = 0.7071$ . The evolution is divided up into four main phases.  $0 < t \leq 4$ , all  $x \in Q_p$ ,  $4 < t \leq 7$  all orbits are in the bin that holds the  $x = -1$  fixed point, though the operator that would be made from the orbits is reducible.  $7 < t \leq 16$ ,  $\Delta\mathcal{L}(t) < 1$  as the PDFs overlap and the information changes.  $t > 16$  the system settles into  $p_0(x)$ .

### 3.4.1 Unstable fixed points and distance to reach $p_0$

Here we will look at how the initial conditions for the logistic map influence the total distance to reach the stationary distribution. Or put another way, are there initial conditions which either quickly or slowly come to be stationary, i.e. have the shortest distance to  $p_0$ ? The answer was hinted at in Fig. (3.5) where we showed that by leaving the vicinity of a fixed point, two consecutive PDFs occupy  $Q_w$ .  $Q_w$  then essentially turns information into wasted heat thus lowering the available work through Eq. (3.33). By starting each ensemble as a delta function we are giving the system the most initial information possible. It then loses this information as it evolves to  $p_0$ . To see how the fixed points influence this loss of information, we uniformly spread ensembles of initial conditions across  $[-1, 1]$ . Plotting the initial position as a function of the total distance it takes to reach  $p_0$ , we see in Fig. (3.7) that there are a small subset of initial

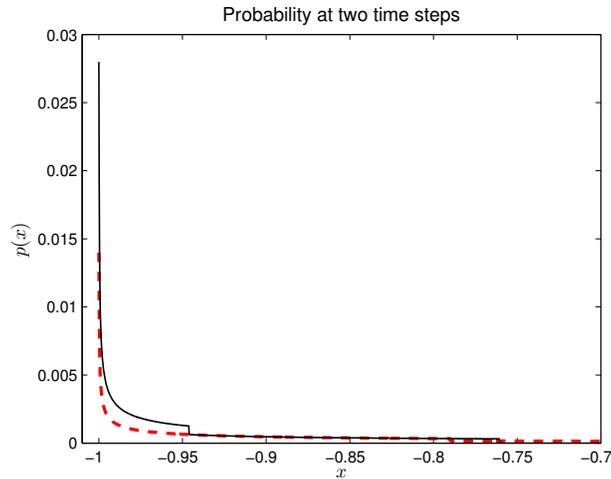


Figure 3.6:  $P(k, 13)$  plotted in black and  $p(k, 14)$  is plotted in red with the dashed line.

conditions which reach the stationary distribution in less distance. The points  $x_0 = [-1, -0.96, -0.708, -0.5, 0, 0.5, 0.708, 0.96, 1]$  are highlighted with circles, each of these initial points reach a bin containing a fixed point in five iterations or less. This shows that it is the unstable fixed points which are most efficiently driving the system to its stationary distribution. An application of this result is if one has a system with an associated cost to reaching its stationary distribution, then by starting the system in an initial condition which will reach an unstable fixed point quickly, the cost of reaching the final distribution will be minimized.

### 3.4.2 Conditions for $\Delta\mathcal{L} = 0$

We saw in Fig. (3.5) that early in the evolution  $\Delta\mathcal{L} = 0$ . Here we will prove that given the system is evolved by a PF operator such that,  $p(x, t') = R_{xy}p(y, t)$ , under what conditions can  $\Delta\mathcal{L} = 0$ . It is easy to see that the lower bound on  $\Delta\mathcal{L}(t) = 0$  occurs when  $p$  is stationary (i.e.  $p(x, t) = p(x, t')$ ). One may ask,

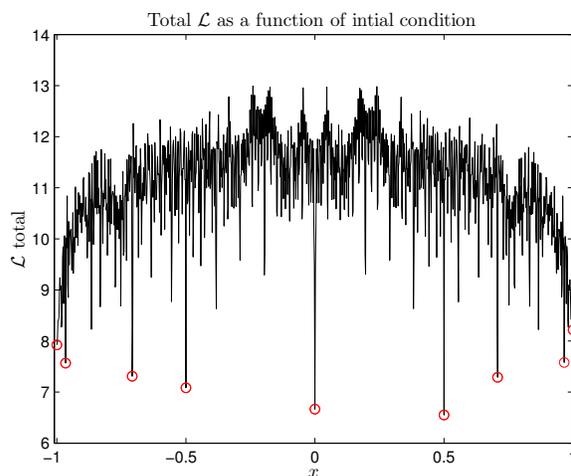


Figure 3.7: The evolution of  $\mathcal{L}$  as a function of time for many initial conditions spread over the domain. Most initial conditions travel a distance of between 13 and 16 before reaching  $p_0(x)$ . The points  $x_0 = [-1, -0.96, -0.708, -0.5, 0, 0.5, 0.708, 0.96, 1]$  whose initial conditions are marked with circles, start at or quickly occupy the bin of a fixed point and thus reach  $p_0(x)$  in a far shorter distance.

is being stationary the only condition for  $\Delta\mathcal{L}(t) = 0$  regardless of whether a system is in equilibrium or out of equilibrium? To answer this question, we utilise the current of probability which flows from state  $y \rightarrow x$  in one time step, which was originally defined for an equilibrium system without summation in Eq. (2.18) as,

$$J_{xy}^{st} = 0 = R_{xy}p_0(y) = R_{yx}p_0(x).$$

Here,  $R_{xy}$  is defined as the non-negative irreducible matrix of transition probabilities, Sect. (2.3), from states  $y$  to  $x$ ,

$$R_{xy} = Pr(\text{state at } (t' > t) \text{ is } x | \text{state at } t \text{ is } y). \quad (3.34)$$

We recall: the distribution  $p_0(x)$  is guaranteed to be a unique stationary distribution of  $R_{xy}$  due to  $R_{xy}$  being irreducible [26].  $\sum_y J_{xy}^{st} = 0$  is guaranteed since  $J_{xy}^{st}$  follows Kirchoff's loop rule that the amount of current into a state is equal to the amount out of a state. We define stationary as the PDF being time independent. The system can be characterised as being reversible or not through  $J_{xy}^{st} = 0$ , or  $J_{xy}^{st} \neq 0$  respectively. In general, we can define a non-equilibrium current  $J_{xy}$  as,

$$J_{xy} = R_{xy}p_t(y) - R_{yx}p_t(x). \quad (3.35)$$

Summing Eq. 3.35 over  $y$  gives,

$$\sum_y J_{xy} = p_{t'}(x) - p_t(x) = \Delta p(x, t). \quad (3.36)$$

This allows us to link the operator  $R_{xy}$  to  $\mathcal{L}$ . Obviously, when  $\sum_y J_{xy} = 0$ ,  $p(x, t') = p(x, t)$ , i.e. the distribution is stationary with  $\Delta\mathcal{L}(t) = 0$ .  $\Delta\mathcal{L}(t) = 0$  is also guaranteed under the stricter condition of detailed balance which defines true equilibrium, i.e. when  $J_{xy} = 0 \forall x, y$  in Eqs. (2.18) or (3.35). Therefore, in view of the uniqueness of  $p_0(x)$ , we can infer that if  $\sum_y J_{xy} = 0 \forall x$  then  $J_{xy} = J_{xy}^{st}$  and  $\Delta\mathcal{L}(t) = 0$ , meaning that the system is stationary. This shows that for any non-detailed balance system, irreducibility is necessary for stationarity to

uniquely imply  $\Delta\mathcal{L} = 0$ . If the system does not have an irreducible operator, then it is possible that  $\Delta\mathcal{L}(t) = 0$  in general, as we saw in Fig. (3.5). This is an interesting result because irreducibility is a global quality of the system. While the intersection of the PDFs is essentially a local quantity, only being affected by the areas of phase space with non-zero values of the PDFs. What we have shown is that being connected across the phase space (irreducibility) puts limitations of the evolution of the system at each time step, in that  $\Delta\mathcal{L} \neq 0$ .

### 3.5 Music through $\mathcal{L}$

Thermodynamics was originally developed to understand the question “how can one best use heating processes to exert forces and to do work” [12]. As a result it developed from empirical observations which were then turned into mathematical relations. Subsequently statistical mechanics was developed and so often dealt with many of the same quantities, such as energy, volume, chemical potential etc. Yet since statistical mechanics is the reverse approach of thermodynamics, in that it is fundamentally a set of mathematical rules which are justified through observation it is not limited to the canonical set of thermodynamic variables and systems. In this section we will use this generality to our advantage and apply the information length and divergence to the study of music.

The key idea is to envision music as a flow of information and to compute its variation from the temporal change in the probability distribution function constructed from a midi-file of the music as it is played. The rate at which information varies is then captured by the velocity in a statistical space where time serves as a parameter. The total distance travelled in this statistical space represents the total accumulative change of information in time and is quantified by the information length  $\mathcal{L}$ . Similarly, the action  $\mathcal{J}$  of the music is computed from the time integral of the energy of the music by using the square of the velocity as kinetic energy. The inequality first shown in Eq. (3.4) for

quasistatic evolutions will be given an interpretation for an arbitrary evolution as the total energy used in statistical space minus the total distance travelled. Given the infinite number of evolutions possible for each musical composition, some comfort will be found in that each piece will follow to a good approximation a simple power law, thereby reclaiming some of the simplicity found in equilibrium statistical mechanics for a non-equilibrium system.

Music plays an intricate part of human life. As a result there is a large body of work devoted to the analysis of music. Going back to the Greeks, “Pythagoras was the first to discover the fundamental connection between mathematics and music” [72]. Since then countless works have been published revealing the structure of music through mathematical language, e.g. see [73, 74, 75, 76]. Of particular interest here are the various power laws that have been found in different measures of music, though not from music itself. This was illustrated by Voss and Clarke [77] who looked at the output voltage of sound recordings and found that for example, the loudness of music and speech follows a power law but the voltage, (time signal itself) does not. The continuous signal from music utilised by Voss and Clarke along with Serrà et.al [78] and digitized music by [79, 80, 81] also demonstrate power law relations in different aspects of music. As it is impossible to encompass the complexity and delicacy of music by any one measure, each approach inevitably has its own advantages and drawbacks in comprehending music. In particular, analysing music via the amplitudes of a continuous signal ignores the exact notes being played, mainly being concerned with the sound created by the performer, which varies from performance to performance. In comparison, digitized music such as midi files has the precision to exactly reproduce the same piece of music each time, since every note in a piece of music is assigned a number that a computer uses to make an exact recreation of the sheet music for a given composition. Consequently, although midi-files contain detailed information about the composition, they sounds synthetic due to the lack of complexity and variation that a human performer brings to a musical performance.

### 3.6 Information variation ( $\mathcal{L}$ and $\mathcal{J}$ )

In this approach to the information length the key physical quantity is the temporal variation in a PDF of the state  $x$ , i.e.,  $p(x, t)$ . Due to the conservation of probability in time, the integral of  $\frac{dp(x,t)}{dt}$  over all states vanishes, i.e.

$$\int dx \frac{dp(x, t)}{dt} = 0. \quad (3.37)$$

We thus quantify the variation of the PDF by using its second moment of  $\frac{dp(x,t)}{dt}$  through the fluctuating energy  $\mathcal{E} = I_F^2$  defined as <sup>1</sup>

$$\mathcal{E}(t) = I_F(t)^2 = \int dx \frac{1}{p(x, t)} \left( \frac{dp(x, t)}{dt} \right)^2. \quad (3.40)$$

$I_F(t)$  in Eq. (3.40) physically represents the effective velocity at which the information varies at time  $t$  while  $\mathcal{E}(t) = I_F(t)^2$  is the associated energy given by the square of this velocity. For the analysis of the evolution of  $p(x, t)$  out of equilibrium where  $p(x, t) = 0$  for some  $x$  and  $t$  (as the system may have explored only a small portion of its state space), we use an alternative form of Eq. (3.40) introduced in Eq. (3.19),

$$\mathcal{E} = I_F^2 = 4 \int dx \left( \frac{dq(x, t)}{dt} \right)^2, \quad (3.41)$$

---

<sup>1</sup>As seen in the previous section, where control parameters  $\lambda_i$ 's ( $i = 1, 2, 3, \dots$ ) of a system are known as a function of time (e.g. in equilibrium), Eq. (3.40) can be recast by using the metric  $g_{ij}$  based on Fisher information (see, e.g. [63, 64, 65])

$$g_{ij} = \int dx p(x, t) \frac{\partial \log p(x, t)}{\partial \lambda^i} \frac{\partial \log p(x, t)}{\partial \lambda^j}, \quad (3.38)$$

as

$$\mathcal{E}(t) = I_F(t)^2 = \sum_{i,j} \frac{d\lambda^i}{dt} g_{ij} \frac{d\lambda^j}{dt}. \quad (3.39)$$

is the metric tensor that gives the Riemannian metric [82] in the parameter space  $\lambda$ 's. Since often the control parameters of a system are not known, it is much more convenient to use Eq. (3.40) directly in terms of PDFs.

which is mathematically well defined.

Thus  $\mathcal{E} = I_F^2$ , measures the total accumulated distance and energy between  $t = 0$  and  $t = T$  by the information length  $\mathcal{L}$  and action  $\mathcal{J}$ , respectively:

$$\mathcal{L} = \int_0^T dt \sqrt{\mathcal{E}(t)} = \int_0^T dt I_F(t), \quad (3.42)$$

$$\mathcal{J} = \int_0^T dt \mathcal{E}(t) = \int_0^T dt I_F(t)^2. \quad (3.43)$$

Eqs. (3.42) and (3.43) quantify the accumulative information variation and energy and are analogous to the relations for the distance and the action for a free particle with unit mass in classical mechanics.

To highlight that  $\mathcal{J}$  has a lower bound related to  $\mathcal{L}$ , we use  $u = 1$  in the following Cauchy-Schwartz inequality

$$\int_0^T I_F^2 dt \int_0^T u^2 dt \geq \left( \int_0^T I_F u dt \right)^2, \quad (3.44)$$

which gives  $\mathcal{J} > \mathcal{L}^2/T^2$ . In the case of constant  $I_F$ , the evolution of the system can be viewed as a ‘free’ motion.

To quantify the difference between  $\mathcal{J}$  and  $\mathcal{L}^2/T$ , it is useful to consider the time average of  $I_F$  and  $I_F^2$  as follows:

$$\langle I_F \rangle_T = \frac{1}{T} \int_0^T dt I_F, \quad \langle I_F^2 \rangle_T = \frac{1}{T} \int_0^T dt I_F^2. \quad (3.45)$$

Writing Eq. (3.42) and Eq. (3.43) in terms of Eq. (3.45), we obtain

$$\mathcal{J} - \frac{\mathcal{L}^2}{T} = T (\langle I_F^2 \rangle_T - \langle I_F \rangle_T^2). \quad (3.46)$$

This illustrates that the time averaged variance of the system is related to the

---

<sup>2</sup>The minimum value of  $\mathcal{J} - \mathcal{L}^2/T$  would be achieved for geodesics in statistical space, meaning  $\mathcal{J} = \mathcal{L}^2/T$  only when  $I_F$  is constant [56] in Eqs. (3.42) and (3.43)

distance it travels,

$$V_T = \langle I_F^2 \rangle_T - \langle I_F \rangle_T^2 = \frac{\mathcal{J}}{T} - \left( \frac{\mathcal{L}}{T} \right)^2. \quad (3.47)$$

We show in the following sections that Eq. 3.46 has the approximate power law  $\mathcal{J} - \mathcal{L}^2/T \propto T^{1+m}$ . Interestingly, Eq. (3.46) is the same result shown for a quasistatic process, Eq. (3.4) only now with equality. Both results are equivalent for  $V_T = 0$ . Unlike the quasistatic evolution our results (3.46)-(3.47) are also general and hold for any arbitrary distribution. This generality will be highlighted next where we calculate  $\mathcal{L}$ ,  $\mathcal{J}$  and  $\mathcal{J}/T - (\mathcal{L}/T)^2$  for famous classical music.

### 3.6.1 Music as a non-equilibrium system

In western music, the musical scale is typically divided into 11 octaves where each octave has a 2:1 relationship between its frequency and the octave below. Every octave is then made up of 12 semitones or half notes. Our key step towards understanding music is to envision a composition (such as Vivaldi's Concerto Summer) as a non-equilibrium system where each note represents a state  $x$  of the system. These are used to construct the probability over all instruments  $p(x, t)$  of a note being played in a coarse grained time interval,  $\Delta t$ . As a piece of music evolves, each instrument then transitions between states leading to information variation, while the simultaneous occupation of a set of states by all instrument creates the sound we hear.

In defining the state  $x$  of a note and  $p(x, t)$ , we utilise music *midi* files as they contain detailed information about the composition. Specifically, the *midi* file format for storing sheet music represents a piece of sheet of music as a series of numbers that are used by a computer in recreating a given composition.<sup>3</sup> Each note from octave 0 to octave 10 is given a *midi number*. Using a *midi* file our state space is then characterized by 129 states. States 0 to 127 correspond

---

<sup>3</sup>Each *midi* file used here is freely available at, [www.classicalmidiconnection.com](http://www.classicalmidiconnection.com).

to each possible note while the state 128 represents a rest (i.e. no note being played). These *midi* numbers and their corresponding notes are shown in Table 3.1.

Octave Number	Notes											
	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
0	0	1	2	3	4	5	6	7	8	9	10	11
1	12	13	14	15	16	17	18	19	20	21	22	23
2	24	25	26	27	28	29	30	31	32	33	34	35
3	36	37	38	39	40	41	42	43	44	45	46	47
4	48	49	50	51	52	53	54	55	56	57	58	59
5	60	61	62	63	64	65	66	67	68	69	70	71
6	72	73	74	75	76	77	78	79	80	81	82	83
7	84	85	86	87	88	89	90	91	92	93	94	95
8	96	97	98	99	100	101	102	103	104	105	106	107
9	108	109	110	111	112	113	114	115	116	117	118	119
10	120	121	122	123	124	125	126	127				

Table 3.1: Each *midi* number corresponds to an octave listed in the left column and a note, listed on the top row.

In order to construct the PDF of the state  $x$  for each note, we now examine time scales in the system and select a suitable time interval  $\Delta t$  for constructing  $p(x, t)$ . The “tick” is the time unit used by the *midi* format. One tick is equivalent to a certain number of milliseconds, specified in the header of each midi file. This allows one to know how many ticks are in a given piece of music. From the total number of ticks in the composition, we can divide the composition into probability distributions where  $p(x, t)$  is the probability of any note  $x$  being played between  $(i - 1)\Delta t \leq t \leq i\Delta t$ ,  $i = 1, 2, \dots, N$ . The choice of  $\Delta t$  is a free parameter and is the coarse graining scale of the system, which is selected to ensure a PDF of the highest quality, as further discussed below.

For our subsequent analysis of music, we use the discrete version of Eqs.

(3.42) and (3.43):

$$\mathcal{L} = 2 \sum_{i=2}^N \Delta t \sqrt{\sum_x \left(\frac{\Delta q}{\Delta t}\right)^2}, \quad (3.48)$$

$$\mathcal{J} = 4 \sum_{i=2}^N \Delta t \sum_x \left(\frac{\Delta q}{\Delta t}\right)^2. \quad (3.49)$$

Here,  $T = N\Delta t$  and  $\Delta t$  is now a discrete time step meaning  $\Delta q = q(x, i\Delta t) - q(x, (i-1)\Delta t)$ . If for example,  $\Delta t$  is chosen to be equal to the time per tick, then  $\mathcal{L}$  and  $\mathcal{J}$  from Eq. (3.48) and (3.49) would often be 0, since  $q(x, i\Delta t) = q(x, (i-1)\Delta t) \forall x$  for many time steps. In the opposite limit where  $\Delta t$  is chosen to be comparable to the length of the composition, all the structure in the music is completely lost due to the time average. There is also the issue that if we let  $\Delta t$  grow too large, the substitution of using  $q(x, t)$  instead of  $p(x, t)$  will give large discrepancies, since for large time steps these quantities differ. After testing different values of  $\Delta t$  between these two extreme limits, we identified  $\Delta t = 0.125 \text{ seconds}$  as an optimal time step, giving the best quality of the PDFs of several famous classical musics studied in this paper. For instance, we checked that there exists a robust power-law scaling of  $\mathcal{L}$  and  $\mathcal{J}$  with respect to time in Vivaldi's Summer for the value of  $\Delta t$  within the interval  $0.01 \text{ seconds} \leq \Delta t < 0.7625 \text{ seconds}$ . The range of power law validity was also checked for each composition analysed in this work, and from this analysis we selected  $\Delta t = 0.125 \text{ seconds}$  well inside the domain of all songs to ensure the best quality PDFs.

The music analysed below is Vivaldi's Summer, Beethoven's Ninth Symphony, 2nd movement, Mozart's Violin Concerto No. 3, and Tchaikovsky's 1812 Overture. These PDFs are of particular interest due to being strongly intermittent as can be seen in a typical example of  $p(x, t)$  in Fig. (3.8) which comes from Tchaikovsky's 1812 Overture. Such PDFs have no resemblance to most commonly studied examples such as Gaussian, Poisson etc. This makes

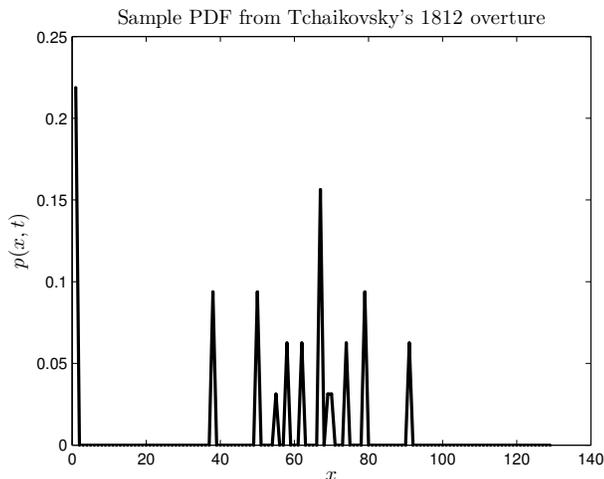


Figure 3.8: A sample PDF from Tchaikovsky’s 1812 Overture.

the power law relations below all the more interesting; that such order can come from apparent randomness.  $\mathcal{L}$  and  $\mathcal{J}$  are shown in Fig. (3.9) and Fig. (3.10) for all compositions.

### 3.6.2 Power-law scalings

Though these figures show the apparent linear nature of the functions, more detailed feature can be seen in Figs. (3.11) and (3.12) which show results for different compositions separately in log-log scales. Power-law indices of  $\mathcal{L}$  and  $\mathcal{J}$  are thus determined by linear fitting to these figures shown in dashed lines and are summarised in Table 3.2. The quality of a linear fit is measured using the standard R-squared value,

$$R^2 = 1 - \frac{\sum_i r_i^2}{(N - 1)Var(x_i)}, \quad (3.50)$$

where  $x_i$ ,  $i = 1, 2, \dots, N$  are the discrete data points and  $r_i = x - x_{fit}$  is the difference between the measurements  $x_i$  and the linear least squares fit  $x_{fit,i}$ . The denominator,  $Var(x_i)$  is the variance of the set  $x_i$ . A value of  $R^2 = 1$

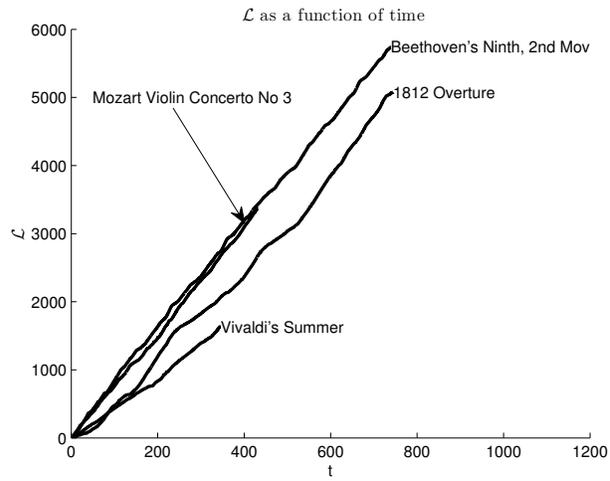


Figure 3.9:  $\mathcal{L}$  for each piece of music.

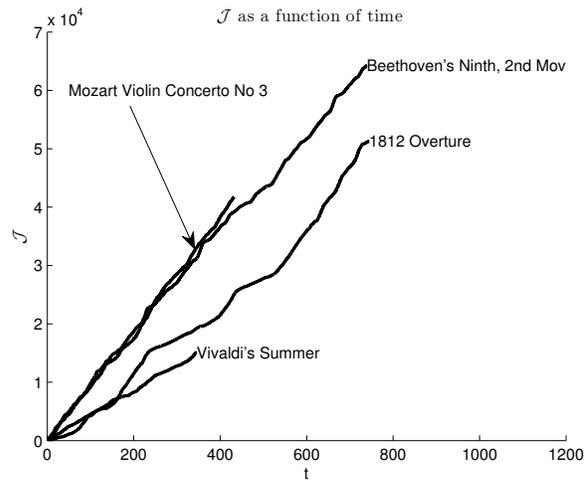


Figure 3.10:  $\mathcal{J}$  for each piece of music.

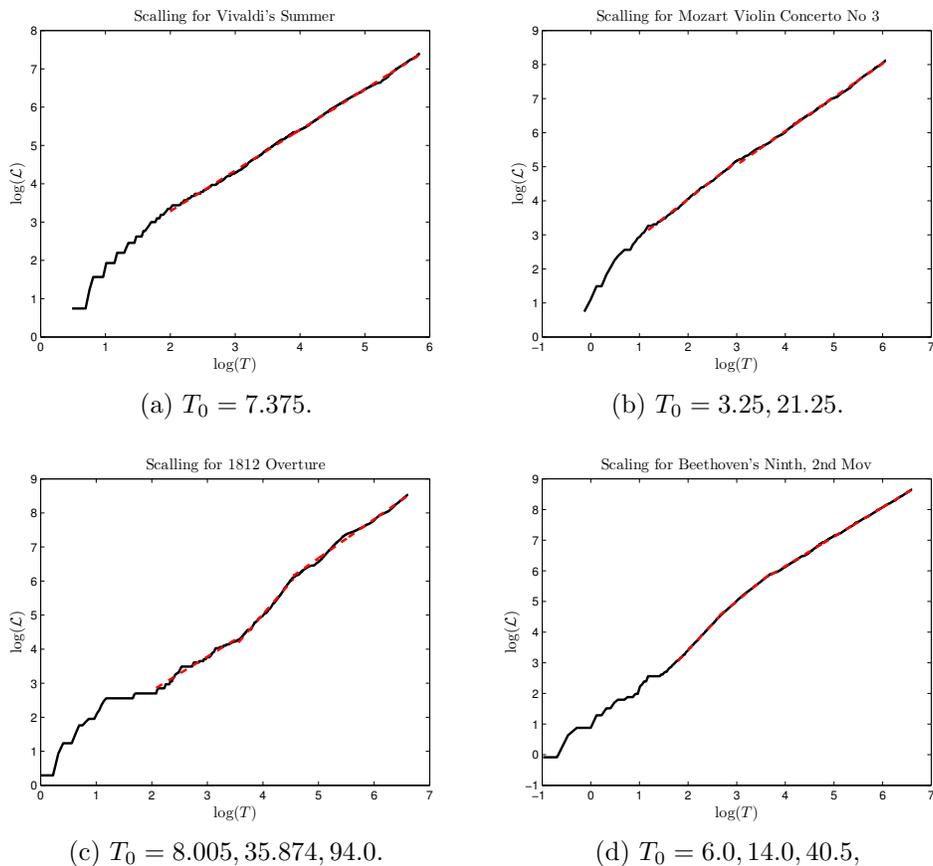
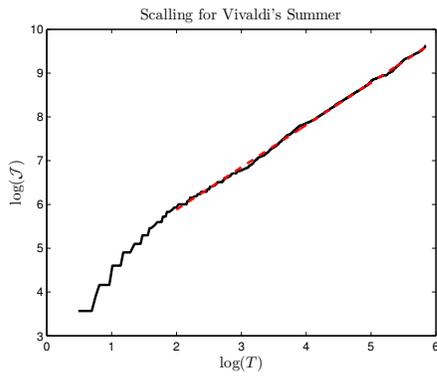


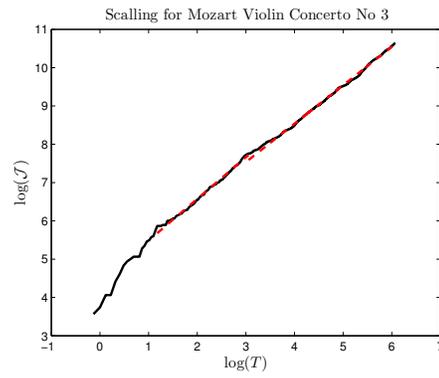
Figure 3.11:  $\log \mathcal{L}$  for each piece of music studied.

represents a perfect fitting. One should note the R-squared values in Tables 3.2-3.3 are all very close to one, meaning power laws are very good approximations. Note that all the compositions exhibit strong initial transient behaviour and the scaling is obtained for  $T > T_0$  where  $T_0$  is chosen to ensure a good scaling. The values of  $T_0$  for different music are shown in Fig. (3.11) and (3.12).

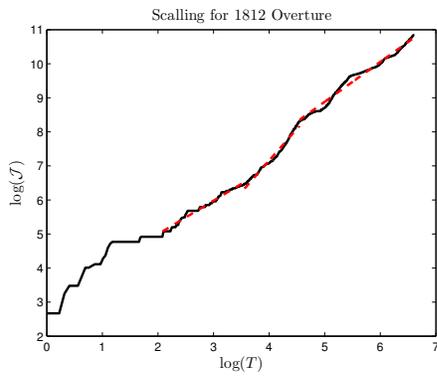
Interestingly, all compositions after an initial transient phase follow power law relations. Furthermore,  $\mathcal{L}$  and  $\mathcal{J}$  become linear in time for each composition barring Tchaikovsky, meaning that there is approximately a constant rate of information change, as the system evolves.



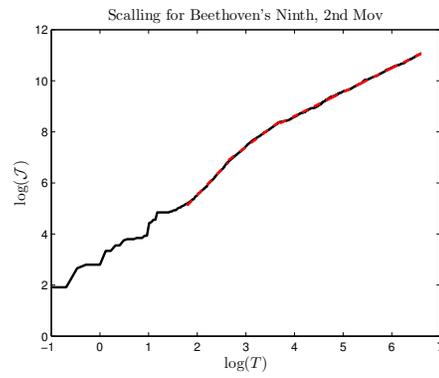
(a)  $T_0 = 1.625$ .



(b)  $T_0 = 3.25, 21.25$ .



(c)  $T_0 = 8.005, 35.874, 94.0$ .



(d)  $T_0 = 6.0, 14.0, 40.5, .$

Figure 3.12:  $\log \mathcal{J}$  for each piece of music studied.

	Mozart		Beethoven		Vivaldi		Tchaikovsky	
	slope	$R^2$	slope	$R^2$	slope	$R^2$	slope	$R^2$
$\log(\mathcal{L})$	1.106	0.9984	1.697	0.9995	1.062	0.9989	0.9969	0.9792
	0.9922	0.999	1.303	0.9984			1.888	0.9938
			0.9582	0.9997			1.141	0.9930
$\log(\mathcal{J})$	1.0867	0.9944	1.9859	0.9969	0.9656	0.9960	0.983	0.9817
	1.003	0.9987	1.4415	0.9957			1.865	0.9745
			0.9387	0.999			1.172	0.9824

Table 3.2: Scalings of linear least squares fittings for log log plots of  $\mathcal{L}$  and  $\mathcal{J}$ , along with accompanying  $R^2$  values. Initial times for each scaling is shown in Fig. (3.11) and (3.12).

The deviation of  $\mathcal{L}$  and  $\mathcal{J}$  in time from an exact linear increase leads to a further interesting behaviour in  $\mathcal{J} - \mathcal{L}^2/T$ . This is shown in Fig. (3.13) by using the same data used in Fig. (3.9) and Fig. (3.10). Specifically, Fig. (3.13) shows the log log plots of  $\mathcal{J} - \mathcal{L}^2/T$  from the initial time  $T_0$  where  $\mathcal{J} \neq \mathcal{L}^2/T$ . The linear least squares fit is shown with the dashed line. Each plot quantitatively shows that  $\mathcal{J} - \mathcal{L}^2/T$  increases linearly in time to leading order, with a (small) time varying exponent  $m$ . The lines of best fit in Fig. (3.13) for Beethoven, Mozart, Tchaikovsky and Vivaldi are given in Table 3.3. The two different values of exponent  $m$  are shown in Table 3.3 for Beethoven's ninth symphony, 2nd movement, which has two distinct scaling regimes of different times.

Composer	$T_0$ , (sec)	$m$	$R^2$
Beethoven	$T_0 = 7.25/23.12$	0.4927/-0.0135	0.8023/0.9988
Mozart	$T_0 = 5.75$	0.0306	0.9991
Tchaikovsky	$T_0 = 1.0$	0.0622	0.9931
Vivaldi	$T_0 = 7.35$	-0.1422	0.9966

Table 3.3: The initial time  $T_0$  when the line of best fit was taken for  $\mathcal{J} - \mathcal{L}^2/T$  for each composition. The exponent  $m$  and the  $R^2$  values are also included.

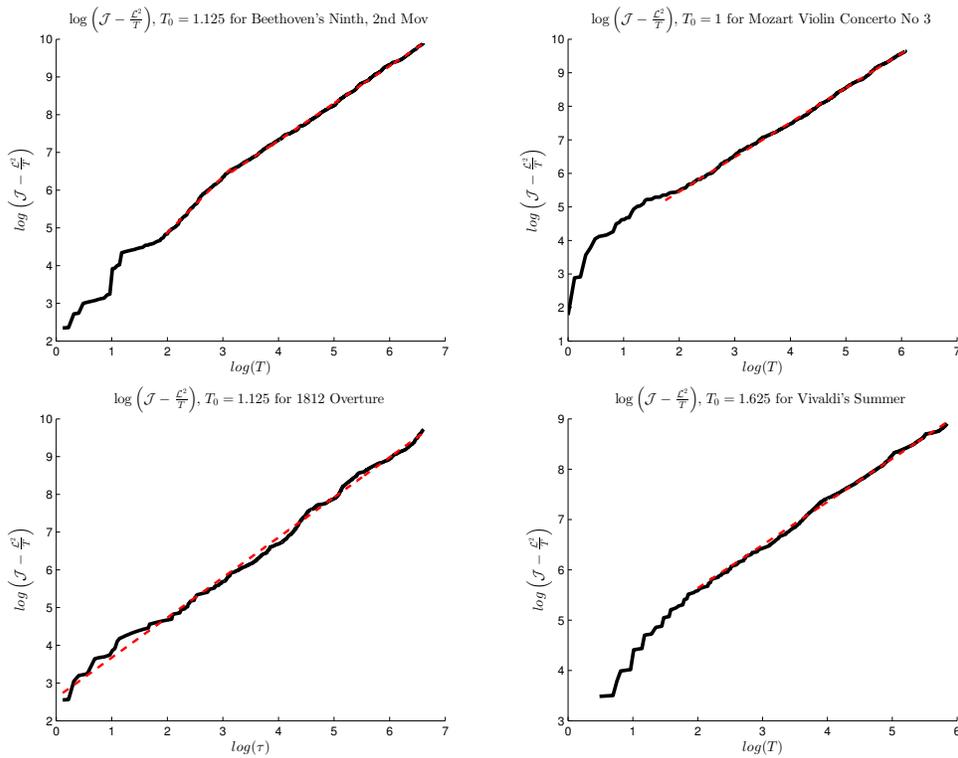


Figure 3.13: log log plots of Eq. (3.46) against  $T$ . Each plot follows an approximate power law.

### 3.6.3 Oscillations

Power-law scaling determined in the previous subsection are approximately leading order behaviours, and just analysing these overlooks some of the most interesting details of the evolutions. To calculate the deviation from power law, we compute  $V_T$  from Eq. (3.47) and plot the fluctuations for each composition in Figs. (3.14)-(3.17). For each piece of music in Figs. (3.14)-(3.17),  $V_T$  initially undergoes a significant increase associated with the beginning of the music and then rapid and damped fluctuations (shown with solid red line). This initial transient is followed by a smoother evolution plotted with the dashed line. To analyse  $V_T$  we take the Fourier transform (FT) of  $V_T$  as  $F[V_T] = \int_{T_0}^T V_T e^{-i2\pi ft} dt$  by using the data taken from  $T_0$  and show the corresponding power spectra  $S(V_T) = |F[V_T]|^2$  in Figs. (3.14)-(3.17). Note that using the entire piece of music for  $\Delta t \leq t \leq T$  would have resulted in one large initial peak of  $S(V_T)$  corresponding to the length of the piece, obscuring all interesting behaviour coming from smaller amplitudes. However, taking the FT over only the dashed regions in Fig. (3.14)-(3.17), i.e. from the minimum of the initial fluctuations enables us to identify the secondary peaks due to the oscillations in  $V_T$ , in addition to the dominant peak corresponding to the inverse of the total time duration of the music. For Vivaldi's Summer, the second main peak occurs at  $f = 9.4937 \times 10^{-3}$  Hz. This oscillation in  $V_T$  is shown in Fig. (3.14) where a corresponding period of  $\tau \approx 105.2$  seconds is marked.

The fluctuations in Beethoven's Ninth symphony, 2nd movement has a distinct oscillation between approximately  $200 \leq T \leq 400$  seconds, which is seen in the power spectrum in Fig. (3.15). The 1812 Overture has the most complicated power spectrum, showing a series of peaks in Fig. (3.16). Mozart reveals one large peak in Fig. (3.17) apart from the dominant low frequency peak.

By interpreting music in terms of a flow of information we were able to compute  $\mathcal{L}$  and  $\mathcal{J}$  for famous classical composers (Mozart, Vivaldi, Tchaikovsky and Beethoven) from midi-files and investigated its temporal variation in PDFs which are strongly intermittent. The fact that the well known relation  $\mathcal{J} - \mathcal{L}^2/T$

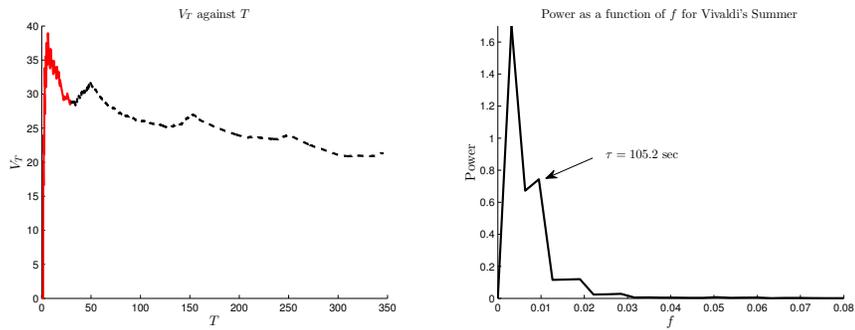


Figure 3.14: Vivaldi's Summer

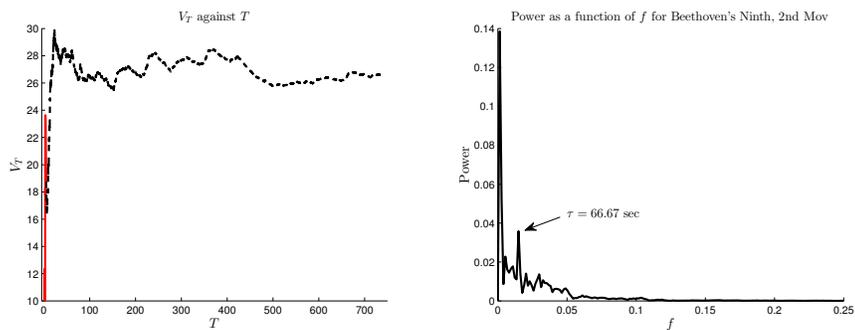


Figure 3.15: Beethoven's Ninth, 2nd Movement.

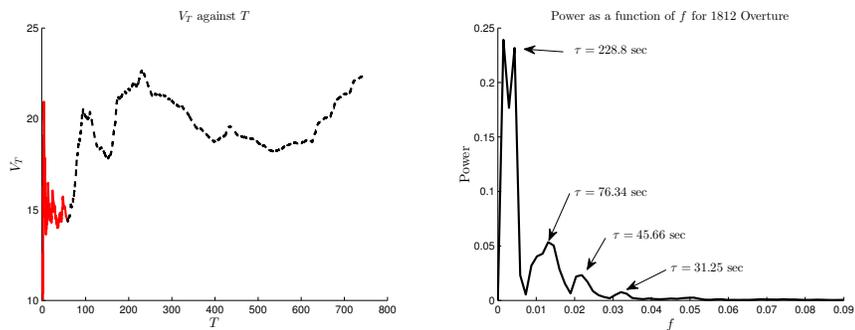


Figure 3.16: Tchaikovsky's 1812 Overture.

follows such a simple relation is quite interesting, especially when one considers the form of the PDFs used to generate the power law. There appears to be

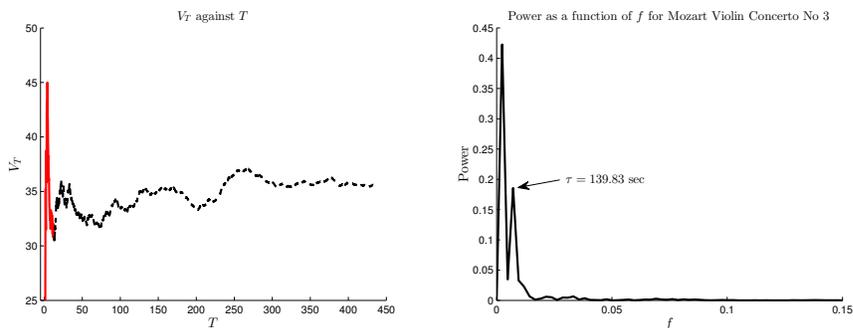


Figure 3.17: Mozart's Violin Concerto No. 3.

a balance between the energy used as the music evolves through statistical space and the distance it travels. Further work will be to determine if this relation is solely applicable to music or if other systems have this balance to their evolutions.

### 3.6.4 Continuous signals

A nice comparison to  $\mathcal{L}$  and  $\mathcal{J}$  made from *midi* files is to use true recordings. This consists of the amplitudes of recorded sound as shown in Fig. (3.18) which are used to instruct a speaker how to vibrate so that we hear a song. To analyse the signal we again have to use a discrete time step, since the signal generated is continuous but the amplitudes used to generate this sound are discrete. Since the time steps between amplitudes are very small,  $dt = 2.265 \times 10^{-5}$  sec, we will have far smaller errors in computing, Eq. (3.19). Each PDF will be made up from five time steps from the amplitudes, giving  $\Delta t = 1.133 \times 10^{-4}$  sec. Using  $\Delta t$  instead of  $dt$  avoids having the trivial case where  $\Delta \mathcal{L}$  is either one or zero since  $dt$  is made up of one point. To compare  $\mathcal{L}$  and  $\mathcal{J}$  generated from the continuous signal to  $\mathcal{L}$  and  $\mathcal{J}$  generated from *midi* files, we analysed recordings of Beethoven's Moonlight sonata performed by Glenn Morrison and Tchaikovsky's 1812 Overture performed by the Herlev Concert Band. All recordings presented below are freely available on the internet at the

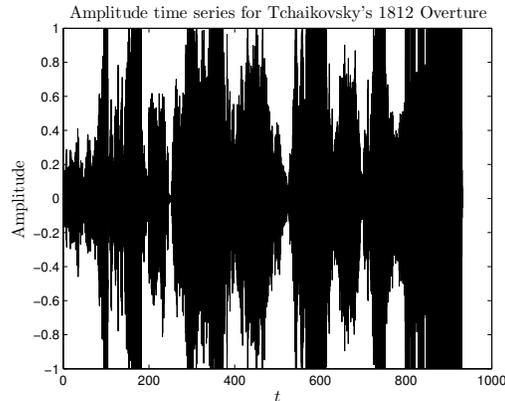


Figure 3.18: The time series for Tchaikovsky's 1812 Overture.

time of writing<sup>4</sup>. For Tchaikovsky's 1812 Overture in Fig. (3.19) we see smaller fluctuations than we did using the midi files. This leads to  $\log(\mathcal{J} - \mathcal{L}^2/T)$  to a very good approximation following the minimum path through statistical space. The log log plot in (a) has respective slopes for  $\mathcal{L}$  and  $\mathcal{J}$  of  $m_{\mathcal{L}} = 1.0763$  and  $m_{\mathcal{J}} = 1.0978$ . In (c) we see that  $\mathcal{J} - \mathcal{L}^2/T$  appears to have a complicated time dependent behaviour. Taking the Fourier transform (FFT) in (d) and plotting the log of the power gives a very good approximation of a power law. The slope of which is  $\alpha = 1.9949$ , which is approximately Brownian motion.

Beethoven's Moonlight Sonata Fig. (3.20) also has an evolution very close to the minimum path, with  $m_{\mathcal{L}} = 1.0469$  and  $m_{\mathcal{J}} = 1.0543$  as shown in (a). (b-c) show the log-log plot and simply  $\mathcal{J} - \mathcal{L}^2/T$  which are both very linear up until the end of the piece, where we see a dramatic increase in the slope. This increase appears to coincide with the performer changing the melody during the final few seconds of the performance. (d) shows the power which results from the FFT. Here again, despite the large increase in slope of  $\mathcal{J} - \mathcal{L}^2/T$ , the fluctuations are extremely close to Brownian motion,  $\alpha = 1.9886$ .

So far all sound files from both continuous recordings and midi files are close to the minimum path. It is of interest to then see if other systems in nature

---

<sup>4</sup>Both musical recordings are available at [www.soundcloud.com](http://www.soundcloud.com).

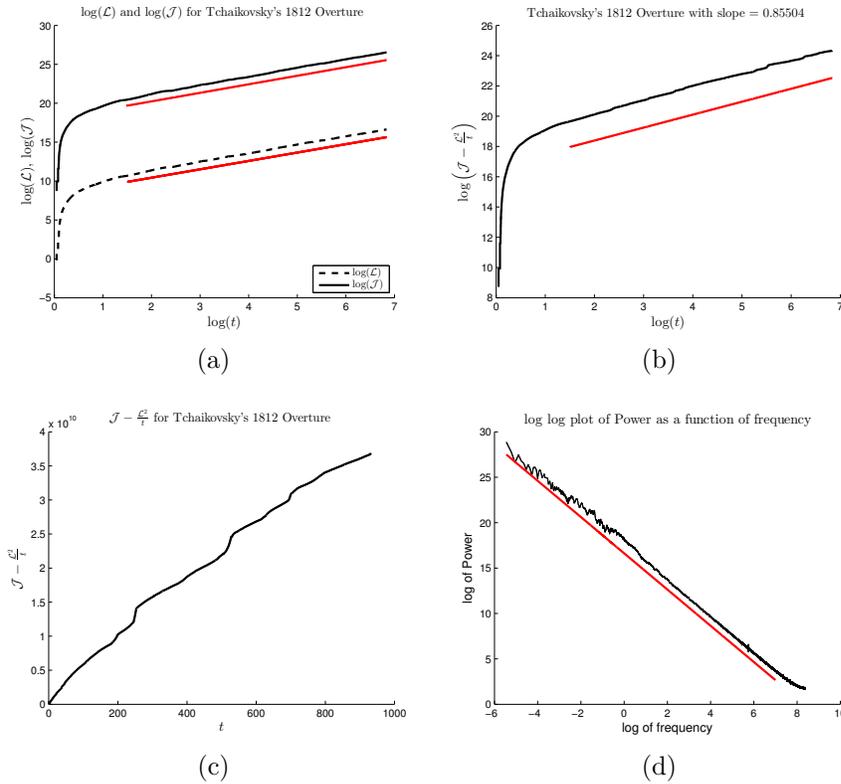


Figure 3.19: (a) Both  $\log(\mathcal{L})$  and  $\log(\mathcal{J})$  show strong power law behaviour as before. The slopes of best fit are  $m_{\mathcal{L}} = 1.0763$  and  $m_{\mathcal{J}} = 1.0978$  respectively. (b) The log log plot of  $\mathcal{J} - \mathcal{L}^2/T$  has a slope close to one, meaning the system takes the minimum path through statistical space. In (c) we see that there is a substantial time dependent term  $m(t)$ . In (d) the log log plot of the power is shown. This has a slope of  $\alpha = 1.9949$ .

also follow this minimum path. Fig. (3.21) shows the results from a recording of humpback whales. This recording was made by submerging a microphone into the sea from a kayak. Though the recording has some background noise the sound from the whales are extremely close to the minimum path. In fact they are an order of magnitude closer than the music files just analysed having an overall slopes for  $\mathcal{L}$  and  $\mathcal{J}$  of  $m_{\mathcal{L}} = 1.0030$  and  $m_{\mathcal{J}} = 1.0027$ . For  $\mathcal{J} - \mathcal{L}^2/T$  in (b-c) we see two distinct slopes. These slopes appear to correspond to the

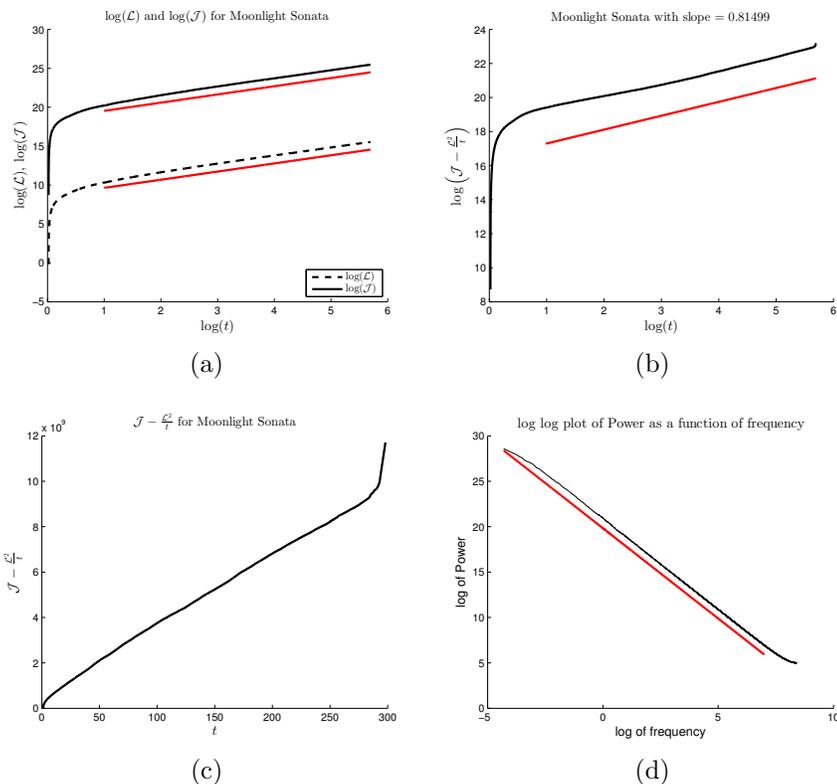


Figure 3.20: (a) log log plots of  $\mathcal{L}$  and  $\mathcal{J}$  both of which follow power law evolutions. The slopes of best fit are  $m_{\mathcal{L}} = 1.0469$  and  $m_{\mathcal{J}} = 1.0543$  respectively. (b) log log plot of  $\mathcal{J} - \mathcal{L}^2/T$ . The slope shows the system is very close to the minimum path. (c)  $\mathcal{J} - \mathcal{L}^2/T$  shows very small fluctuations. (d) Taking log of the fluctuations show a slope of  $\alpha = 1.9886$ .

animals using different vocalizations. The fluctuations also follow a power law as before, though this time they appear to be slightly farther from Brownian motion at,  $\alpha = 1.9580$ .

The three systems studied have followed  $\mathcal{J} - \mathcal{L}^2/T \propto T^{1+m(t)}$ . The fluctuations though all being small  $m(t) \ll 1$ , each approximately follow the power law relation  $1/f^2$ . Since each system has had small  $m(t)$ , it would be nice to compare these results to a completely different system and see if it might have a different fluctuations. The system we have chosen is a recording of a seven

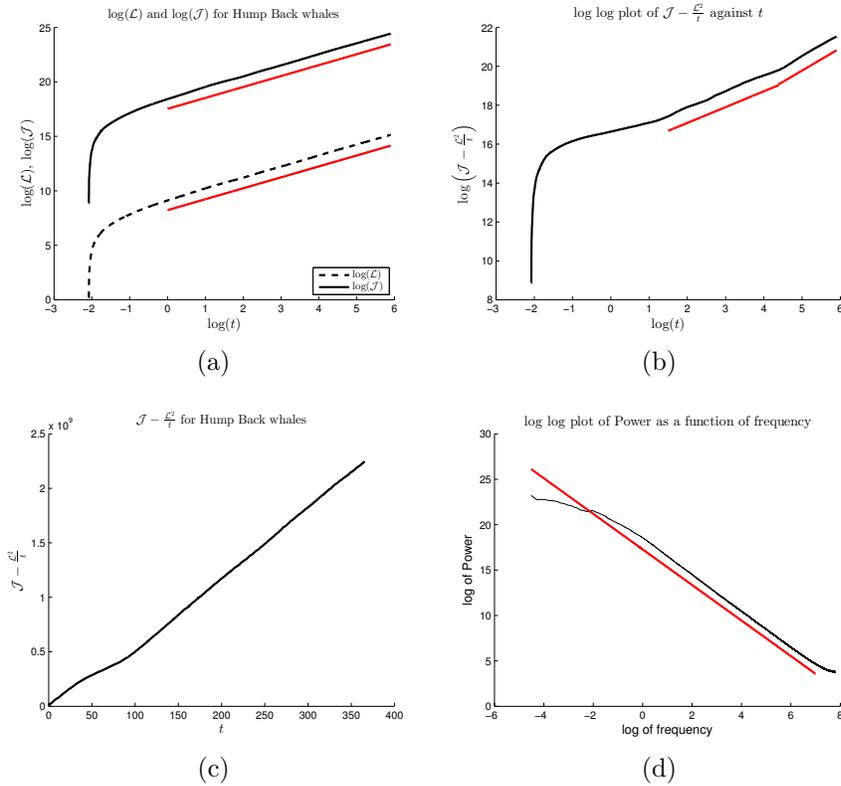


Figure 3.21: (a) log log plots of  $\mathcal{L}$  and  $\mathcal{J}$  both of which follow power law evolutions. The slopes of best fit are,  $m_{\mathcal{L}} = 1.0030$  and  $m_{\mathcal{J}} = 1.0027$  respectively (b) log log plot of  $\mathcal{J} - \mathcal{L}^2/T$ . There are two distinct linear slopes,  $m = -0.1881$  and  $m = 0.1552$ . (c) Both of the trends in  $\mathcal{J} - \mathcal{L}^2/T$  are clearly seen. (d) shows the log log of the Power as a function of frequency, and here the slope is slightly farther from Brownian motion,  $\alpha = 1.9580$ .

minute walk through the city of Sheffield England. This included the wind blowing into the microphone, buses and cars passing by, birds singing and human conversation. One can think of this system then as a collection of organized information that is randomly combined together. Looking at Fig. (3.22) (a) we see that to a good approximation  $\mathcal{L}$  and  $\mathcal{J}$  again follow an approximate power law, with  $m_{\mathcal{L}} = 1.0839$  and  $m_{\mathcal{J}} = 1.1153$ . (b-c) show  $\mathcal{J} - \mathcal{L}^2/T$ , which has an exponent that is very small,  $m(t) = 0.0431$ . Given the small exponent, we

might expect (d) to show slope near 2 as the others do. Essentially the power plotted in log log form confirms this though the slope is slightly farther from Brownian motions at  $\alpha = 1.9145$ .

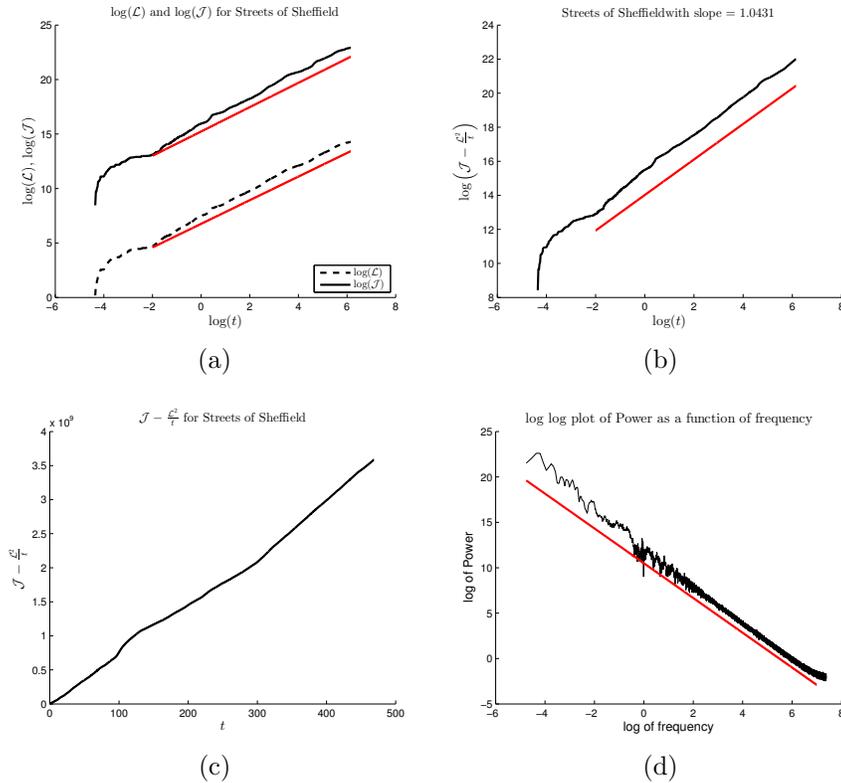


Figure 3.22: (a) shows  $\log(\mathcal{L})$  and  $\log(\mathcal{J})$ . Both have power law behaviour with  $m_{\mathcal{L}} = 1.0839$  and  $m_{\mathcal{J}} = 1.1153$ . (b) shows  $\log(\mathcal{J} - \mathcal{L}^2/T)$  and a time varying exponent of  $m(t) = 0.0431$ . (c) shows  $\mathcal{J} - \mathcal{L}^2/T$  which has two odd trends. The log log plot of the power spectrum is given in (d) with  $\alpha = 1.9145$ .

The continuous signal from each system is the result of sound being combined from each instrument or source. When we look at the power spectrum from these diverse systems they all are approximately Brownian motion, as their frequency spectrum is very near  $1/f^2$ . Yet if we look directly at the fluctuations we see marked differences between the music recording, whales and city noise. Both

musical compositions and whales have a large maximum at the beginning of the signal with a then more or less steady decrease. The city recording though has multiple local maxima and minima.

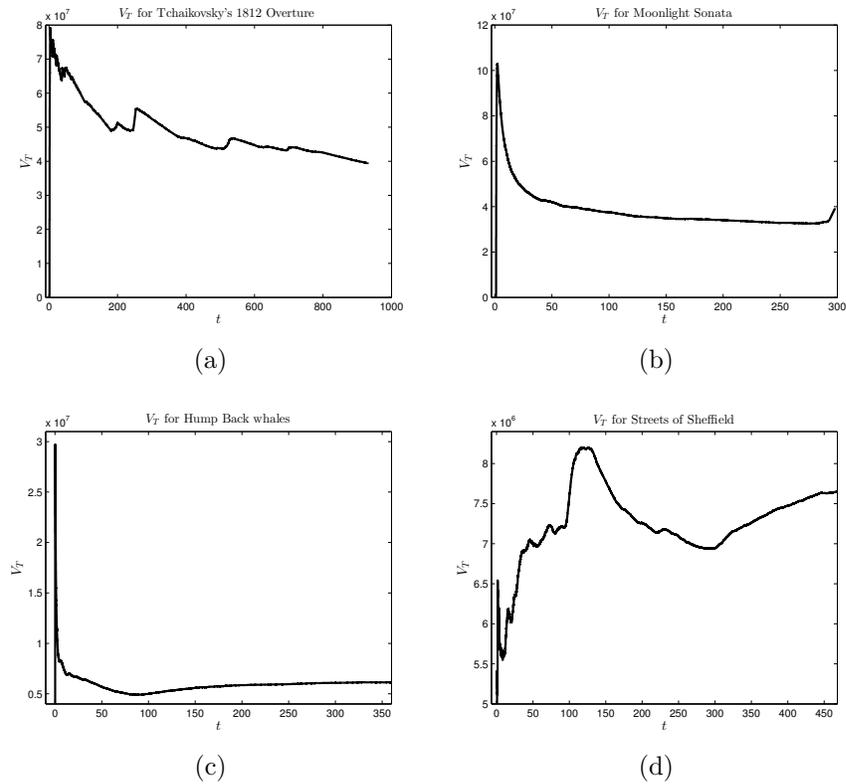


Figure 3.23: Tchaikovsky's 1812 Overture (a), Bethoven's Moonlight Sonata (b) and Humpback whales (c) all have strong initial fluctuations which then decrease with time. The walk through Sheffield on the other hand has a large fluctuations but then has multiple local minimums and maximums in the coarse of the recording. This arises from the recording being made up of several independent sources (see text for examples).

It seems music and at least hump back whales often follow a very precise path through configuration space, namely the minimum path. This minimum path is arguably the path of minimum dissipation out of equilibrium and is certainly the path of minimum dissipation in equilibrium. The only way to

differentiate each system is then through its fluctuations. Yet as diverse as our system's are, all of them have time dependent exponents,  $m(t)$  which appear to follow  $1/f^2$  distribution despite their very different appearances in Fig. (3.23). The caveat to following a  $1/f^2$  distribution is that each power spectrum has a high frequency cut off ( $1/T_{min}$ ) where the power law fails. This cutoff is around the time step used for the system,  $T_{min} = \Delta t$ .

# Chapter 4

## Conclusion

Statistical mechanics has proven an invaluable tool in almost every branch of physics. Yet until recently, progress in non-equilibrium statistical mechanics has been slow to catch up. In this work we have shown that with the added complexity inherent in non-equilibrium systems, order and structure can emerge. Specifically the Observable Representation and its irreversible extension the NOR, allow us to organize the information of the Perron-Frobenius operator into a space of distances. Though we do not have the complete story of the NOR, we have illustrated that for representation and control of complex systems the NOR can be a valuable tool. The thermodynamic length, while having a longer history than the NOR, is also still an active field of research. This is perhaps where we will realize one of the simplest goal of non-equilibrium statistical mechanics, that of knowing in general how to calculate the average of a set of observables as the system evolves. For if we can understand the rule of how a path is formed in statistical space, we may be able to explicitly calculate the distributions which give this path.

# Bibliography

- [1] F. Weinhold. Metric geometry of equilibrium thermodynamics. *Journal of Chemical Physics*, 63:2479, 1975.
- [2] C. Jarzynski. *Physical Review Letters*, 78:2690–2693, 1997.
- [3] Gavin E. Crooks. *Physical Review E*, 60:2721–2726, 1999.
- [4] B Gaveau and L.S. Schulman. *Physics Letters A*, 229:347, 1997.
- [5] B. Gaveau and L.S. Schulman. *Journal of Mathematical Physics*, 39:1517, 1998.
- [6] John Gribbin. *Deep Simplicity*, volume 1. Penguin, 2004.
- [7] S.D.M White and M.J. Rees. *Monthly Notices Royal Astronomy Society*, 183:341–358, 1978.
- [8] Sami K. Solanki. *The Astronomy Astrophysics Review*, 11:153–286, 2003.
- [9] V. Ramanathan and W. Collins. *Nature*, 351:27–32, 1991.
- [10] Nitzan Rosenfeld, Jonathan W. Young, Uri Alon, Peter S. Swain, and Michael B. Elowitz. *Science*, 307:1962–1965, 2005.
- [11] Kerson Huang. *Statistical Mechanics*, volume 2. John Wiley and Sons, 1983.
- [12] Ralph Baierlein. *Thermal Physics*. Cambridge University Press, 2000.

- [13] Richard Feynman. *The Feynman lectures on physics*.
- [14] Alan Slomson. *An Introduction to Combinatorics*, volume 1. Chapman and Hall Mathematics, 1991.
- [15] Hans-Otto Georgii. Probabilistic aspects of entropy. *International Symposium on Entropy*, pages 1–15, 2000.
- [16] Leonard Susskind. the theoretical minimum, Statistical Mechanics. <http://theoreticalminimum.com/courses/statistical-mechanics/2013/spring>, 2015. [Online; accessed 13-February-2015].
- [17] E.T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 4, 1957.
- [18] P.J. Morrison. Hamiltonian description of the ideal fluid. *Modern Physics*, 70, 1998.
- [19] Cosma Rohilla Shalizi. The backwards arrow of time of the coherently bayesian statistical mechanic. *ArXiv*.
- [20] Edward Ott. *Chaos in Dynamical Systems*, volume 1. cambridge university press, 1993.
- [21] Predrag Cvitanović, Roberto Artuso, Per Dahlenqvist, Ronnie Mainien, Gregor Tanner, Gábor Vattay, Niall Whelan, and Andreas Wirzba. *Chaos Classical and Quantum*. ebook, chaosbook.org.
- [22] C. Beck and F. Schlögl. *Thermodynamics of chaotic systems*. Cambridge University Press, 1993.
- [23] J.P Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Review of Modern Physics*, 57:617–656, 1985.
- [24] Marco Sandri. Numerical calculation of lyapunov exponents. *The Mathematica Journal*, 1996.

- [25] V.I. Oseledet. A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems. *Transactions of The Moscow Mathematical Society*, 19:197–231, 1968.
- [26] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2009.
- [27] Jr Frederick W. Byron and Robert W. Fuller. *Mathematics of Classical and Quantum Physics*, volume 1. Dover, 1992.
- [28] Steven H. Strogatz. *Nonlinear dynamics and Chaos*, volume 1. Westview Press, 2000.
- [29] Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20:130–141, 1963.
- [30] Lord Rayleigh. On convection currents in a horizontal layer of fluid when the higher temperature is on the under side. *Philosophical Magazine and Journal of Science*, 32:529–546.
- [31] E.L. Koschmieder. *Bénard Cells and Taylor Vortices*. Cambridge Monographs on Mechanics and Applied Mathematics, 1993.
- [32] C.E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:623–656, 1948.
- [33] A.I. Khinchin. *Mathematical Foundations of Information Theory*. Dover, 1957.
- [34] Christian Beck. Generalized information and entropy measures in physics. *Contemporary Physics*, 50:495–510, 2009.
- [35] Thomas M. Cover and Joy A. Thomas. *Elements of Information theory*, volume 2. Wiley, 2006.

- [36] Alfréd Rényi. On measures of entropy and information. *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability*, pages 547–561, 1960.
- [37] Constantino Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal of Statistical Physics*, 52:479–487, 1988.
- [38] G. Baris and Ugur Tirnakli. On the way towards a generalized entropy maximization procedure. *Physics Letters A*, 373:3230–3234, 2009.
- [39] A.G. Bashkirov. Renyi entropy as a statistical entropy for complex systems. *Theoretical and Mathematical Physics*, 149(2):1559–1573, 2006.
- [40] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical statistics*, 22:79–86, 1951.
- [41] Annick Lesne. Shannon entropy: a rigorous mathematical notion at the crossroads between probability, information theory, dynamical systems and statistical physics. *Mathematical Structures in Computer Science*, 24, 2014.
- [42] B. R. Frieden. *Science from Fisher Information*, volume 2. Cambridge University Press, 2004.
- [43] A.T. Bharucha-Reid. *Elements of the Theory of Markov Processes and their Applications*, volume 2nd. Dover Publications Inc, 1988.
- [44] Suanrong Chen and Jiu Ding. A piecewise constant method fro frobenius-perron operators via delta function approximateions. *Applied Mathematics and Computations*, 219:1047–1052, 2012.
- [45] B. Gaveau and L.S. Schulman. Dynamical distance: coarse grains, pattern recognition and network analysis. *Bulletin Des Sciences Mathématiques*, 129:631–642, 2005.

- [46] R. Graham and H. Haken. Generalized thermodynamic potential for markoff systems in detailed balance and far from thermal equilibrium. *Z. Physik*, 243, 1971.
- [47] N.G. Van. Kampen. *Stochastic processes in physics and Chemistry*, volume 3. Elsevier, 2007.
- [48] L.S. Schulman and B. Gaveau. Course grains: The emergence of space and order. *Foundations of Physics*, 31:713–731, 2001.
- [49] B Gaveau, L.S. Schulman, and L.J. Schulman. *J. Phys. A*, 39:10307, 2006.
- [50] L.S Schulman and J.P. Bagrow. Visualizing relations using the observable representation. *Advances in Complex Systems*, 14:829–851, 2011.
- [51] S.B. Nicholson, L.S. Schulman, and Eun jin Kim. Deciphering interactions of complex systems that do not satisfy detailed balance. *Physics Letters A*, 377:1810–1813, 2013.
- [52] S. Nicholson, E.Kim, and L.S. Schulman. A master equation approach to deciphering non-detailed balance systems. *Chaotic Modeling and Simulation*, 4:569–574, 2012.
- [53] J.R. Norris. *Markov Chains*. Cambridge University Press, 1997.
- [54] P. Tallapragada. Computation of finite time lyapunov exponents using the perron-frobenius operator. *arXiv:1101.4338*, 2011.
- [55] Gary Froyland and Michael Dellnitz. Detecting and locating near-optimal almost-invariant sets and cycles. *Siam J. Sci Computation*, 6:1839–1863.
- [56] J. Nulton, P. Salamon, B Andresen, and Qi Anmin. Quasistatic processes as step equilibrations. *J Chem. Phys*, 83:334–338, 1985.
- [57] Gavin E. Crooks. Measuring thermodynamic length. *Physical Review Letters*, 99, 2007.

- [58] I. Procaccia and R. D. Levine. Potential work: A statistical-mechanical approach for systems in disequilibrium. *The Journal of Chemical Physics*, 65:3357–3364, 1976.
- [59] W.K. Wootters. Statistical distance in hilbert space. *Physical Review D*, 23:357–362, 1981.
- [60] J.W. Gibbs. *The collected works*, volume 1. Longmans, Green and Co, New York, 1928.
- [61] G. Rupeiner. Thermodynamics: A riemannian geometric model. *Phys. Rev. A.*, 20:1608–1613, 1979.
- [62] P. Salamon, J. Nulton, and E. Ihrig. On the relation between entropy and energy versions of thermodynamic length. *Journal of Chemical Physics*, 80:436–437, 1983.
- [63] Matteo Polettini and Massimiliano Esposito. Nonconvexity of the relative entropy for markov dynamics: A fisher information approach. *Physical Review E.*, 88, 2013.
- [64] Edward H. Feng and Gavin E. Crooks. Far-from-equilibrium measurements of thermodynamic length. *Physical review E*, 79, 2009.
- [65] Sergio Bordel. Non-equilibrium statistical mechanics: partition functions and steepest entropy increase. *Journal of Statistical Mechanics: Theory and Experiment*, 2011.
- [66] A. Plastino A.R. Plastino, M. Casas. Fisher’s information, kullback’s measure, and h-theorems. *Physics Letters A*, 246, 1998.
- [67] B.R. Frieden. Estimation of distribution laws and physical laws. *Physica A*, 198, 1993.
- [68] S.B. Nicholson and Eun jin Kim. Investigation of the statistical distance to reach stationary distributions. *Physics Letters A*, 379:83–88, 2014.

- [69] Peter Salamon and R. Stephen Berry. Thermodynamic length and dissipated availability. *Physical Review Letters*, 51, 1983.
- [70] B. R. Frieden. Lagrangians of physics and the game of fisher-information transfer. *Phys. Rev. E.*, 52:2274–2286, 1995.
- [71] E. Lutwak, D. Yang, and G. Zhang. Cramer-rao and moment-entropy inequalities for renyi entropy and generalized fisher information. *IEEE Transactions on Information Theory*, 51:473–478, 2005.
- [72] Marcus du Sautoy. *The music of the primes*. Harper Perennial, 2003.
- [73] Tzanetakis G and Cook P. Musical genre classification of audio signals. speech and audio processing. *IEEE transactions*, 10:293–302, 2002.
- [74] C. Madden. *Fractals in Music: Introductory Mathematics for Musical Analysis*. High Art Press, 1999.
- [75] Manaris B, Romero J, Machado P, Krehbiel D, and Hirzel T. Zipf’s law, music classification and aesthetics. *Computer Music Journal*, 29:55–69, 2005.
- [76] Liu X, Tse C, and Small M. Complex network structure of musical compositions: Algorithmic generation of appealing music. *Physica A*, 389:126–132, 2010.
- [77] Richard F. Voss and John Clarke. ‘1/f noise’ in music and speech. *Nature*, 258:317–318, 1975.
- [78] Joan Serrà, Tan Hakan, Özaskan, and Josep Lluís Arcos. Note onset deviations as musical piece signatures. *PLOS ONE*, 8, 2013.
- [79] Zhi-Yuan Su and Tzuyin Wu. music walk, fractal geometry in music. *Physica A*, 2007.

- [80] Kenneth J., Hsü, and Andrew Hsü. Self-similarity of the “1/f noise” called music. *Proceedings of the National Academy of Science*, 88:3507–3509, 1991.
- [81] Lu Liu, Jianrong Wei, Huishu Zhang, Jianhong Xin, and Jiping Huang. A statistical physics view of pitch fluctuations in the classical music from bach to chopin: Evidence of scaling. *PLOS ONE*, 8, 2013.
- [82] Shun ichi Amari. Information geometry on hierachy of probability distributions. *IEEE Transactions on Information Theory*, 47, 2001.